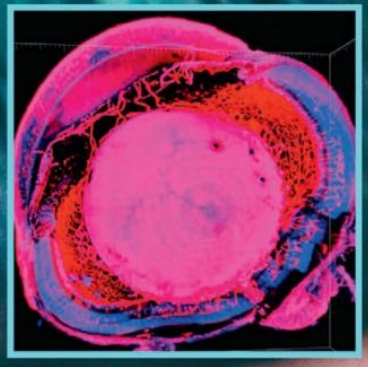
The Eye

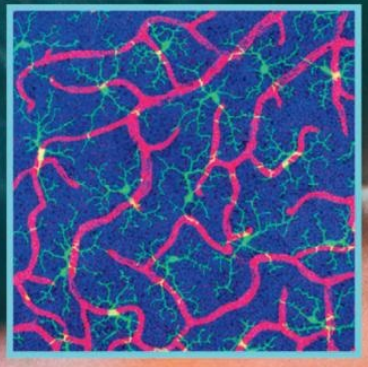
Basic Sciences in Practice

Fifth Edition

John V. Forrester
Andrew D. Dick
Paul G. McMenamin
Fiona Roberts
Eric Pearlman







ELSEVIER

The Eye

Basic Sciences in Practice

Fifth Edition

John V. Forrester

MB ChB MD FRCS(Ed) FRCP(Glasg)(Hon) FRCOphth(Hon) FMedSci FRSE FARVO

Emeritus Professor of Ophthalmology, Department of Ophthalmology, University of Aberdeen, Aberdeen, UK; Section of Immunology and Infection, University of Aberdeen, UK; Ocular Immunology Program, The University of Western Australia, Australia; Centre for Experimental Immunology, Lions Eye Institute, Western Australia, Australia

Andrew D. Dick

BSc MB BS MD FRCP FRCS FRCOphth FMedSci FRSB FARVO

Professor of Ophthalmology, University of Bristol; Duke-Elder Professor of Ophthalmology and Director, UCL-Institute of Ophthalmology, University College London, London, UK

Paul G. McMenamin

BSc MSc(MedSci) DSc (Med) PhD FARVO FRSB

Emeritus Professor, Department of Anatomy and Developmental Biology, Monash University, Melbourne, Victoria, Australia

Fiona Roberts

BSc MBChB MD FRCPath

Consultant Ophthalmic Pathologist, Queen Elizabeth University Hospital, Glasgow, UK

Eric Pearlman

BSc PhD

Director, Institute of Immunology; Professor, Departments of Ophthalmology and Physiology and Biophysics, University of California, Irvine



ELSEVIER

For additional online content visit ExpertConsult.com

ALGrawany

© 2021, Elsevier Limited. All rights reserved.

First edition 1996
Second edition 2002
Third edition 2008
Fourth edition 2016
Fifth edition 2021

The right of John V. Forrester, Andrew D. Dick, Paul G. McMenamin, Fiona Roberts and Eric Pearlman to be identified as authors of this work has been asserted by them in accordance with the Copyright, Designs and Patents Act 1988.

No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or any information storage and retrieval system, without permission in writing from the publisher. Details on how to seek permission, further information about the publisher's permissions policies and our arrangements with organizations such as the Copyright Clearance Center and the Copyright Licensing Agency can be found at our website: www.elsevier.com/permissions.

This book and the individual contributions contained in it are protected under copyright by the publisher (other than as may be noted herein).

Notices

Practitioners and researchers must always rely on their own experience and knowledge in evaluating and using any information, methods, compounds or experiments described herein. Because of rapid advances in the medical sciences, in particular, independent verification of diagnoses and drug dosages should be made. To the fullest extent of the law, no responsibility is assumed by Elsevier, authors, editors or contributors for any injury and/or damage to persons or property as a matter of products liability, negligence or otherwise, or from any use or operation of any methods, products, instructions, or ideas contained in the material herein.

ISBN: 978-0-7020-7993-1

Content Strategist: Jeremy Bowes
Content Development Specialist: Helen Leng
Content Coordinator: Gill Cloke
Project Manager: Andrew Riley
Design: Amy Buxton
Illustration Manager: Deanna Sorenson

Printed in the UK

Last digit is the print number: 9 8 7 6 5 4 3 2 1

ELSEVIER Book Aid International Working together to grow libraries in developing countries
www.elsevier.com • www.bookaid.org

Contents

Preface vi
Acknowledgements vii

1. Anatomy of the eye and orbit 1
2. Embryology and early development of the eye and adnexa 106
3. Genetics 135
4. Biochemistry and cell biology 162
5. Physiology of vision and the visual system 275
6. General and ocular pharmacology 349
7. Immunology 380
8. Microbial infections of the eye 464
9. Pathology 492

Addendum

Covid-19 and the eye 541

Index 543

eBook Content

For the fifth edition of *The Eye*, additional content is provided in the eBook (https://expertconsult.inkling.com/) and accessed at the point where the icon appears.

The authors have prepared video clips to help explain and expand on aspects of the basic science and these can be accessed where the video icon appears.

All suggestions for further reading are in the eBook.

Preface

The Eye has once again been fully updated in this new fifth edition. This textbook is aimed at all students and trainees, scientists and researchers, physicians and optometrists who wish to acquire a foundation in the basic sciences of ophthalmology and vision. This is a very broad field since it encompasses all aspects of the life sciences. Indeed, *The Eye* may also be of interest to established academic ophthalmologists and vision scientists who wish to get up to date with areas of knowledge which are not directly the remit of their clinical practice or research program. In its initial conception, *The Eye* was put together to fill a need for a single, readable, concise text which could be used by a widely diverse group of individuals dipping their toes into the ocean of knowledge that comprises the science behind ophthalmology and vision. This clearly includes the core disciplines of anatomy, embryology, and vegetative physiology of the eye and neurophysiology of vision. However, in order to gain knowledge relevant to disease processes and therapeutic interventions in the eye, broader information from other disciplines such as pathology, pharmacology, microbiology, genetics, and immunology is essential to a clear understanding.

Our understanding of ocular function and disease has advanced exponentially since the publication of the fourth edition. Inevitably this affects some disciplines more than others: advances in genetics have allowed developments in gene therapy and gene editing as well as revealing the important role of epigenetics; microbiology and immunology have been revolutionized by the continuing discoveries relating to the microbiome and in particular the role of the gastrointestinal (GI) tract as an immune organ; drug delivery to the eye remains a major challenge for the pharmacologist; new knowledge of the fluid flow within the eye and the brain has important implications for our concept of anatomical barriers; developments in pathology are providing improved aids to diagnosis. Further, emerging blinding diseases caused by mosquito-borne viral infections including Zika, Dengue and Chikungunya viruses are joining a long list of known pathogens that affect the eye and the CNS. Further, COVID-19 has emerged as a major worldwide pandemic, and while the SARS-CoV-2 virus has not yet been shown to affect the retina, it does invade neurons in the brain.

This new edition of *The Eye* has undergone a major revision in several of the above fields and offers up-to-date, cutting-edge information which will appeal to its core readership, namely trainees in ophthalmology and optometry. We also hope that *The Eye* will continue to appeal to ophthalmology and vision scientists by providing easily accessible basic information in fields of knowledge which are not their primary discipline. We have retained the original format and subsections, since these appear to work well, and we have further developed the eBook material in a more user-friendly and accessible fashion, while the video material remains unchanged. We also hope that the new edition occupies a space in eye research laboratories across the globe where it can act as a quick "wiki-guide" to basic terminology and understanding. This information in its correct context will be particularly valuable to new recruits to the fascinating world of ophthalmic and visual science.

John V. Forrester
Andrew D. Dick
Paul G. McMenamin
Fiona Roberts
Eric Pearlman

Acknowledgements

We would like to thank Professor William R. Lee, who co-authored the first two editions of *The Eye*, for his guidance and support and generous provision of material for the current edition. We are also grateful to the anonymous reviewers who commented on the draft text.

John V. Forrester
Andrew D. Dick
Paul G. McMenamin
Fiona Roberts
Eric Pearlman

Dedication

To Anne, Lindsey, Christine (dec) and Lucy

ALGrawany

Chapter 1

Anatomy of the eye and orbit

CONTENTS

Anatomical terms of reference
Osteology of the skull and orbits
- General arrangement and features of the skull
- Osteology of the orbit
- Cranial cavity (Fig. 1.8A,B and eFig. 1.1)

Structure of the eye
- General shape, size and position of the eye
- The cornea
- The sclera
- Limbus and aqueous outflow pathways (Fig. 1.18)
- The iris
- The ciliary body
- The lens and zonular apparatus
- Anterior and posterior chambers
- The vitreous
- Retina and retinal pigment epithelium (Fig. 1.29)
- The choroid
- Optic Nerve

Orbital contents
- General arrangement
- Periorbita and orbital fibroadipose tissue (Figs 1.43 and 1.44)
- Extraocular muscles
- Blood vessels of the orbit (Fig. 1.50)

Cranial nerves associated with the eye and orbit
- General functional arrangement
- Oculomotor nerve (cranial nerve III) (Figs 1.52 and 1.53)
- Trochlear nerve (cranial nerve IV)
- Abducent nerve (cranial nerve VI)
- Trigeminal nerve (cranial nerve V) (Figs 1.58 and 1.59)
- Facial nerve (cranial nerve VII) (Figs 1.62 and 1.63)

Ocular appendages (adnexa)
- Muscles of the eyelids and adjacent face
- The eyelids
- The conjunctiva
- Lacrimal apparatus (Fig. 1.68)

Anatomy of the visual pathway
- Intracranial portion of the optic nerve (Figs 1.51 and 1.71)
- Optic chiasma
- Optic tracts
- Lateral geniculate bodies
- Optic radiations (geniculocalcarine tracts)
- Primary visual cortex (area 17)
- Secondary visual association areas (areas 18 and 19)
- Frontal eye field
- Retinotopic organization of the visual pathway and visual pathway disturbances (Figs 1.76–1.82)
- Blood supply of the visual pathway

Anatomical terms of reference

The internationally accepted terminology for description of the relations and position of structures in the body requires reference to a series of imaginary *planes* (Fig. 1.1). Thus, relative positions of anatomical structures are referred to in terms of *medial* (nearer the median or mid-sagittal plane) and *lateral* (away from this plane); *anterior* and *posterior* refer to the front and back surfaces of the body; *superior* (cranial or rostral) or *inferior* (caudal) refer to position in the vertical; *superficial* and *deep* specify distance from the surface of the body. A combination of terms can be used to describe the relative position of structures that do not fit exactly any of the other terms, e.g. ventrolateral, posteromedial, etc.

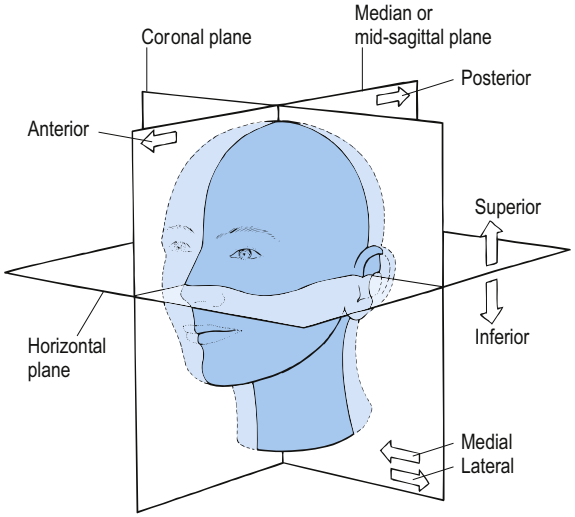


Fig. 1.1 Diagram illustrating anatomical planes of reference.

Osteology of the skull and orbits

General arrangement and features of the skull

The skull is divided into two parts: an upper part shaped like a bowl, which contains the brain, known as the *cranium* or *neurocranium* and a lower part, the *facial skeleton* or *viscerocranium*. The cranium can be further subdivided into the cranial vault and cranial base.

The skull is composed of a large number of separate bones that are united by *sutures* (fibrous immovable joints). The cranium consists of eight bones (only two are paired), the facial skeleton consists of 14 bones, of which only two are single (Fig. 1.2A,B). The skull contains a number of cavities that reflect its multiple functions:

- *cranial cavity*: houses, supports and protects the brain
- *nasal cavity*: concerned with respiration and olfaction
- *orbits*: contain the eyes and adnexa
- *oral cavity*: start of gastrointestinal tract, responsible for mastication and initial food processing; houses taste receptors.

Many of the cranial bones contain air-filled spaces, the paranasal sinuses (Fig. 1.3). Most of the anatomical features of the whole skull relevant to the study of the eye and orbits are indicated in Fig. 1.2A,B (*norma frontalis* and *norma lateralis*). (See Video 1.1.)

Osteology of the orbit

The orbital cavities, situated between the cranium and facial skeleton, are separated from each other by the nasal cavity and the ethmoidal and sphenoidal air sinuses (Fig. 1.3A–C). Each bony orbit accommodates and protects the eye and adnexa and serves to transmit the nerves and vessels that supply the face around the orbit. Parts of the following bones contribute to the walls of the orbit: maxilla, frontal, sphenoid, zygomatic, palatine, ethmoid and lacrimal (Figs 1.4 and 1.5A,B). Each orbit is roughly the shape of a quadrilateral pyramid whose base is the *orbital margin* and whose apex narrows at the *optic canal*. Each orbit has a floor, roof, medial wall and lateral wall (Fig. 1.4). (See Videos 1.2. and 1.3)

The floor tapers off before reaching the apex; therefore, the apex of the pyramid is triangular in shape. The orbit is widest approximately 1.5 cm behind the orbital margin. The medial walls are approximately parallel to the mid-sagittal plane, while the lateral walls are oriented at an angle of approximately 45 degrees to this plane. The *orbital aperture* is directed forwards, laterally and slightly downwards, a characteristic of primates and indeed predators that require binocular vision. Nerves and muscles passing from the apex into the orbit pass forward and laterally (Fig. 1.3A,B). The orbit is approximately 40 mm in height, 40 mm in width and 40 mm in depth. The volume is approximately 30 mL, of which one-fifth is occupied by the eye.

The walls of the orbit

The bones that make up the roof, floor and medial and lateral walls are summarized in Fig. 1.4.

Features of the orbital roof.

- *Fossa for the lacrimal gland:* lies in the anterolateral aspect of the roof behind the zygomatic process of the frontal bone.

A L G r a w a n y

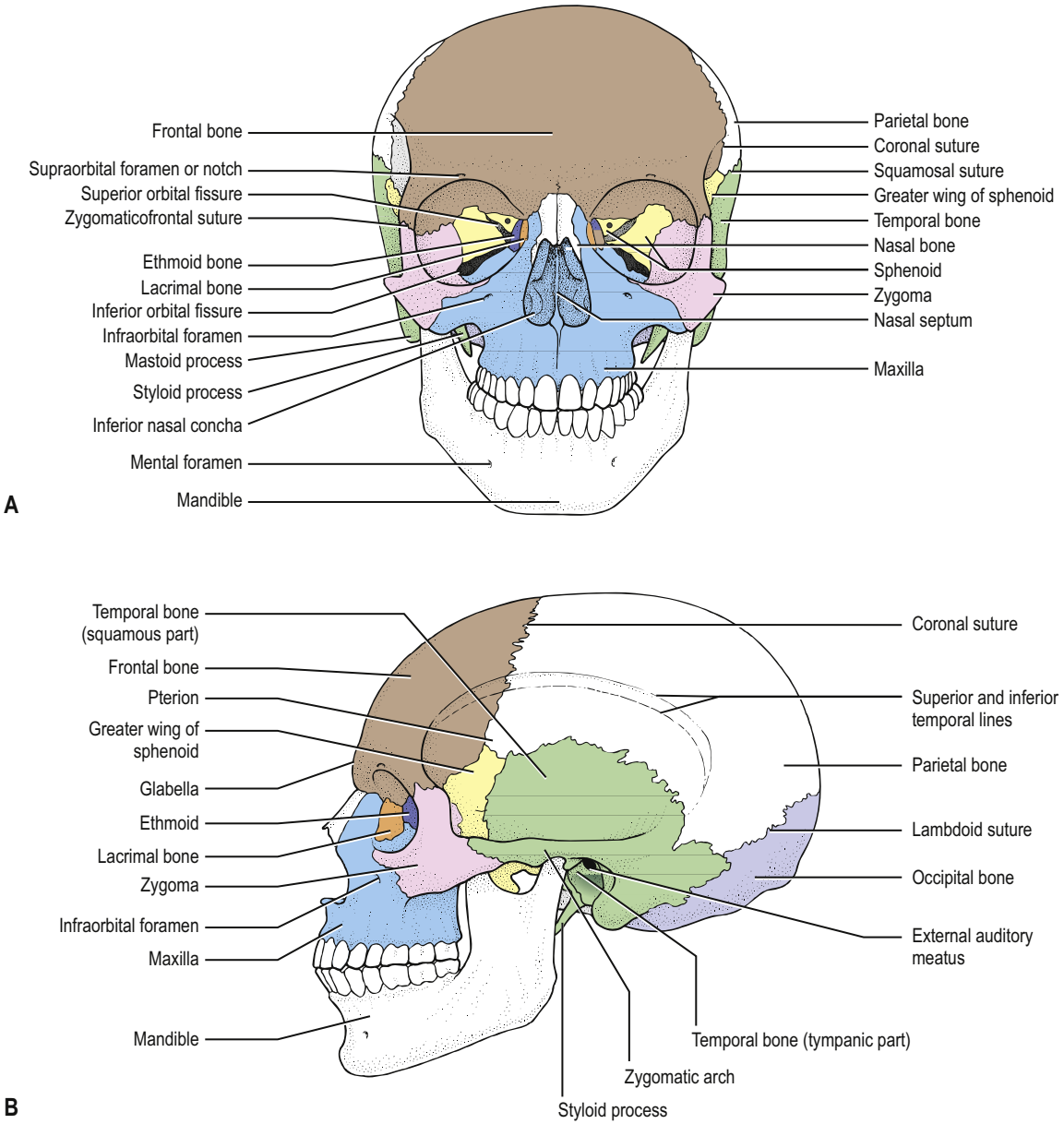


Fig. 1.2 Osteology of the skull. Two views of the skull: (**A**) norma frontalis; and (**B**) norma lateralis to illustrate the individual bones and important anatomical landmarks.

- *Trochlear fossa (fovea):* lies in the anteromedial aspect of the roof, 4 mm from the margin, and is the site at which the trochlea (small pulley) is attached. The tendon of the superior oblique passes through the trochlea.
- *Anterior and posterior ethmoidal canals:* positioned at the junction of roof and medial wall above the frontoethmoidal suture (Fig. 1.5A). They transmit the anterior and posterior ethmoidal nerves and vessels.

Relations. The roof, which is thin and translucent except at the lesser wing of the sphenoid, separates the orbit from the anterior cranial fossa and frontal lobes of the brain. Anteriorly, the frontal sinus lies above the orbit.

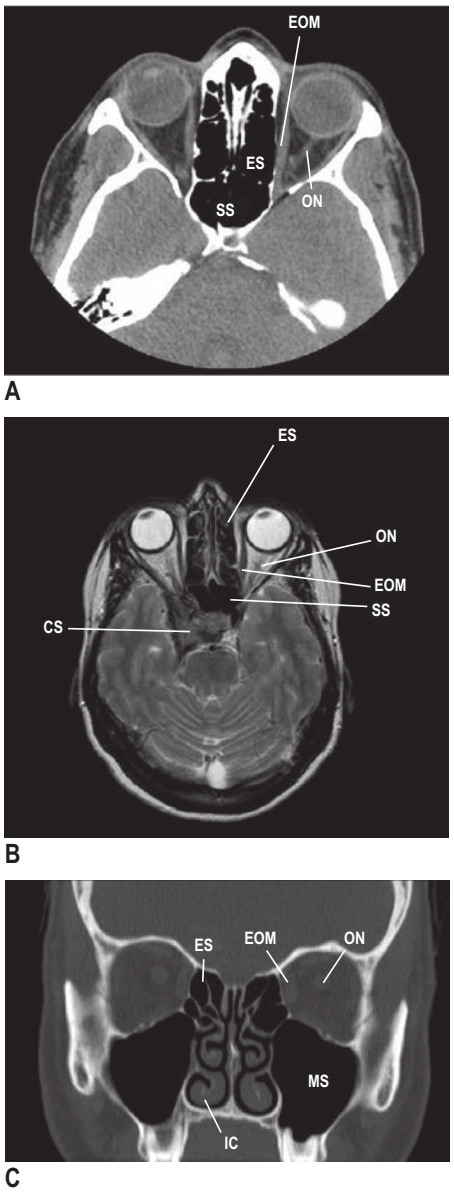


Fig. 1.3 Transverse (**A** and **B**) and coronal (**C**) computed tomography (CT: **A** and **C**) and magnetic resonance imaging (MRI: **B**) scans of the head displaying the major relations of the orbits. Features identifiable in the scans include ethmoid air cells/sinuses *(ES)*, maxillary sinus *(MS)*, sphenoid sinus *(SS)*, nasal cavity *(NC)*, inferior nasal concha *(IC)*, extraocular muscle *(EOM)*, optic nerve *(ON)* and cavernous sinus *(CS)*.

Features of the medial orbital wall.

- This wall is oblong and thin (0.2–0.4 mm). The four bones that comprise this wall are separated by vertical sutures (Figs 1.4 and 1.5A).
- *Lacrimal fossa* for the lacrimal sac: it is bound by anterior and posterior lacrimal crests and is continuous below with the *nasolacrimal canal* (Fig. 1.5B).

Relations. This is the thinnest of the walls and is largely transparent or semi-transparent: the ethmoidal air sinuses can easily be seen through this wall in a dried skull (Fig. 1.5A,B). Medial to this wall in an anterior to posterior sequence lie the anterior, middle, posterior ethmoidal air cells and the sphenoidal sinus. Infection can readily spread through the thin walls of the air cells causing orbital cellulitis (see Box 1.1 Clinical correlates).

Features of the orbital floor.

- The floor slopes slightly downwards from the medial to the lateral wall.
- It is crossed by the *infraorbital groove*, which runs forward from the *inferior orbital fissure*. Before it reaches the orbital margin this fissure becomes the *infraorbital canal*, which opens as the *infraorbital foramen* 4 mm below the orbital margin on the anterior surface of the maxilla (Figs 1.3C, 1.4 and 1.5A).

Relations. The floor separates the orbit from the maxillary sinus, the bone being only 0.5–1 mm in thickness (Fig. 1.3C). The floor is the most likely wall of the orbit to be damaged following trauma such as in orbital blow-out fractures (Box 1.2 Clinical correlates).

Features of the lateral orbital wall (Fig. 1.4).

- *Spina recti lateralis:* a small bony spine on the greater wing of the sphenoid near the apex of the orbit which gives origin to part of the lateral rectus.
- *Zygomatic foramen:* transmits zygomatic nerve and vessels to temporal fossa and cheek (zygomaticotemporal nerve and zygomaticofacial nerve) (Fig. 1.5B).
- *Lateral orbital tubercle:* forms the attachment of the check ligament of the lateral rectus, suspensory ligament (Lockwood's) of the eye, superior transverse ligament (Whitnall's) and aponeurosis of levator palpebrae superioris.
- Foramina for small veins that communicate with middle cranial fossa.

Relations. Laterally: skin, temporal fossa and middle cranial fossa in an anterior–posterior sequence (Fig. 1.3A).

Orbital margin, fissures and optic canal

Orbital margin. This is a thickened rim of bone that helps protect the orbital contents. It is made up of three bones: the frontal, zygomatic and maxilla (Figs 1.4 and 1.5A,B). The lateral margin does not reach as far anteriorly as the medial margin (see Figs 1.2B and 1.5B). The medial margin is sharp and distinct in its lower half because of the anterior lacrimal crest but is indistinct superiorly (Figs 1.4 and 1.5B).

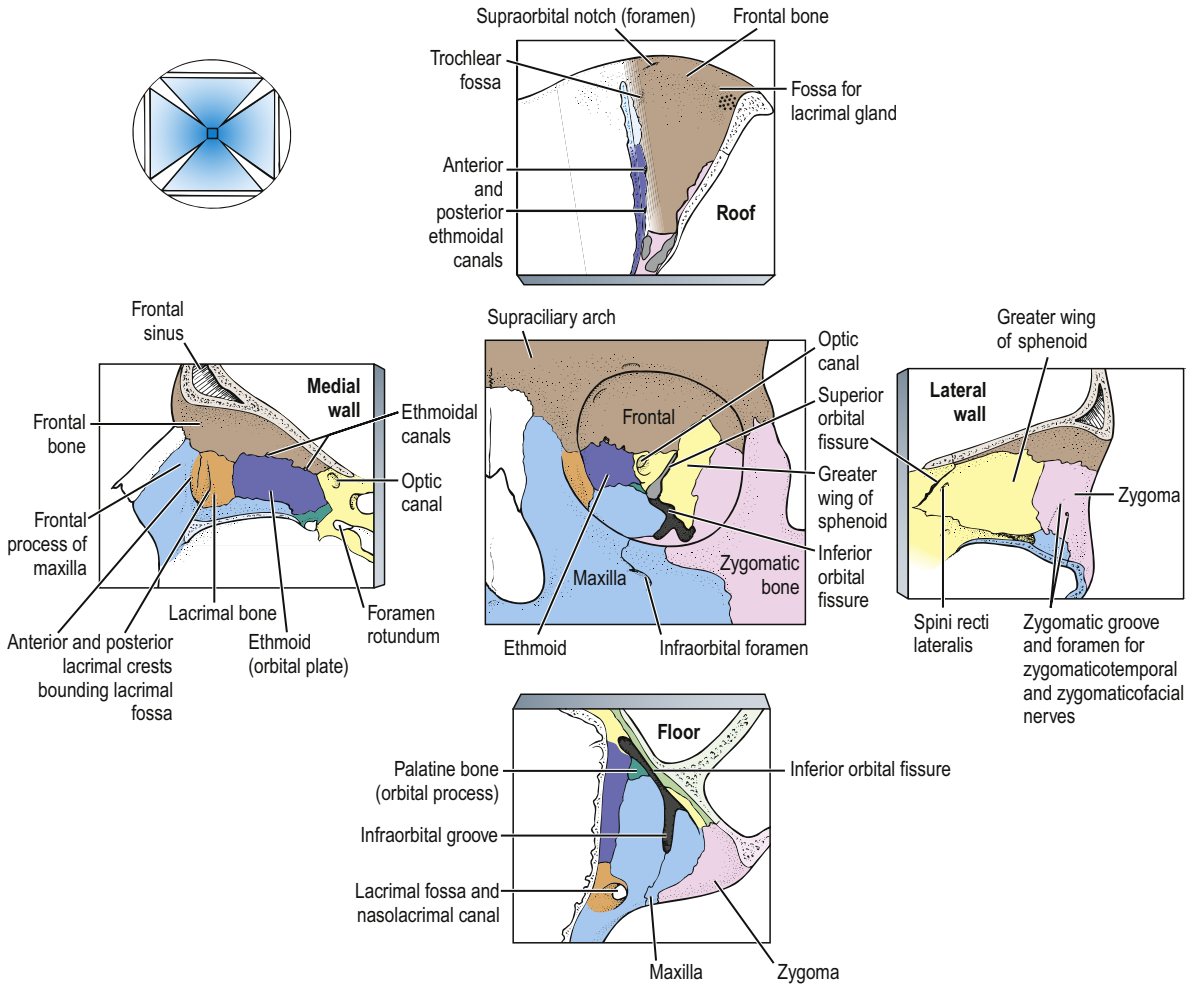
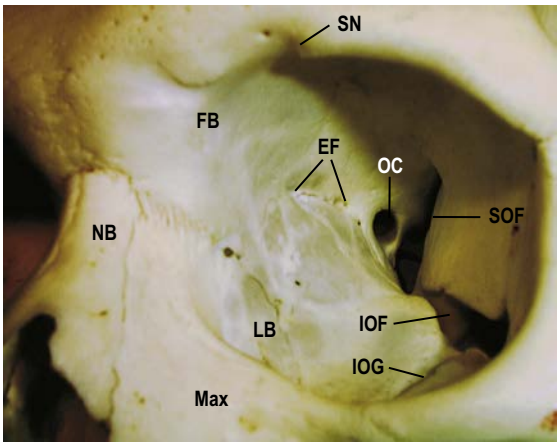
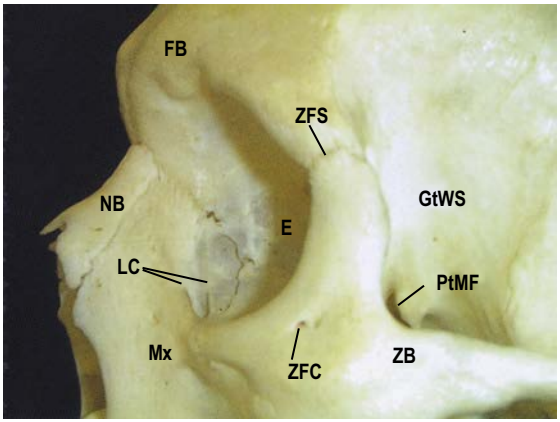


Fig. 1.4 Osteology of the orbit. The central diagram illustrates the anterior view of the intact left orbit, the four surrounding diagrams (exploded orbit: see inset, *top left*) show the individual bones that form the roof, floor, medial and lateral walls and other noteworthy features. *Roof:* orbital plate of the frontal bone and small area of lesser wing of the sphenoid. *Medial wall:* frontal process of the maxilla, lacrimal bone, orbital plate of the ethmoid and the body of the sphenoid. *Floor:* orbital plate of the maxilla, orbital surface of the zygoma and the orbital process of the palatine bone. *Lateral wall:* orbital surfaces of greater wing of sphenoid posteriorly and zygomatic bone anteriorly.

Superior orbital fissure. This communication between the orbital and cranial cavities lies between the roof and lateral wall of the orbit and is bounded by the lesser and greater wings of sphenoid (Figs 1.4, 1.5A and 1.6). It is wider at its medial end and narrowest at its lateral end. It is around 22 mm long and is separated from the optic foramen above by the posterior root of the lesser wing of the sphenoid. The part of the *common tendinous ring* that gives origin to the lateral rectus spans between the narrow and wide parts of the fissure. Structures passing above or outside the tendinous ring or annulus include the lacrimal nerve, frontal nerve, trochlear nerve, superior ophthalmic vein and recurrent branch of the lacrimal artery. The latter anastomoses with the orbital branch of the middle meningeal artery and may more commonly travel in a small cranio-orbital foramen lateral to the superior orbital fissure. Structures passing within the ring, and thus within the apex of the muscle cone, include the oculomotor nerve (superior and inferior divisions), abducent nerve, nasociliary nerve, sympathetic root of the ciliary ganglion and variably the inferior ophthalmic vein (Fig. 1.6).



A



B

Fig. 1.5 (**A**) Anterior view of the bony orbit showing important osteological features of the apex including the relation of the superior orbital fissure *(SOF)*, optic canal *(OC)* and inferior orbital fissure *(IOF)*. (**B**) Lateral view of the orbit. *E*, Ethmoid bone; *EF*, anterior and posterior ethmoidal foramina; *FB*, frontal bone; *GtWS*, greater wing of the sphenoid; *IOG*, inferior orbital groove; *LB*, lacrimal bone; *LC*, lacrimal crest; *Max*, maxilla; *NB*, nasal bone; *PtMF*, pterygomaxillary fissure; *SN*, supraorbital notch; *ZB*, zygomatic bone; *ZFC*, zygomaticofacial canal; *ZFS*, zygomaticofrontal suture.

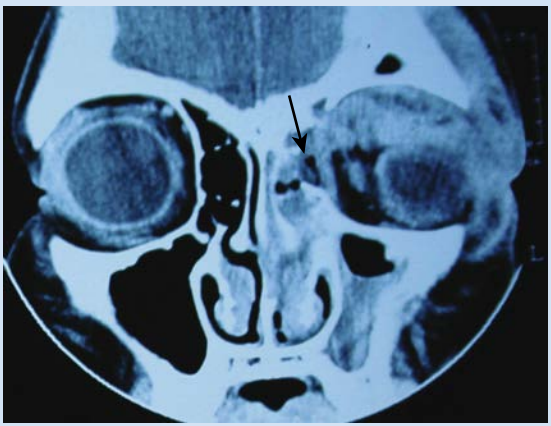
Box 1.1 **Clinical correlates**

Orbital cellulitis

This condition may be a consequence of infection spreading from the air sinuses to the orbit via the paper-thin medial wall (*lamina papyracea*) that separates the two. An example is shown of a patient with orbital cellulitis following creation of a drainage fistula (**A**). A coronal CT (**B**) illustrates the communication with the ethmoidal sinuses and nasal cavity (*arrow*).



A



B

Images courtesy Dr Alan McNab.

Inferior orbital fissure. This fissure lies between the lateral wall and floor of the orbit below the superior orbital fissure. It forms a communication between the orbit and the infratemporal fossa and pterygopalatine fossa. It runs forward and laterally for approximately 20 mm and ends 20 mm from the orbital margin (Figs 1.4 and 1.5A). The fissure is narrowest in the middle section and in life is covered by periorbita and a sheet of smooth muscle of unknown function, the orbitalis or 'muscle of Müller'. It transmits the infraorbital nerve, the zygomatic nerve and branches from the pterygopalatine ganglion and the inferior ophthalmic vein may communicate with the pterygoid venous plexus below.

Optic canal. This is a bony channel in the sphenoid that passes anteriorly, inferiorly and laterally (36 degrees) from the middle cranial fossa to the apex of the orbit. The

Box 1.2 **Clinical correlates**

Orbital blow-out fractures

The floor, although thicker than the medial wall, is more often involved in orbital blow-out fractures, probably because it lacks the buttress-like supports of the ethmoidal air cells and the protection of the nose. Tumour spread to or from the maxillary sinus may occur via the floor of the orbit.



A



B

An example of ocular motility being compromised in left eye following blow-out fracture (**A**) and a coronal CT (**B**) showing the herniation of orbital contents into the roof of the maxillary sinus (*arrow*).

Images courtesy Dr Alan McNab.

canal is formed by the two roots of the lesser wing of the sphenoid. The optic canals are 25 mm apart posteriorly and 30 mm anteriorly. Each is funnel-shaped and narrowest anteriorly where its opening into the orbit is oval with sharp upper and lower borders and a prolonged roof (10–12 mm in length). The opening at the cranial aspect is oval with a prolonged floor. The sphenoidal and posterior ethmoidal air sinuses are important medial relations and the olfactory tracts are superior relations of the canal. The canal transmits the *optic nerve* with its meningeal coverings and the *ophthalmic artery*, which lies below and then lateral to the nerve within the dural sheath for part of its course (Fig. 1.6). Sympathetic nerve fibres, originating from the superior cervical ganglion that form a plexus around the internal carotid artery, accompany the ophthalmic artery.

Paranasal sinuses

The paranasal sinuses (Figs 1.3 and 1.7 and Video 1.4) comprise the frontal, ethmoidal, sphenoidal and maxillary sinuses. They are air-filled cavities in the skull that are in communication with the nasal cavity via a series of apertures. Infection commonly spreads from the nasal cavity into the sinuses. The sinuses function to warm and moisten the air, add resonance to the voice and lighten the skull. They vary in size and shape between individuals and are commonly involved in facial fractures.

Additional content on the paranasal sinuses and orbital floor fractures is available here https//expertconsult.inkling.com

Cranial cavity (Fig. 1.8A,B and eFig. 1.1)

The cranial cavity houses the brain, cerebral vessels, the meninges, meningeal vessels and the intracranial portions of the cranial nerves. The base of the cranial cavity can be subdivided for descriptive purposes into three fossae: anterior, middle and posterior. Accounts of the detailed anatomy of these fossae can be found in standard anatomy texts; therefore, only features of relevance to the eye and orbit in the anterior and middle cranial fossae will be described.

Cranial fossae

Anterior cranial fossa. The anterior cranial fossa is limited in front and laterally by the frontal bone and posteriorly by the lesser wing of the sphenoid. Its floor is formed by the orbital plate of the frontal bone, the cribriform plate of the ethmoid (with a median crest-like ridge, the crista galli, which forms the anterior attachment of the falx cerebri) and the lesser wings and anterior part of the body (jugum) of the sphenoid. The perforations of the cribriform plate transmit the olfactory nerves. The orbital plate of the frontal bone separates the orbit below from the frontal lobes of the cerebral hemispheres, whose sulci and gyri cause surface impressions on the bone. Projecting posteriorly from the lesser wings of the sphenoid are the anterior clinoid

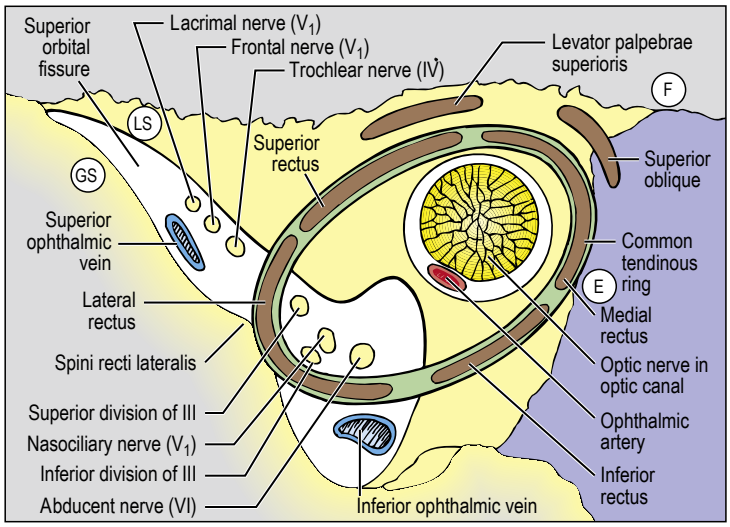


Fig. 1.6 Schematic diagram of the superior orbital fissure and optic canal in the right orbit. Note the origins of the extraocular muscles from the common tendinous ring and the relative position of the cranial nerves and vessels as they enter or exit the orbit. The positions of the veins are variable. The first letters of each of the structures passing through the superior orbital fissure (LFTSNIA) form a well-known mnemonic. *E*, Ethmoid; *F*, frontal bone; *GS*, greater wing of sphenoid; *LS*, lesser wing of sphenoid.

processes that overhang the middle cranial fossa and give attachment to the free edge of the tentorium cerebelli.

Middle cranial fossa. The middle cranial fossa lies at a lower plane than the anterior cranial fossa but is higher than the posterior cranial fossa. Its floor is shaped like a butterfly, with a narrow central or median part and expanded lateral parts. It is bound anteriorly by the posterior free edge of the lesser wing of the sphenoid, the anterior clinoid processes and the anterior margin of the sulcus chiasmaticus (Fig. 1.8A,B). Posteriorly it extends to the superior borders of the petrous temporal bones and dorsum sellae of the sphenoid and laterally it is bound by the squamous part of the temporal bone, part of the parietal bones and the greater wings of the sphenoid. Features and foramina of the floor of the middle cranial fossa and the structures that they transmit are summarized in Table 1.1.

Additional discussion on the posterior cranial fossa is available online at https//expertconsult.inkling.com

Pituitary fossa. The *pituitary fossa* (hypophyseal fossa) is an indentation in the roof of the body of the sphenoid bone in the middle cranial fossa. It is bound anteriorly by the *tuberculum sellae*, in front of which lies the *sulcus chiasmatica*, and posteriorly by the *dorsum sellae*, a ridge of bone at either end of which lies the *posterior clinoid processes*. The pituitary fossa houses the pituitary gland or *hypophysis cerebri*. This is connected by a thin stalk—the pituitary stalk (or tuber cinereum)—to the brain. The fossa is roofed by a sheet of dura mater, the *diaphragma sella* (Fig. 1.9D,E), which is attached in front to the tuberculum and behind to the dorsum sellae. The pituitary stalk passes through a small opening in the roof. The right and left cavernous sinuses are important lateral relations (see later).

The meninges and dural venous sinuses (Fig. 1.9A–C)

The brain and spinal cord are surrounded by three layers of meninges: a tough *pachymeninx*, the dura mater and the *leptomeninges* consisting of the arachnoid mater and pia mater. Between the arachnoid and pia is the *subarachnoid* space filled with cerebrospinal fluid.

Dura mater. The *dura mater* is theoretically 'divided' into an endosteal layer (really the periosteum on the inner surface of the skull) and a meningeal layer; however, on the whole they are fused except where they separate to form *dural venous sinuses* and *dural folds* (Fig. 1.9A,C). The latter are connective tissue septae that extend into the cranial cavity and serve to subdivide it into compartments. In association with the cerebrospinal fluid, they aid in providing physical support and protection for the brain. The position and form of the dural folds are summarized diagrammatically in Fig. 1.9A. and the relationship with cerebral arteries and the venous system of the brain and the dural venous sinuses is illustrated in the image of a 3D print (Fig 1.9B).

The *dural venous sinuses* are valveless, highly specialized, firm-walled veins within the cranial cavity that drain venous blood from the brain and cranial bones (Fig. 1.9C). Note that the term sinus can be applied to the air-filled paranasal sinuses as described previously (p.7 and Fig 1.7). In common with other veins, the sinuses are lined by vascular endothelial cells; however, their walls contain no smooth muscle cells. The arrangement of the sinuses is summarized in Fig. 1.9A,B.

ALGrawany

Frontal sinuses (Figs 1.4 AND 1.7)

The frontal sinuses are paired and lie behind the superciliary arches within the frontal bone (eFig. 1.1). They are separated from each other or further subdivided by thin bony septa that are not necessarily in the midline. The sinuses may extend as far laterally as the zygomatic process of the frontal bone. Each is approximately triangular and extends highest above the medial end of the eyebrow (Fig. 1.7). Each sinus opens into the middle meatus of the nasal cavity, either through the ethmoidal infundibulum or directly via the frontonasal duct. The mucosal lining is supplied by the supraorbital nerves and vessels; hence, referred pain from frontal sinusitis is experienced along the course of the supraorbital nerve.

Recent geometric morphometric studies (elliptic Fourier analysis) of the outlines of frontal sinuses from large numbers of radiographic images have confirmed a long-held belief that each individual's frontal sinus is distinct and unique. This may have important applications for personal identification in the context of forensics.

Ethmoidal sinuses (air cells) (Figs 1.3A–C and 1.7)

These thin-walled sinuses are for the most part situated in the lateral mass of the ethmoid, although frontal, maxillary, lacrimal, sphenoidal and palatine bones contribute to the walls. They are variable in number and are grouped into anterior, middle and posterior. The general pattern of drainage of the sinuses is as follows: the anterior opens into the hiatus semilunaris, the middle on to the bulla ethmoidalis (both middle meatus) and the posterior into the superior meatus. They are related to the frontal sinus anteriorly, the sphenoidal sinus posteriorly, the nasal cavity medially and below and laterally the orbit (Fig. 1.7A).

Sphenoidal sinus (Figs 1.3A,B and 1.7)

This sinus lies within the body of the sphenoid bone and possesses an indented roof because of the pituitary fossa that lies above and houses the pituitary gland (Fig. 1.8). It may be divided by a variable midline septum. A transverse ridge in the lateral wall marks the position of the internal carotid artery (within the cavernous sinus). Other important relations of the sinus include the optic chiasma and nerves above, the nasal cavity below, the ethmoidal sinuses anteriorly and the paired cavernous sinuses laterally (Fig. 1.9C). The sphenoidal sinus drains into the superior meatus or sphenoethmoidal recess. Surgical access to the pituitary gland may be gained via the nasal cavity and sphenoidal sinus; hence, surgeons must be aware of the above-mentioned relations.

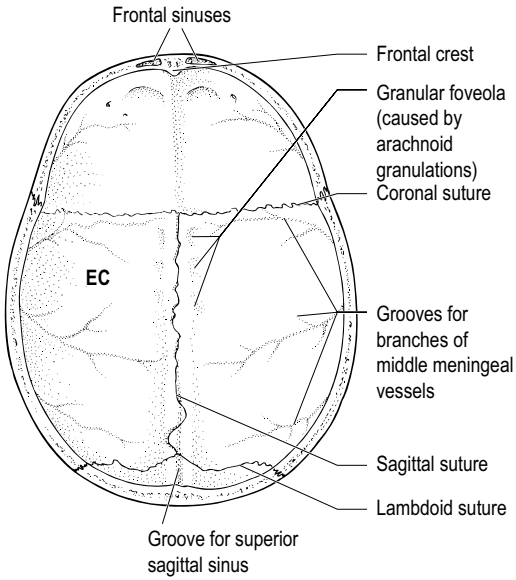
Maxillary sinus (Figs 1.3C and 1.7)

These are the largest of the paranasal air sinuses. They are pyramidal in shape and lie within the body of the maxilla. The *base* forms part of the lateral wall of the nasal cavity and the *apex* is within the zygomatic process. Each sinus is in communication with the middle meatus of the nasal cavity via an aperture, the maxillary hiatus, on its base, which empties into the lower part of the hiatus semilunaris. The opening is positioned high on this wall and therefore does not facilitate gravitational drainage in the upright position. The *nasolacrimal duct* lies in a thin bony canal in the anterior part of the base. The orbital plate forms the *roof* of the sinus and floor of the orbit.

Orbital floor fractures

Rapid traumatic compression of the orbital contents, such as occurs during squash ball injuries, can lead to blow-out fractures; orbital contents may herniate into the maxillary sinus. It was once thought that orbital contents, including extraocular muscles, became trapped in the fractured floor, thus restricting range of movement and explaining the diplopia suffered by these patients. However, recent studies have indicated that in many cases only orbital fibroadipose tissue is trapped in the damaged floor of the orbit.

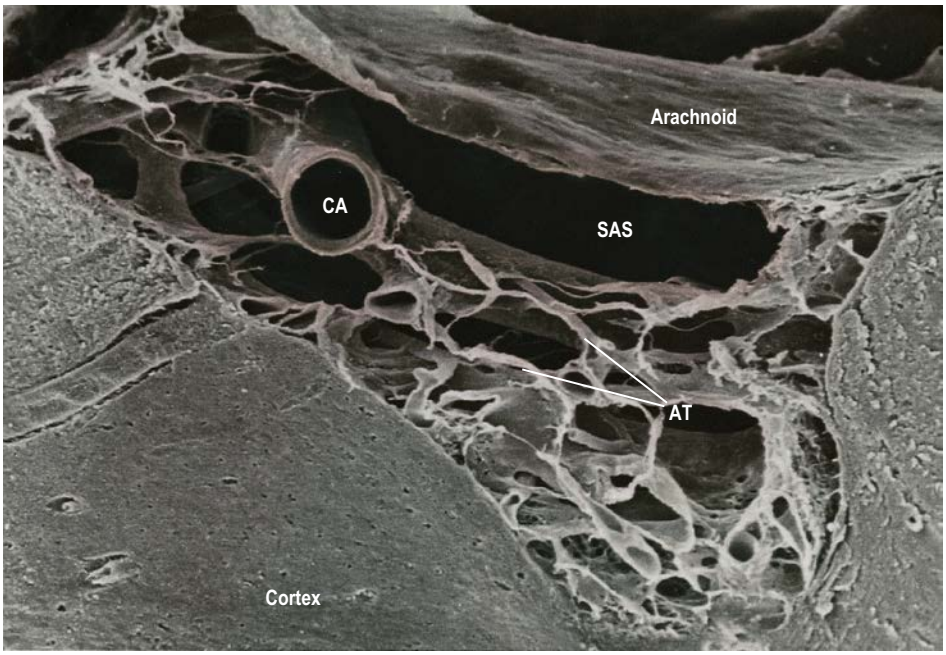
The *floor* of the maxillary sinus is formed by the alveolar process housing a variable number of the roots of the first and second molars that protrude into the sinus and may be separated from the sinus by only a thin covering of bone or mucous membrane. Thus, sinusitis may present as referred pain such as toothache and vice versa. In addition, abscesses in the maxillary sinus may result from infection of these roots. The *anterior/lateral wall* is directed on to the face and access for drainage of maxillary obstructions or other surgical procedures in the sinus may be gained by this route. The *posterior wall* faces the infratemporal fossa.



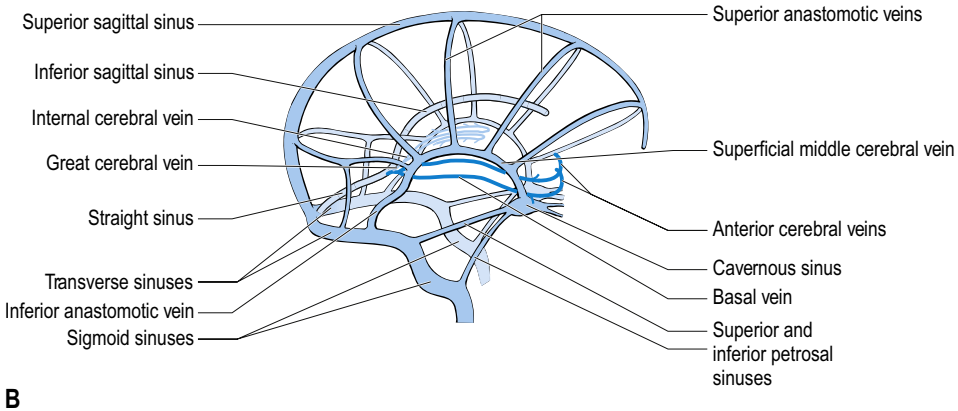
Internal aspect of the calvaria illustrating sutures and surface impressions made by meningeal arteries, dural venous sinuses and arachnoid granulations.

Posterior cranial fossa

This is the deepest of the three cranial fossae, its floor lying below the level of the middle fossa. Its roof is formed by the tentorium cerebelli. It lodges the hindbrain: the cerebellum, pons and medulla oblongata. The fossa is bound anteriorly by the superior border of the petrous temporal bone and the dorsum sella and surrounds the foramen magnum, the cerebellum being housed in the cerebellar fossae on the squamous part of the occipital bone. Features and openings on the floor of the posterior cranial fossa are not as relevant to the eye and orbit as those in the anterior or middle fossae; however, readers should be able to identify the following: foramen magnum, jugular foramen, hypoglossal canal, internal acoustic meatus, grooves for the sigmoid and transverse sinuses, internal occipital protuberance and clivus (Fig. 1.8A,B).



A



B

A. Scanning electron micrograph of the primate brain cortex and surrounding meninges. CA - cerebral artery; AT - arachnoid trabeculae. B, diagram of the arrangement of the dural venous sinuses.

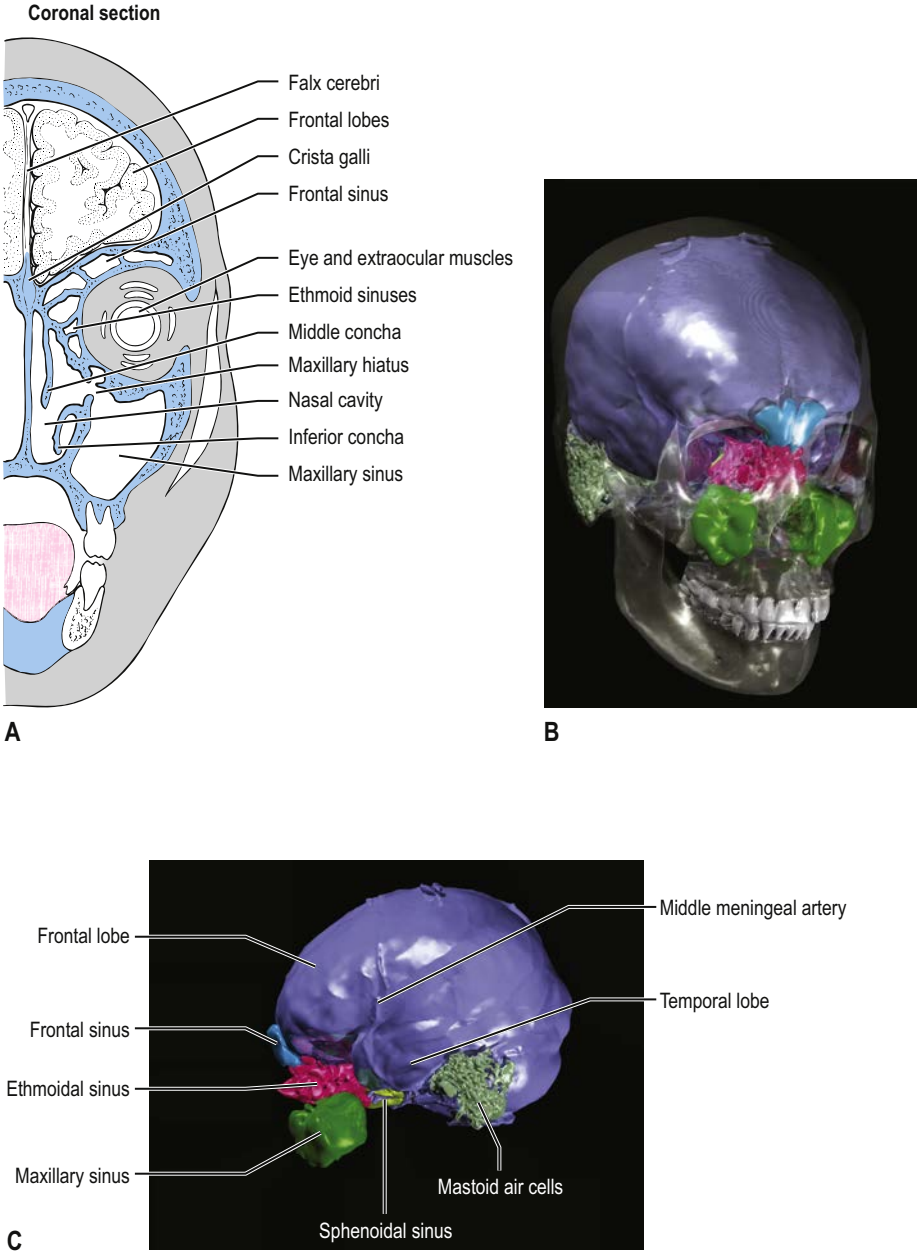


Fig. 1.7 (**A**) Diagram of a coronal section of the head revealing most of the paranasal sinuses except the sphenoid sinus. (**B**) 3D visualization of paranasal sinuses as seen from anterior aspect and (**C**) lateral aspect. See http://www.oucom.ohiou.edu/dbms-witmer/3D_human.htm or https://expertconsult.inkling.com/ for 3D pdf movies of skull and more details of paranasal sinus anatomy.

A L G r a w a n y

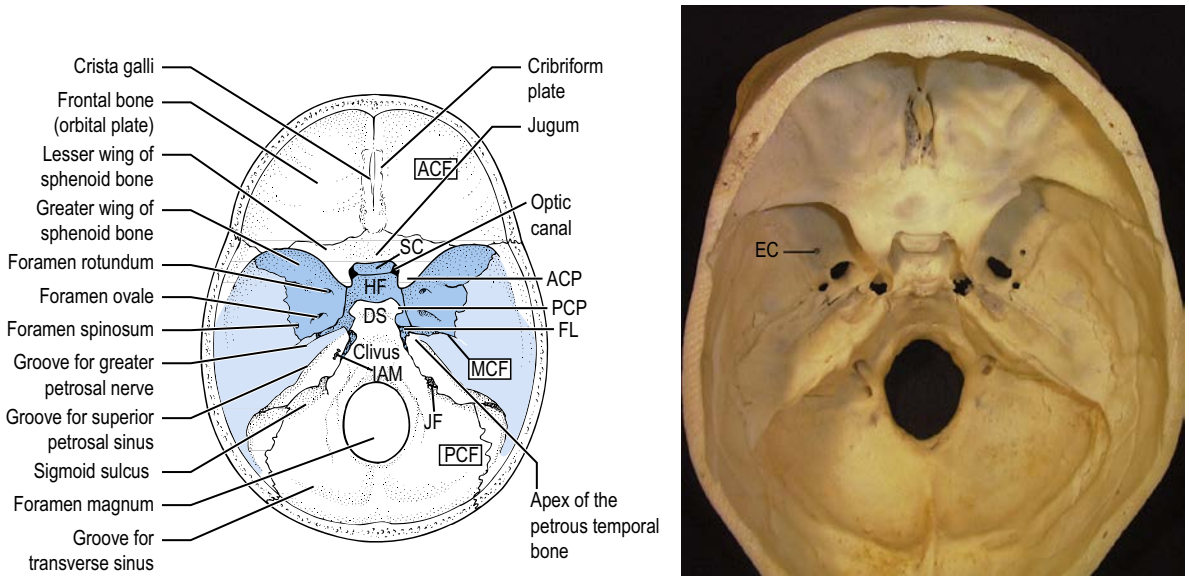


Fig. 1.8 Osteology of the cranial cavity. (**A**) The boundaries of the anterior *(ACF)*, middle *(MCF)* and posterior *(PCF)* cranial fossae together with major foraminae and important landmarks on the base of the skull. (**B**) Photograph of cranial cavity to illustrate features shown in (**A**). *ACP*, Anterior clinoid process; *DS*, dorsum sella; *FL*, foramen lacerum; *HF*, hypophyseal fossa; *IAM*, internal auditory meatus; *JF*, jugular foramen and fossa; *PCP*, posterior clinoid process; *SC*, sulcus chiasmaticus.

Of particular note to those studying the eye and orbit is the pair of cavernous sinuses lying either side of the body of the sphenoid (Fig. 1.D,E).

The importance of the *cavernous sinuses* (Fig. 1.9D,E) lies in their *position, relations* and extensive *communications.* Each cavernous sinus is around 2–3 cm long in the sagittal plane and consists of a series of incompletely fused venous channels or a single venous channel partially subdivided by *trabeculae.* It has walls of dura mater, like other venous sinuses.

Position. There is one cavernous sinus on either side of the body of the sphenoid. The sinus extends from the superior orbital fissure in front to the apex of the petrous temporal bone behind.

Relations. These are summarized in Fig. 1.9D,E (coronal section).

Communications. The sinuses communicate with each other via the anterior and posterior intercavernous sinuses. Tributaries draining into the sinuses anteriorly include the superior and inferior ophthalmic veins (which drain the eye and orbit as well as areas of skin around the periorbital region of the face and nose) and the sphenoparietal sinuses. The superficial cerebral vein from the brain drains into the sinus from above.

Blood from each sinus may, depending on relative pressures, drain via the superior and inferior petrosal sinuses either directly to the internal jugular veins (inferior petrosal) or to the transverse sinuses and thus to the internal jugular veins. Other exits include venous plexi around the internal carotid artery or veins traversing the foramen ovale or sphenoidal emissary foramen to communicate with the pterygoid plexus and other veins in the region of the skull base. Communications with the vertebral venous plexus in the epidural space also exist via the basilar venous plexus on the clivus. The importance of the close relationship of the internal carotid within the cavernous sinuses is illustrated by the clinical condition of arteriovenous fistulas and cavernous sinus thrombosis (Box 1.3 Clinical correlates).

The major dural folds (Fig. 1.9A) are falx cerebri, tentorium cerebelli, cavum trigeminale and diaphragma sella.

A description of the various dural folds is available online at https//expertconsult.inkling.com

The meningeal arteries lie within the inner (or periosteal) layer of dura with their accompanying veins (Fig. 1.9C) and are responsible for the many fine grooves that ramify over the inner surface of the cranium (see Fig. 1.8B). Here they are vulnerable to damage in head trauma, which can result in extradural and subdural haematomas (Box 1.4

The dural folds

Falx cerebri

A sickle-shaped fold with its attached border in the mid-sagittal plane from the crista galli to the tentorium cerebelli behind. It lies in the vertical fissure between the two cerebral hemispheres, its lower border lying above the corpus callosum. The superior sagittal sinus is situated in the attached border and the inferior sagittal sinus is in the lower free border of the falx cerebri.

Tentorium cerebelli. This fold lies approximately in a horizontal plane at 90 degrees to the falx, although it is elevated centrally (hence 'tent-like'). It separates the occipital lobe of each cerebral hemisphere above from the cerebellum in the posterior cranial fossa below. The free edge forms the boundary of the tentorial notch, which separates the forebrain from the hindbrain and 'houses' the midbrain. Sinuses related to the tentorium include the straight sinus, right and left transverse sinuses, superior petrosal sinuses and cavernous sinuses.

Cavum trigeminale. A blind-ended dural recess whose entrance is in the posterior cranial fossa. It is formed by an invagination of the dura beneath the free edge of the tentorium and is roofed by dura on the floor of the middle cranial fossa. It houses the trigeminal ganglion, which sits in a shallow hollow on the apex of the petrous temporal bone, and some accompanying vessels. The ganglion is surrounded by cerebrospinal fluid continuous with the subarachnoid space of the posterior cranial fossa.

Diaphragma sella. A small circular fold of dura over the sella turcica that is pierced centrally by the infundibulum (Fig. 1.9D). It blends laterally with the roof of the cavernous sinus (Fig. 1.9E).

The area above the tentorium is known as the supratentorial compartment; that below is the infratentorial compartment. The cranial dura of the supratentorial compartment is innervated by sensory branches of the trigeminal nerve and stimulation of these nerves (stretching, inflammation, compression) gives rise to frontal or parietal headache. The infratentorial compartment is supplied by branches of the upper cervical nerves and stimulation of these sensory nerves may therefore manifest as occipital and neck pain.

The neck rigidity accompanying acute meningitis of the infratentorial region is most likely the result of reflex contractions, or spasm, of posterior neck musculature in response to stretching of the inflamed cranial and spinal cord meninges.

ALGrawany

Table 1.1 **Summary of features on the floor of the middle cranial fossa**

Feature/foramen	Position	Relevance
Sulcus chiasmaticus	Between the two optic canals anterior to tuberculum sella	Optic chiasma rarely lies in contact with this region
Sella turcica (Turkish saddle)	Central part of sphenoid body between the two cavernous sinuses	The central hollow, the hypophyseal fossa, houses the pituitary gland. Anterior and posterior clinoid processes give attachment to the free and attached margins of the tentorium cerebelli
Optic canal	Between the two roots of the lesser wing of the sphenoid	Transmits optic nerve, ophthalmic artery, sympathetic nerves and meningeal coverings
Superior orbital fissure	Between the lesser and greater wings of the sphenoid. Lies at apex of cavernous sinus	Transmits trochlear, abducent and oculomotor nerves and terminal branches of ophthalmic nerve
Foramen rotundum	Pierces greater wing of sphenoid	Transmits maxillary nerve and small veins from cavernous sinus
Foramen ovale	Pierces greater wing of sphenoid	Transmits mandibular nerve, accessory meningeal artery and occasionally the lesser petrosal nerve
Foramen spinosum	Posterolateral to foramen ovale	Transmits middle meningeal artery and vein and meningeal branch of the mandibular nerve
Foramen lacerum	At apex of petrous temporal bone	The upper end transmits the internal carotid artery before it enters the cavernous sinus. Also transmits sympathetic nerves and a small plexus of veins. The lower end is covered by connective tissue and pierced only by small branches of the ascending pharyngeal artery
Trigeminal impression	Anterior surface of petrous temporal bone behind foramen lacerum	Occupied by trigeminal ganglion in trigeminal cave. Joined on lateral aspect by grooves for the greater and lesser petrosal nerves
Tegmen tympani and arcuate eminence	Tegmen is a thin plate of temporal bone over middle ear cavity. Arcuate eminence is produced by superior semicircular canal in petrous temporal bone	Infections in middle ear may spread through thin plate of bone to middle cranial fossa and temporal lobe of the brain

Clinical correlates). The dura also contains lymphatics that travel alongside the superior sagittal sinus and are especially numerous in the dura that lines the floor of the cranial cavity. It is also richly endowed with tissue resident immune cells including macrophages, dendritic cells and mast cells.

The largest and most important of the meningeal arteries is the middle meningeal artery, which enters the skull through the foramen spinosum. These arteries supply the meninges and diploë (bone marrow of cranial bones), but they *do not* supply the brain.

Arachnoid mater. The arachnoid (*Gk.* spider) is a delicate fibrocellular layer beneath the dura (separated by potential subdural space) that is connected to the pia mater covering the brain by numerous fibrocellular bands that cross the cerebrospinal fluid-filled subarachnoid space. This arrangement has led some to consider the leptomeninges as a conjoined pia–arachnoid membrane. The arachnoid bridges over the sulci, gyri and other irregularities on the brain surface, thus creating the *subarachnoid cisterns* or enlargements in the subarachnoid space (Fig. 1.9C, inset).

Specialized regions of arachnoid, the *arachnoid villi* and *granulations* (fibrous aggregations of villi), project into several of the dural venous sinuses (Fig. 1.9B) and act as one-way pressure-sensitive valves allowing cerebrospinal fluid to drain from the subarachnoid space into the dural venous sinuses and possibly also the lymphatics. Structures passing to and from the brain to the skull or its foramina, such as cranial nerves, must traverse the subarachnoid space. In addition, all cerebral arteries and veins lie in this space.

Since the arachnoid fuses with the perineurium of cranial nerves, the cerebrospinal fluid-containing subarachnoid space extends for a short distance around all cranial nerves. In particular it surrounds the optic nerve in a cuff-like manner as far as the posterior surface of the eye.

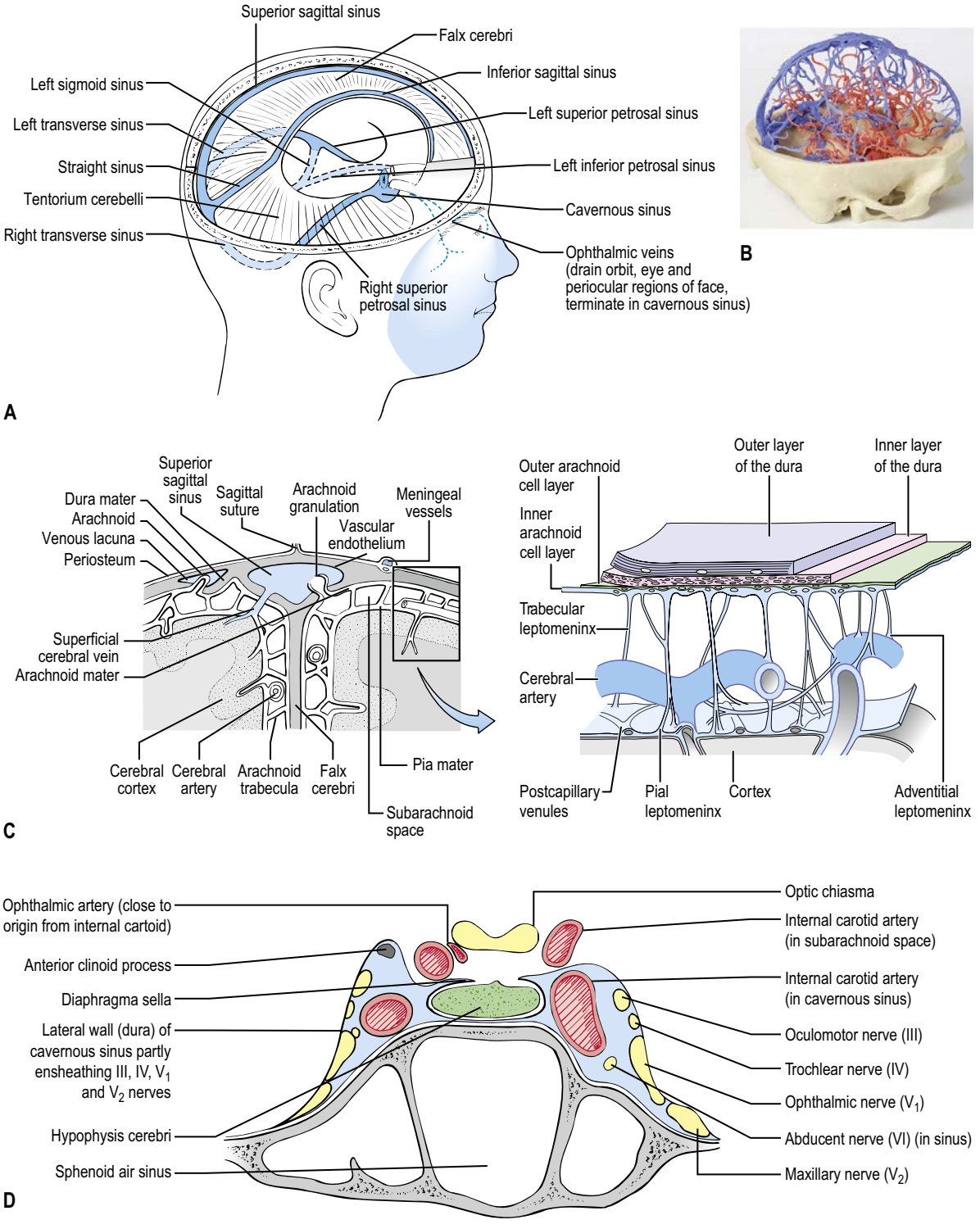
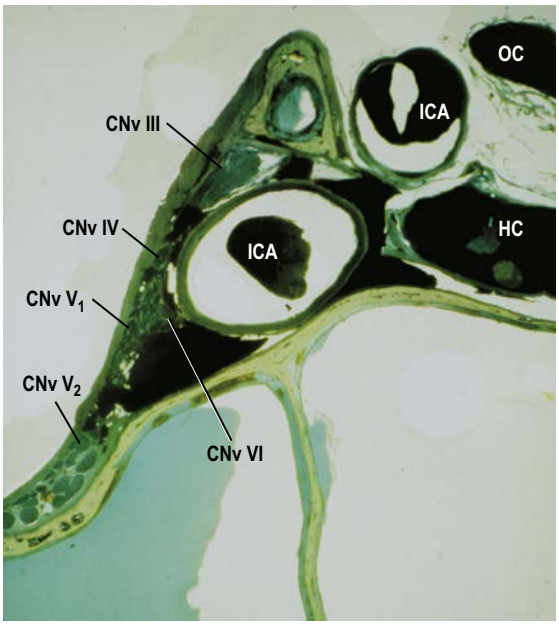


Fig. 1.9 (**A**) The interior of the cranium with the brain removed to illustrate the arrangement of the dural folds and some of the related dural venous sinuses. (**B**) A 3D print of the dural venous sinuses, cerebral veins and cerebral arteries in the cranial cavity. The vault of the skull has been digitally removed. The data are derived from CT angiography. (**C**) The meninges as seen in coronal section in the region of the superior sagittal sinus. *Inset:* higher-power diagrammatic representation of the meningeal layers. (**D**) Coronal section approximately midway along the body of the sphenoid bone to reveal the paired cavernous sinuses, one on either side. Note the position of the cranial nerves (II, III, IV, V_1, V_2 and VI), internal carotid artery (cut in two places, within and above the sinus) and hypophysis cerebri (pituitary gland



E

Fig. 1.9 Cont'd (**E**) High-power view of the left cavernous sinus (coronal plane, 100 μm thick section of low-viscosity nitrocellulose resin-embedded specimen) upon which (D) was based. *HC*, Hypophysis cerebri; *ICA*, internal carotid artery; *OC*, optic chiasma.

Pia mater (*Fig. 1.9C*). The pia mater, a vascular fibrocellular membrane that is thicker than the arachnoid, closely follows the contours of the brain. Vessels entering or leaving the brain parenchyma carry a pial sheath with them. Pial tissue is rich in astrocytes, which extend along the vessel walls and form an important component of the blood–brain barrier. The perivascular spaces surrounding these vessels, the so-called Virchow–Robin spaces, are potentially in communication with the cerebrospinal fluid of the subarachnoid space and may be dilated in pathological states.

Structure of the eye

The eye (Fig. 1.10) is a highly specialized organ of photoreception, the process by which light energy from the environment produces changes in specialized nerve cells in the retina, the rods and cones. These changes result in nerve action potentials, which are subsequently relayed to the optic nerve and then to the brain, where the information is processed and consciously appreciated as vision. All the other structures in the eye are secondary to this basic physiological process, although they may be part of the system necessary for focusing and transmitting the light on to the retina, for example cornea, lens, iris and ciliary body, or they may be necessary for nourishing and supporting the tissues of the eye, for example the choroid, aqueous outflow system and lacrimal apparatus.

Box 1.3 **Clinical correlates**

Arteriovenous fistulas in the cavernous sinus

These cause a variety of symptoms including pulsating protrusion of the globe and congestion of the vessels of the lids and conjunctiva owing to raised venous pressure. Patients complain of hearing noises resembling rushing water, probably because of increased flow rates in the labyrinthine plexus, which is in communication with the cavernous sinus via the superior petrosal sinus.

Cavernous sinus thrombosis

Cavernous sinus thrombosis as a sequel to infection spreading to the sinus, from such diverse initial sites as the nose, lids, behind the ear, bony labyrinth, pharynx and temporomandibular joint, can give rise to a variety of symptoms explainable on the basis of structures affected in and around the sinus. Facial pain may be the result of the involvement of the ophthalmic nerve (V_1). Lateral rectus paralysis may follow involvement of the abducent nerve. Involvement of the other oculomotor nerves is less common because they are more protected in the lateral wall of the sinus. Thrombosis is usually bilateral because of the communications via the intercavernous sinuses. Papilloedema may result from obstruction of central retinal venous return.

Box 1.4 **Clinical correlates extradural and subdural haematomas**

Damage to middle meningeal vessels, especially the frontal branch of the middle meningeal artery and vein (the latter lying closest to the bone), can result from blows to the head, especially in the temporal region (the pterion; see Figs 1.2B, 1.5B, 1.8B) where the bones are thinnest and most likely to fracture. Slow venous, or more rapid arterial, bleeding will lead to an extradural (epidural) haematoma with a resultant rise in intracranial pressure. Coma and death will occur if such a haematoma is not drained as soon as possible after symptoms of raised intracranial pressure, including papilloedema, manifest. A subdural haematoma can occur if the trauma results in brain laceration or tearing of intradural veins.

A L G r a w a n y

General shape, size and position of the eye

The eye is approximately a sphere 2.5 cm in diameter with a volume of 6.5 mL. However, in reality it comprises the parts of two spheres, a smaller one anteriorly, the cornea, that has a greater curvature than the sclera, which constitutes the large sphere. The cornea forms one-sixth of the circumference of the globe and has a radius of 7.8 mm; the remaining five-sixths is formed by the sclera, which has a radius of 11.5 mm. There is variation in size between individuals but the average axial length of the globe is 24 mm (range 21–26 mm). The diameter is 23 mm and the horizontal length approximately 23.5 mm. Small eyes (<20 mm) are hyperopic or hypermetropic, while large eyes (26–29 mm) are myopic. The eye is situated in the anterior portion of the orbit, closer to the lateral than the medial wall and nearer the roof than the floor. The eye is made up of three basic layers or coats, often known as tunics (Fig. 1.10). These are the fibrous (corneoscleral) coat, the uvea or uveal tract (composed of choroid, ciliary body and iris) and the neural layer (retina). The coats surround the contents, namely the lens and the transparent media (aqueous humour and vitreous body).

The cornea and sclera together form a tough fibrous envelope that protects the ocular tissues. The fibrous coat also provides important structural support for intraocular contents and for attachment of extraocular muscles. The cornea meets the sclera at a region known as the limbus or corneoscleral junction.

The cornea

The surface of the cornea (air–tissue interface) and associated tear film is responsible for most of the refraction of the eye. The transparency of the cornea is its most important property, although because of its highly exposed position it must also present a tough physical barrier to trauma and infection. Corneal transparency is the result of a number of related factors: its avascularity; the regularity and smoothness of the covering epithelium; and the regular arrangement and spacing of the extracellular and cellular components in the stroma, which is dependent on the state of hydration, metabolism and nutrition of the stromal elements.

Shape

The cornea is smaller in the vertical (10.6 mm) than in the horizontal (11.7 mm) diameter; however, viewed from behind, the circumference appears circular. The central radius is 7.8 mm with the peripheral corneal curvature being less marked. The cornea is also thicker at the periphery (0.67 mm) than in the centre (0.52 mm). The shape of

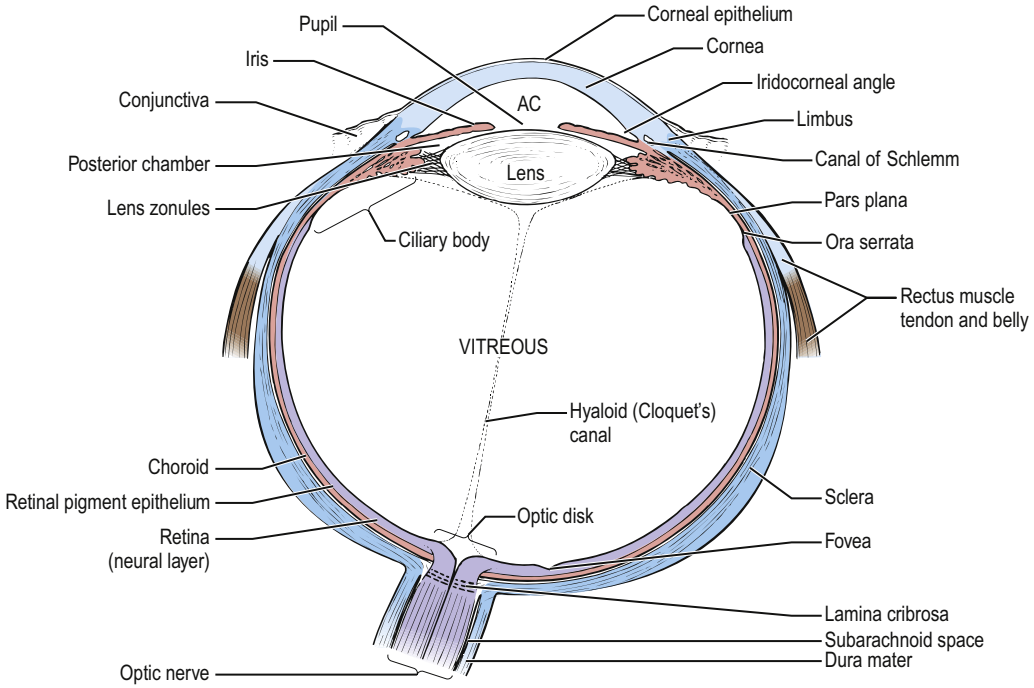


Fig. 1.10 Schematic diagram of the human eye in horizontal section revealing the major components and the arrangement of the three layers. The corneoscleral envelope (*blue*), the uveal tract (*orange/red*) and the inner neural layer (*purple*). *AC*, Anterior chamber

the cornea can be visualized by computerized video keratography as a series of colour-coded topography maps (Fig. 1.11) and is a frequently used method to visualize and to diagnose astigmatism (Box 1.5 Clinical correlates).

Structure

The cornea is composed of five layers (Fig. 1.12A).

Corneal epithelium (*Fig. 1.12B*). The corneal epithelium is a 50–60 μm thick layer of *stratified* (possessing five or six layers) *squamous non-keratinized* epithelial cells (the superficial cells are flattened, nucleated and non-keratinized). The barrier formed by the superficial cells is due to extensive tight junctions, whereas adjacent cells are held together by numerous *desmosomes* and to the underlying basal lamina by *hemidesmosomes* and anchoring filaments (Fig. 1.12B). The anterior surface of the corneal epithelium is characterized by numerous microvilli and microplicae (ridges) whose glycocalyx coat interacts with, and helps stabilize, the precorneal tear film. New cells are derived from mitotic activity in the limbal basal cell layer (see p. 22) and these displace existing cells both superficially and centripetally in an X, Y, Z migratory pattern. The corneal epithelium responds rapidly to repair disruptions in its integrity by amoeboid sliding movements of cells on the wound margin followed by cell replication.

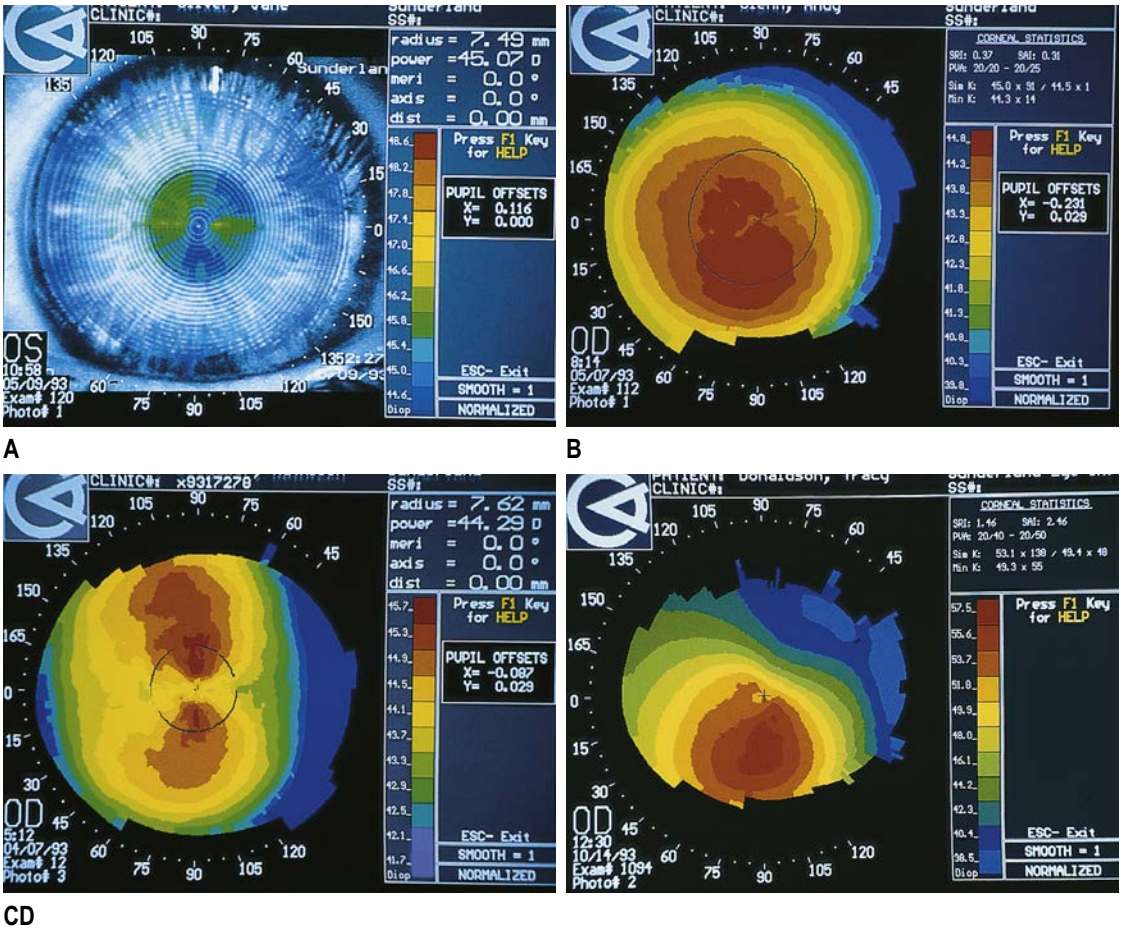


Fig. 1.11 Computerized video keratography (CVK). (**A**) This method involves the projection of over 6000 points of light on to the corneal surface in the form of placido rings. The images are analysed by the computer and complex colour-coded topographical/dioptric maps can be constructed. The scale or key is shown alongside: 'hotter' colours represent higher dioptric values. (**B**) A normal or round topographic map. (**C**) Regular 'with the rule' astigmatism in a normal healthy cornea with +1.5 dioptres of astigmatism at 90 degrees. (**D**) Corneal topographic map of a patient with early keratoconus. CVK analysis is particularly useful in identifying early keratoconus. In this case the higher dioptric values are concentrated in the infratemporal region. (Photographs kindly provided by Prof. C. McGhee.)

The basal epithelial cells rest on a thin, but prominent, *basal lamina* (lamina lucida, 25 nm; lamina densa, 50 nm). Corneal epithelial adhesion is maintained by a basement membrane complex, which anchors the epithelium to *Bowman's layer* via a complex mesh of anchoring fibrils (type VII collagen) and anchoring plaques (type VI collagen), which interact with the lamina densa and the collagen fibrils of Bowman's layer. The corneal epithelium is devoid of melanocytes. Myeloid-derived major histocompatibility complex (MHC) class II antigen-positive dendritic cells (sometimes referred to as Langerhans cells) are abundant in the limbal and peripheral corneal epithelium (Fig. 1.15) and are found in low numbers (approximately 20–25 cells/mm^2) in the central cornea. Some studies report MHC class II-negative dendritic cells in the mouse central corneal epithelium. However, these findings have been difficult to verify. Recent advances in *in vivo* confocal microscopy (IVCM) (Fig. 1.13) suggests that the normal human central corneal epithelium contains dendritic cells although their immunophenotype cannot be ascertained in the clinic. The comparative paucity of potential antigen-presenting cells, such as dendritic cells, and the avascular nature of the cornea are considered factors crucial to the success of corneal grafting (see Chapter 7).

Box 1.5 **Clinical correlates**

Astigmatism

Astigmatism is usually the result of differences in the radius of curvature in the vertical and horizontal meridians. Abnormalities in corneal curvature can be readily demonstrated by computerized video keratography, which applies the principle of projecting placido rings onto the corneal surface from which topographic maps can be constructed (Fig. 1.11).

Anterior limiting lamina (Bowman's layer). Bowman's layer (a modified acellular region of the stroma; 8–12 μm thick) consists of fine, randomly arranged, collagen fibrils (20–30 nm diameter, types I, III, V and VI). The anterior surface is well delineated and is separated from the epithelium by the thin basal lamina, while the posterior boundary merges with the stroma (Fig. 1.12A). Bowman's layer terminates abruptly at the limbus.

Substantia propria or corneal stroma (*Fig. 1.14A–C*). The corneal stroma is a dense connective tissue of remarkable regularity. It makes up the vast majority of the cornea and consists predominantly of 2 μm thick, flattened, *collagenous lamellae* (200–250 layers) oriented parallel to

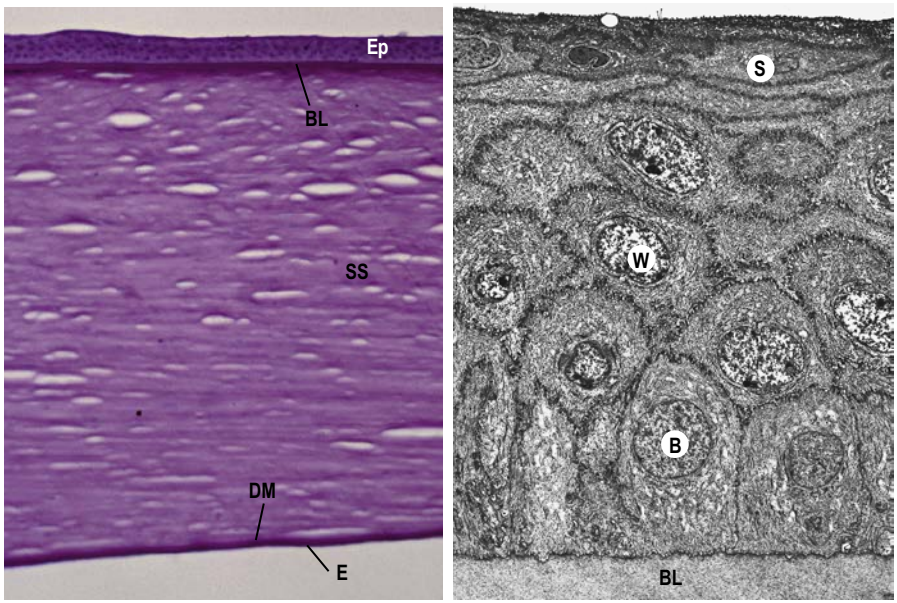


Fig. 1.12 Histology and ultrastructure of the cornea and its constituent layers. (**A**) Low-power light micrograph showing the five layers of the human cornea. (**B**) Electron micrograph of the corneal epithelium. *B*, Basal cell layer; *BL*, Bowman's layer; *DM*, Desçemet's membrane; *E*, endothelium; *Ep*, epithelium; *S*, superficial cells; *SS*, substantia propria or stroma; *W*, wing cells . Original magnifications: **A** × 80; **B** × 3000. (Fig. 1.12B courtesy W.R. Lee and D. Aitken.)

the corneal surface and continuous with the sclera at the limbus. Between the lamellae lie extremely flattened, modified fibroblasts known as *keratocytes*. These cells are stellate in shape with thin cytoplasmic extensions containing conspicuously few distinctive organelles (Fig. 1.14A–C) when viewed in conventional cross-sections. However, frontal sections reveal an abundance of organelles and a novel network of fenestrations on their surface which may facilitate the diffusion of metabolites or the mechanical 'anchoring' or attachment of collagen bundles (Fig. 1.14A). The density of keratoctyes in the anterior stroma is 20,000–24,000 cells/mm^2 and that density decreases posteriorly before

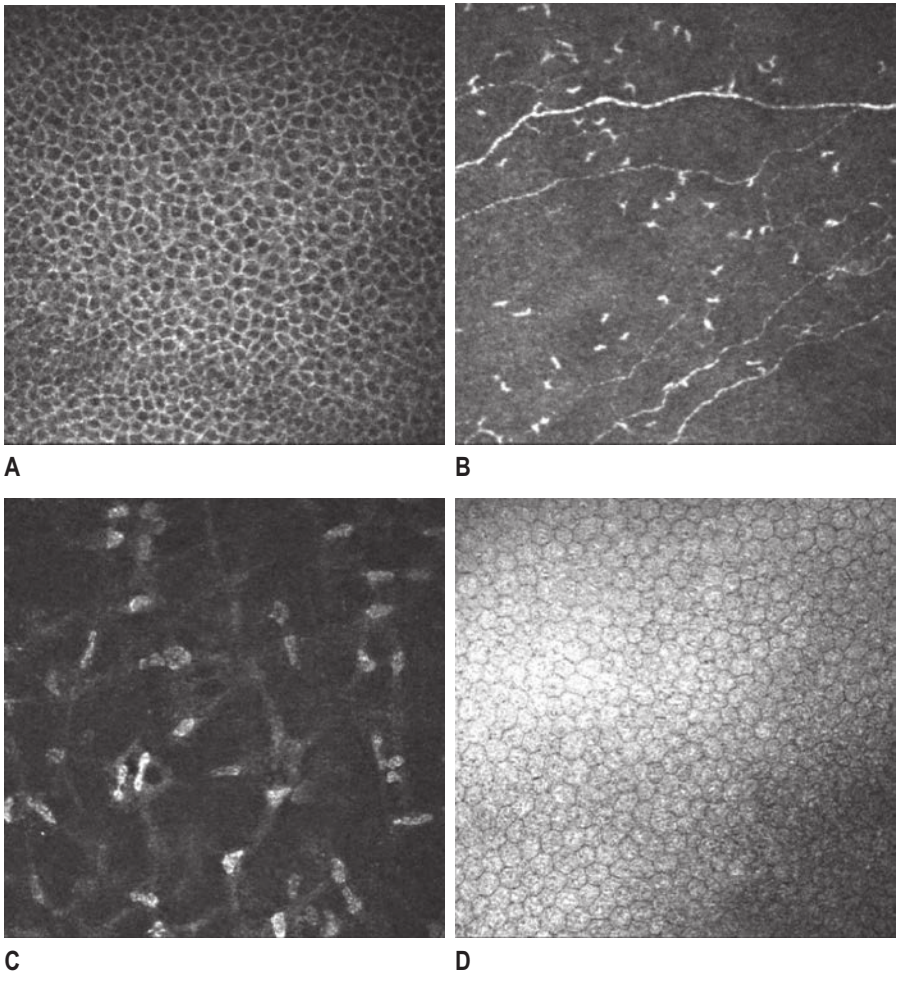


Fig. 1.13 *In vivo* confocal microscopy (IVCM) of the human cornea is a powerful non-invasive instrument used in the clinical evaluation of corneal abnormalities and normal structure of the tear film, cornea and conjunctiva. IVCM images are obtained by performing optical sections of the cornea using non-coherent white light. Cells and matrix components with differing reflective properties within the transparent cornea can be imaged. The advantage of a confocal approach is that only information in a narrow focal plane, approximately 4–25 μm in thickness, is analysed or collected by the microscope and scattering of light from structures outside the focal plane is thus minimized. The optics allow the light beam to be scanned (in the *x* and *y* axes) in a narrow area at one focal plane before shifting in depth to another plane of focus (*z* axis) where the scan is repeated. Thus, a series of optical 'slices' of high lateral resolution (1–2 μm) can be obtained from the entire cornea and, because of small differences in brightness/contrast, cellular detail can be visualized. This provides information that is normally the realm of conventional light microscopic and *ex vivo* laser scanning fluorescence confocal microscopic studies of processed tissues and whole mounts. The images are slices at differing depths in the cornea from superficial to deep: (**A**) epithelium; (**B**) sub-basal nerve plexus and dendriform cells, which may represent Langerhans cells; (**C**) keratocytes in the posterior stroma; (**D**) corneal endothelial cells. (Images courtesy Prof. C. McGhee.)

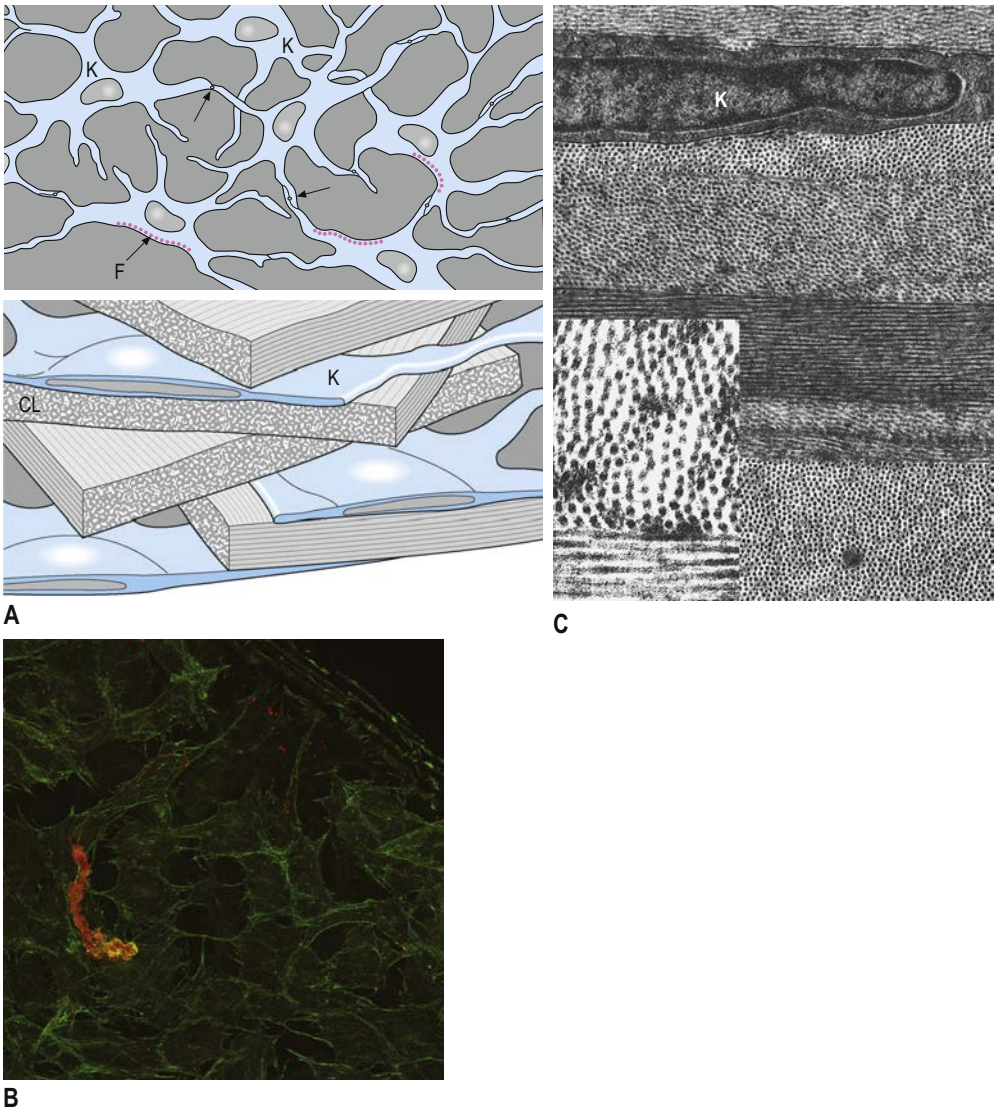


Fig. 1.14 (**A**) Schematic diagram showing the arrangement of the collagenous lamellae *(CL)* and the interposed keratocytes *(K)*. (**B**) Confocal microscopic view of keratocytes in the mouse cornea. *Green*: F-actin; *red*: MHC class II labelled myeloid cell. (**C**) Electron micrograph illustrating a keratocyte among regularly spaced collagenous lamellae. *Arrows*, Gap junctions; *F*, fenestrations. (*Inset*: higher power to show collagen fibres.) Original magnification: **C** × 20 000. (Fig.1.14C courtesy W.R. Lee and D. Aitken.)

increasing again near Desçemet's membrane (Fig. 1.13C). Keratocytes are connected by gap junctions to their neighbouring cells and arranged in a corkscrew pattern spiralling from the epithelium to the endothelium. The collagenous lamellae form a highly organized orthogonal ply, adjacent lamellae being oriented at right angles, with the exception of the anterior third in which the lamellae display a more oblique orientation. The collagen fibres (Fig. 1.14C, inset) are predominantly of type I (30 nm diameter, 64–70 nm banding) with some type III, V and VI also present. The transparency of the cornea is highly dependent on the regular diameter (influenced by the presence of type V collagen in particular) and spacing of the collagen fibres (interfibrillary distance), which in turn is regulated by negatively charged glycosaminoglycans (GAG) and proteoglycans forming bridges between the collagen fibrils.

ALGrawany

The GAGs in the human cornea are predominantly keratan sulphate and chondroitin (dermatan) sulphates (see Chapter 4 page 213). The corneal stroma normally contains no blood or lymphatic vessels, but sensory nerve fibres are present in the anterior layers en route to the epithelium (see later and Fig. 1.13). Studies using transgenic mice in which eGFP (enhanced green fluorescent protein) is expressed on all CX_3CR1 positive monocyte-derived cells have revealed extensive populations of resident tissue macrophages throughout the corneal stroma, some of which have recently been shown to communicate via novel cell–cell membrane nanotubes (Fig. 1.15F,H).

Posterior limiting lamina (Descemet's membrane) (Figs 1.12A and 1.16A). This is a thin, homogeneous, discrete, periodic acid–Schiff-positive layer between the posterior stroma and the endothelium, from which it can become detached. It is 8–12 μm in thickness and represents the modified basement membrane of the corneal endothelium. It consists of two parts, an anterior third that is banded and a homogeneous or non-banded posterior two-thirds. It is rich in basement membrane glycoproteins, laminin and type IV collagen. The anterior banded region is reported to contain type VIII collagen. Types V and VI collagen may be involved in maintaining adherence at the interface of Descemet's membrane with the most posterior lamellae of the stroma. Descemet's membrane is continuous peripherally with the cortical zone of the trabeculae in the trabecular meshwork. Microscopic wart-like protuberances (Hassall–Henle bodies) containing 'long banded' (100 nm) deposits of unknown nature appear in the periphery of Descemet's membrane with age. It is frequently thickened at its peripheral termination (Schwalbe line, the anterior limit of the trabecular meshwork). If disrupted, Descemet's membrane tends to curl inwards towards the anterior chamber.

Corneal endothelium. The corneal endothelium, a *simple squamous epithelium* on the posterior surface of the cornea, has a critical role in maintaining corneal hydration and thus transparency.

Fluid is constantly being lost via evaporation at the ocular surface, a fact illustrated by increased corneal thickness after a night of lid closure and when an impermeable contact lens is placed over the epithelium. The endothelial cells rest on Descemet's membrane (Fig. 1.16A) and form an uninterrupted polygonal or hexagonal array, or mosaic (Fig. 1.16B), which can be clearly seen *in vivo* with the aid of *specular microscopy* and *in vivo confocal microscopy* (Fig. 1.13). The cells are 5–6 μm in height and 18–20 μm in diameter. Their lateral surfaces are highly interdigitated and possess apical junctional complexes that, together with abundant cytoplasmic organelles including mitochondria (Fig. 1.16A), are indicative of their crucial role in active

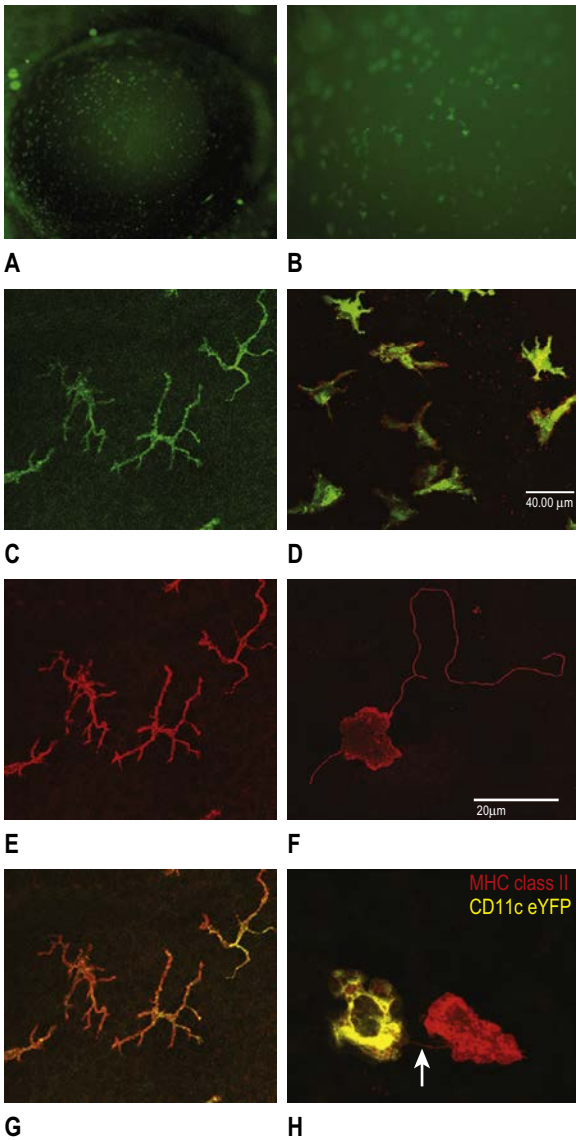


Fig. 1.15 (**A**, **B**) Low- and high-power *in vivo* fluorescent microscopy of the cornea of a normal CX3CR1-GFP transgenic mouse in which all myeloid cells (macrophages and dendritic cells) are labelled with *green* fluorescent protein (GFP). Note the regular array of corneal stromal macrophages. (**C–H**) Confocal microscopy of immune cells in the normal cornea. **C**, **E**, **G** show the same field of corneal epithelium from a corneal flatmount from a CX3CR1-GFP mouse. **C**: GFP *(green)*; **E**: MHC class II *(red)*; and **G**: combined image. (**D**) $CD11b^+$ *(red)* stromal macrophages in the same mouse cornea. (**F**) MHC class II^+ putative dendritic cell (DC) (see Chapter 7, page 425-429 for details of DC function) with a membrane nanotube extending from the cytoplasm. (**H**) Putative DC joined to an MHC class II^+ macrophage.

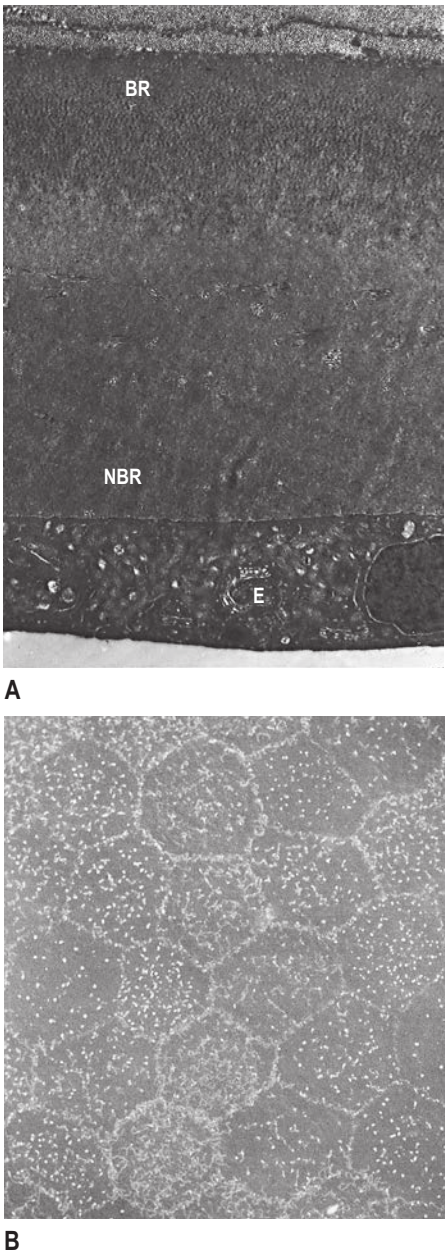


Fig. 1.16 (**A**) Electron micrograph of Desçemet's membrane and corneal endothelium *(E)* to illustrate the banded region *(BR)* and non-banded region *(NBR)*. (**B**) An *en face* view of the inner surface of the corneal endothelium as seen by scanning electron microscopy. Note the homogeneous hexagonal array of endothelial cells. A similar but less detailed view of the endothelium can be achieved in the living patient with the aid of specular microscopy. Original magnifications: **A** × 7000; **B** × 1200. (Courtesy W.R. Lee and D. Aitken; Fig. 1.16A courtesy Springer-Verlag.)

Box 1.6 **Ageing changes**

There are approximately 350,000 endothelial cells per cornea (3000–4000 cells/mm^2 at birth, falling to 2500 cells/mm^2 in middle age and 2000 cells/mm^2 in old age). Consequently, with age, the dense, regular hexagonal arrangement typical of the young cornea is replaced by fewer cells of more heterogeneous size and shape.

transport of bicarbonate and sodium, which controls transport of fluid from the stroma into the aqueous humour.

In the normal human cornea, endothelial cells are generally considered to have low regenerative capacity and lost cells are quickly replaced by spreading of adjacent cells. However, there is some evidence that putative endothelial stem cells are located in centripetally arranged 'niches' or columns in the extreme peripheral cornea. Damage to corneal endothelial cells and density below 800 cells/mm^2 leads rapidly to oedema and swelling of the stroma, with resultant loss of transparency (see Chapter 4, Box 4.10). The density of endothelial cells decreases with age (Box 1.6 Clinical correlates) and a density lower than 1500 cells/mm^2 in a potential donor cornea is considered unsuitable for transplantation.

Nerve supply of the cornea

The cornea is richly supplied by sensory fibres derived from the ophthalmic division of the *trigeminal nerve*, mainly via the *long ciliary* nerves. Occasionally the inferior cornea receives some branches from the maxillary division of the trigeminal. Nerve bundles enter the peripheral cornea in a radial manner and as they travel centrally within the anterior one-third of the stroma and approximately 1 mm from the limbus they lose their perineurium and myelin sheaths. This alteration in myelination is thought to be related to the importance for transparency. They divide into smaller branches and begin to change direction towards the epithelium where they must pierce Bowman's layer, whereupon they further divide into smaller bundles to form the *subepithelial* or *sub-basal plexus* in the interface between Bowman's layer and the basal aspect of the corneal epithelium. There are apparently no specialized end organs associated with these terminal axons, which are predominantly within the size range 0.1–0.5 μm consistent with A-delta and C fibres that function to transmit the sensory modalities of pain and temperature. Individual beaded fibres penetrate the epithelial layers and terminate in the superficial layers in the form of an intraepithelial plexus. There are approximately 7000 nociceptors per mm^2 in the human corneal epithelium

rendering it exquisitely sensitive. The sensory nerves in the corneal epithelium are of great interest to researchers as they can be readily viewed non-invasively using IVCM and act as the conduits for the spread of herpes simplex virus to the cornea (Box 1.7 Clinical correlates) and can indicate the presence of a range of ophthalmic and systemic diseases including dry eye disease and diabetes.

The sclera

The sclera (Fig. 1.17A–D) forms the principal part of the outer fibrous coat of the eye and functions both to protect the intraocular contents and to maintain the shape of the globe when distended by intrinsic intraocular pressure. The globe shape is maintained even during contraction of the extraocular muscles, whose tendons insert on its surface. However, in infants the corneoscleral envelope is not mature and can increase in size (Box 1.8). The sclera is relatively avascular and in adults appears white externally. The viscoelastic nature of the sclera (great tensile strength, extensibility and flexibility) allows only limited distension and contraction to accommodate minor variations in intraocular pressure.

Box 1.7 **Clinical correlates**

Larger myelinated nerve fibres can often be seen during slit-lamp examination as fine whitish fibres radiating into the cornea from the limbus. *In vivo confocal microscopy* has greatly added to our understanding of the distribution of corneal nerves in healthy and diseased corneas (see Fig. 1.13). As a result of their position in the anterior stroma and their radial arrangement, many are damaged during refractive procedures such as *LASIK* (Excimer laser *in situ* keratomileusis). Damage to the corneal epithelium and intraepithelial nerve terminals can cause a great deal of pain.

Reactivation of latent herpes simplex virus in the trigeminal ganglion occurs following damage to nerve terminals (cold, exposure to ultraviolet light, trauma, corticosteroids) and activated virus is transmitted to the cornea along sensory nerve branches, leading to recurrent herpes simplex keratitis and superficial corneal ulceration.

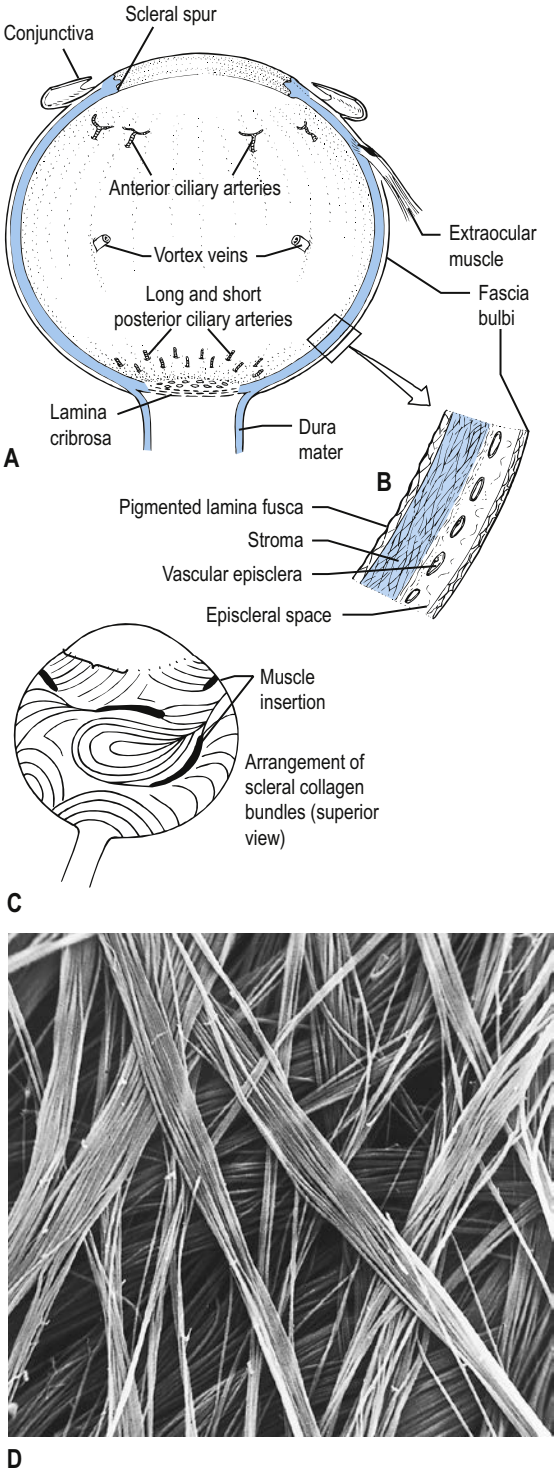


Fig. 1.17 (**A**) Schematic diagram of the isolated sclera and structures that blend with it (muscle tendons and optic nerve dura) or traverse its substance. (**B**) The scleral layers. (**C**) The pattern of orientation of the collagen bundles in the scleral stroma in relation to the extraocular muscle tendinous insertions. (**D**) Collagen bundles in sclera. Original magnification: **D** × 7000. (Fig. 1.17D courtesy Dr A. Thale, from Thale and Tillmann, 1993.)

Box 1.8 **Clinical correlates**

Buphthalmos

The corneoscleral envelope of children with congenital glaucoma responds to raised intraocular pressure by irreversibly stretching, owing to the immaturity of the collagen fibres, thus producing the characteristically enlarged buphthalmos ('ox-eye') of this condition.

Box 1.9 **Clinical correlates**

The yellowing of the eyeball in jaundice is the result of bilirubin deposition in the conjunctiva and not the sclera. Abnormal thinning of the sclera, such as occurs in some connective tissue disorders, e.g. Ehlers-Danlos syndrome, may also lead to a blue tinge. Localized thinning of the stromal collagenous layers may lead to staphyloma (bulging).

The sclera is thickest posteriorly (1 mm) and thinnest (0.3–0.4 mm) behind the insertions of the aponeurotic tendons of the extraocular muscles. It is covered by the *fascia bulbi* posteriorly and the conjunctiva anteriorly. The sclera consists of dense irregular connective tissue comprising extracellular matrix and matrix-secreting fibroblasts. The matrix consists principally of collagen type I, although types III, IV, V, VI, VIII, XII and XIII have been identified. Unlike the cornea, the scleral collagenous lamellae are irregularly arranged (Fig. 1.17D) and are interspersed with elastic fibres, each consisting of an elastin core surrounded by longitudinally arranged microfibrils composed of a number of glycoproteins including fibrillin. The opaque nature of the sclera, in contrast to the transparency of the cornea, can be partly ascribed to this irregular arrangement of the collagen fibres (Fig. 1.17D), but also to the variable fibre diameter (25–250 nm), variable and irregular fibrillar spacing, higher water content and the reduced coating of GAGs on collagen fibres. Indeed, the sclera contains one-quarter of the proteoglycan and GAG content of the cornea. Dermatan sulphate and chondroitin sulphate proteoglycans are the most abundant in the sclera.

Collagen fibrils take up tensile force and are aligned with the direction of greatest tensile strength. The arrangements of scleral collagen can be studied using the 'splitline' technique, which has revealed that the collagen fibrils in the outer sclera are arranged in bundles that course in whorls, loops and arches, particularly around the muscle insertions and optic nerve (Fig. 1.17C). The collagen fibrils on the internal aspect of the sclera are arranged in a rhombic pattern (Fig. 1.17D) (see also Chapter 4, page 221).

Box 1.10 **Ageing changes**

The blue sclera of infants is the result of the underlying choroidal pigment showing through the thin collagenous stroma. In elderly individuals, fat deposition in the sclera may produce a yellowish hue.

The sclera extends anteriorly from the *limbus* to the *lamina cribrosa* posteriorly (Figs 1.17A and 1.42). The scleral collagen fibrils are arranged in circles or figure-of-eight patterns at the lamina cribrosa. Structures that transverse the sclera are shown in Fig. 1.17A. Histologically the sclera has three layers: the lamina fusca, stroma and episclera (Fig. 1.17B). Colour changes in the white sclera and overlying conjunctiva can indicate systemic disease such as jaundice or age-related changes (Box 1.9 and Box 1.10 Clinical correlates).

Limbus and aqueous outflow pathways (Fig. 1.18)

It is becoming increasingly appreciated that the *limbus* (Fig. 1.18A–C) is more than the border zone between the cornea and sclera; it has multiple functions including nourishment of the peripheral cornea, corneal wound healing, immunosurveillance of the ocular surface and hypersensitivity responses; it contains the pathways of aqueous humour outflow and is thus involved in the control of intraocular pressure. Surgical incisions to access the anterior chamber for cataract and glaucoma surgery are made at the limbus. Indeed, surgeons define the limbus differently to conventional anatomical descriptions (Box 1.11 Clinical correlates). The limbus is 1.5–2.0 mm in width and the change in the radius of curvature between the sclera and cornea produces a shallow *external scleral sulcus* and an *internal scleral sulcus*; the latter is deepened by the *scleral spur* and houses the canal of Schlemm and trabecular meshwork. The longitudinal ciliary muscle fibres attach to the posterior aspect of the *scleral spur* and its anterior surface gives rise to the corneoscleral trabeculae.

Several important transitions take place at the limbus (Fig. 1.18A–C):

- The regularly arranged corneal lamellae give way to the more random array of lamellae in the sclera. The corneal termination is V-shaped (Fig. 1.18B–C).
- The stratified squamous non-keratinized corneal epithelium with its parallel internal and external surfaces gives way to conjunctival epithelium, characterized by a folded basal surface and interdigitating subepithelial connective tissue (sometimes forming distinct papillae) (Fig. 1.18B).

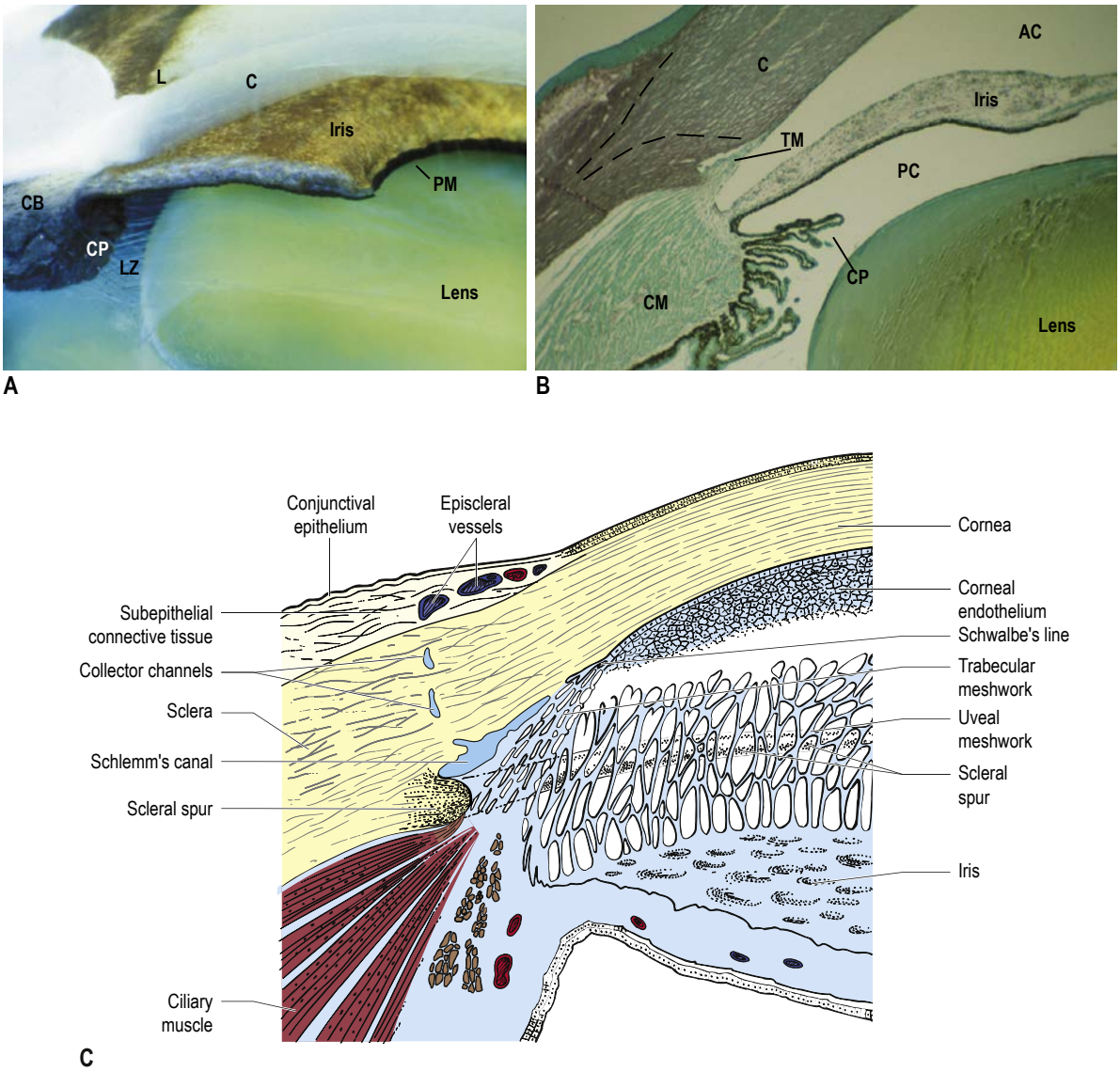


Fig. 1.18 (**A**) Macroscopic photograph of the anterior segment of a primate eye, which is almost identical to the human eye. Note the heavily pigmented limbus *(L)* characteristic of primates. (**B**) Histology section of the primate anterior segment (Van Gieson stain). *Dotted line* indicates the corneoscleral junction. (**C**) Three-dimensional schematic diagram of important features in and around the iridocorneal angle and corneoscleral limbus. *AC*, Anterior chamber; *C*, cornea; *CB*, ciliary body; *CM*, ciliary muscle; *CP*, ciliary processes; *LZ*, lens zonules; *PC*, posterior chamber; *PM*, pupillary margin; *TM*, trabecular meshwork. Original magnifications: **A** and **B** × 50.

- The conjunctival epithelium contains goblet cells and a rich network of MHC class II$^+$ CD11c$^+$ dendritic (Langerhans) cells (see section on conjunctiva, pp. 87-91).
- Loops or arcades of conjunctival capillaries (derived from the anterior ciliary arteries) and lymphatic capillaries terminate at the limbus. The smaller vessels are not under neuronal control and are particularly susceptible to the effects of vasoactive amines (e.g. histamine, leukotrienes, prostaglandins) released by local immune cells (see later).
- Desçemet's membrane and Bowman's layer of the cornea terminate in this region.

Box 1.11 Definitions of anatomical and surgical limbus

Definitions of the limits and markings of the limbus vary among anatomists, pathologists and surgeons. The anatomical (histological) limbus is defined by a line that follows the V-shaped transition of corneal lamellae to scleral lamellae (Fig. 1.18B). Pathologists define the limbus as a block of tissue bordered anteriorly by a line passing through the termination of Schwalbe's line and the junction of the conjunctival and corneal epithelium (corneolimbal junction) and posteriorly by a line from the scleral spur perpendicular to the tangent of the external surface. Surgeons usually cut close to the blue–grey transition zone seen on external examination and incisions made here will pass anterior to the trabecular meshwork and Schlemm's canal (Fig. 1.18B,C).

- The loose conjunctival subepithelial vascularized connective tissue (substantia propria), containing immunocompetent cell types such as mast cells, plasma cells and lymphocytes, tapers off at the limbus and is absent in the cornea (Fig. 1.18B).

A recently described 'layer' anterior to Descemet's membrane has been described as Dua's layer (see Chapter 4, page 217) which represents a condensation of the stromal collagen fibres and is identified by dissection of the cornea using a bubble of air during preparation of the cornea for corneal transplantation.

Aqueous outflow pathways

In the chamber or iridocorneal angle, partially nestled in the internal scleral sulcus, lies a complex wedge-shaped circumferential band of specialized, sponge-like, connective tissue, the *trabecular meshwork*, with the *canal of Schlemm* (sinus venosus sclerae) on its outer aspect (Figs 1.18B–C and 1.19A). The base of the trabecular meshwork is formed posteriorly by the scleral spur, the anterior face of the ciliary muscle and the iris root. The apex of the meshwork terminates anteriorly at Schwalbe's line and the adjacent innermost corneal lamellae (Figs 1.18 and 1.19A). The trabecular meshwork can be further subdivided into three anatomical zones: the innermost *uveal meshwork* with cord-like trabeculae; the *corneoscleral meshwork* with flattened sheet-like trabeculae (Fig. 1.19D); and the outermost *cribriform meshwork* beneath the inner wall of Schlemm's canal (Fig. 1.19E). The cribriform meshwork, unlike the rest of the trabecular meshwork, is not arranged in lamellae but consists of trabecular cells enmeshed in a loose extracellular matrix of collagen (types I, III and IV), elastic-like fibres and proteoglycans. This layer is thought to be the main site of resistance to aqueous outflow. The elastic cores of the trabeculae are continuous with the elastic fibres in the cribriform meshwork, which are in turn connected to the inner wall of Schlemm's canal via 'connecting fibres'. The anterior ciliary muscle fibres terminate in the elastic cores of the trabeculae. As the ciliary muscle contracts and moves inwards, the three-dimensional trabecular meshwork can be expanded, which results in an increase in the amount of 'free' spaces in the cribriform meshwork. This in turn allows greater aqueous outflow, thus increasing aqueous outflow facility.

Aqueous humour passes from the anterior chamber through the *intertrabecular* and *intratrabecular spaces*, which are lined by trabecular cells. These cells envelop the trabeculae (Fig. 1.19D) and maintain the state of hydration of the connective tissue core in a similar manner to corneal endothelium. In addition, trabecular cells are also phagocytic, trapping and removing debris from the aqueous humour as it percolates through the tortuous intertrabecular and intratrabecular spaces which narrow as Schlemm's canal is approached (Figs 1.18C and 1.19A).

Schlemm's canal (*sinus venosus sclerae*) is an endothelium-lined 36 mm long circumferential channel filled with aqueous humour. The canal measures 200–400 μm in the anteroposterior axis, is seldom more than 50–60 μm deep and is often septate. It is drained by 25–35 *collector channels* (20–90 μm in diameter) and between two and eight *aqueous veins* (of Ascher) (up to 100 μm in diameter). These either join deep, intrascleral and episcleral venous plexuses which drain into conjunctival veins or, in the case of aqueous veins, may drain directly into superficial conjunctival veins. The majority of aqueous humour (70%–90%) leaves the anterior chamber through the trabecular meshwork and Schlemm's canal (*'conventional' outflow pathways*). The inner wall of the canal is characterized, in well-preserved and properly fixed eyes, by transcellular channels or giant vacuoles (Fig. 1.19B,C). There is good evidence to suggest that these intracellular vacuoles, with openings on both the trabecular and luminal aspects, function to drain the great bulk of aqueous humour. The number and size of vacuoles and their openings or pores vary in a pressure-sensitive manner. Small quantities of aqueous may also pass between endothelial cells in the canal wall. Whilst it was thought that the endothelial lining of Schlemm's canal and the aqueous veins is functionally and phenotypically analogous to lymphatic endothelium, recent data dispute this and have shown that prospero homeobox protein 1 (Prox-1), the master control gene for lymphatic development, is expressed in Schlemm's canal but that the endothelial lining of Schlemm's canal is distinguishable from typical lymphatic vessels by lack of lymphatic vessel endothelial hyaluronan receptor (LYVE-1) expression but CD31+ and that supporting cells in its wall may have contractile properties.

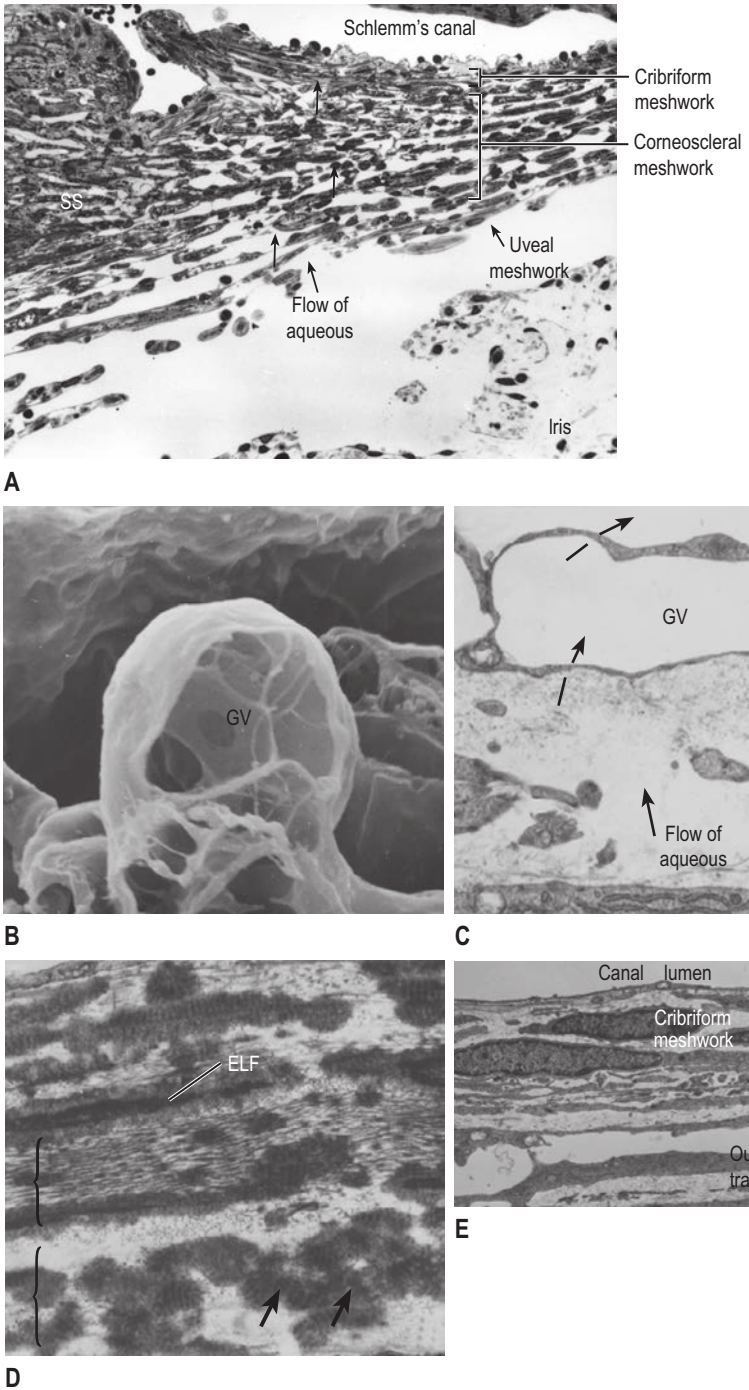


Fig. 1.19 Histology and ultrastructure of the trabecular meshwork and Schlemm's canal. (**A**) Light micrograph showing the scleral spur *(SS)*, Schlemm's canal and the three zones of the meshwork: the uveal, corneoscleral and cribriform meshworks. The path of aqueous through the inter- and intratrabecular spaces is indicated by *arrows*. (**B**) Giant vacuole *(GV)* in the inner wall of Schlemm' s canal as seen by scanning electron microscopy. (**C**) Transmission electron micrograph of a giant vacuole. The flow of aqueous is indicated by the *arrows*. This particular section does not include the basal or luminal pores seen in some giant vacuoles. (**D**) Electron micrograph of a trabecula cut in cross-section showing the layered arrangement of the extracellular components. *Arrows* indicate 'long spacing collagen'. (**E**) High-power micrograph of the cribriform meshwork or layer showing the lack of trabecular organization. Fibrocyte-like cells are loosely arranged in various types of extracellular matrix. *CL,* Cat; *CZ*, cortical zone; *ELF*, elastic-like fibres. Original magnifications: **A** × 340; **B** and **C** × 7000; **D** × 13 000; **E** × 5000. (From McMenamin, Lee and Aitken, 1986, with permission.)

Box 1.12 **Clinical correlates**

Glaucoma

Glaucoma is defined as a progressive optic nerve neuropathy. For most forms of glaucoma, elevated intraocular pressure and ageing remain important risk factors, although low tension or normal tension forms of glaucoma are common. However, in many forms of this condition pathological changes in the trabecular meshwork and Schlemm's canal may be responsible for increased resistance to aqueous outflow and raised, or diurnal fluctuations in, intraocular pressure (IOP). In *congenital glaucoma* there is malformation of the complex three-dimensional arrangement of the trabeculae and excess extracellular matrix in the outer meshwork. Physical blockage of the inner surface of the chamber angle by the iris occurs in *closed-angle glaucoma*; this may be a primary or a secondary process (see Chapter 9, p. 517). Various forms of obstruction in the trabecular meshwork may give rise to *open-angle glaucoma* (see Chapter 9, p. 517); the cause of the primary form of this condition is unknown, although there is evidence to indicate that excessive deposition of extracellular elements may occur in the cribriform meshwork. Secondary forms of open-angle glaucoma may be the result of debris such as lens proteins, melanin, macrophages and haemorrhagic products physically obstructing the intertrabecular and intratrabecular spaces, thus causing raised IOP (see Chapter 9, p. 517).

A proportion (10%–30%) of aqueous humour drains via the *'non-conventional' aqueous outflow pathways*. This route is not pressure sensitive and consists of the intercellular spaces between ciliary muscle fibres and the loose connective tissue of the suprachoroidal space. From here, aqueous traverses the sclera via the connective tissue sheaths of nerves and vessels that pierce its substance (see Fig. 1.17A). Recent evidence has also emerged that there is a greater exchange of aqueous with the iris stroma and thus the iris vasculature than previously known (see Chapter 4 p.223). Disturbances in the normal function of the outflow pathways in glaucoma are discussed in Box 1.12 Clinical correlates.

Uveal tract or uvea

The uveal tract (L. *uva* = grape), the middle vascular pigmented layer of the eye, consists of the *iris, ciliary body* and *choroid* (see Fig. 1.10). These three components are continuous with one another and have an opening anteriorly, the *pupil,* and posteriorly the choroid is deficient at the *optic nerve canal.* The uveal tract is analogous to the vascular pia-arachnoid of the brain and optic nerve, with which it anastomoses at the optic nerve head. The choroid is described on p. 53; the iris and ciliary body are described in the following sections.

The iris

The iris (Fig. 1.20A,B,) is a thin, heavily pigmented, contractile circular disk analogous to the diaphragm of a camera. It is suspended in the frontal or coronal plane anterior to the lens and ciliary body and is surrounded by aqueous humour. The iris separates the anterior and posterior chambers, which are in continuity through an opening, the *pupil,* which lies slightly inferonasal to the centre of the iris. The iris is attached by its *root* at the angle (iridocorneal) of the anterior chamber where it merges with the ciliary body and trabecular meshwork. The free edge is known as the *pupillary margin.*

The iris is 12 mm in diameter with a circumference of 37 mm. It is cone-shaped with the pupil margin positioned more anteriorly than the root. The pupil margin rests on the lens, without whose support, for example in aphakic patients, it becomes tremulous (iridonesis). The size of the pupil regulates the amount of light entering the eye and is dependent on the state of contraction of the intrinsic pupillary muscles, the *dilator* and *sphincter* pupillae. The pupil may vary from 1 to 8 mm in diameter and there may be a slight degree of asymmetry between right and left eyes in normal individuals.

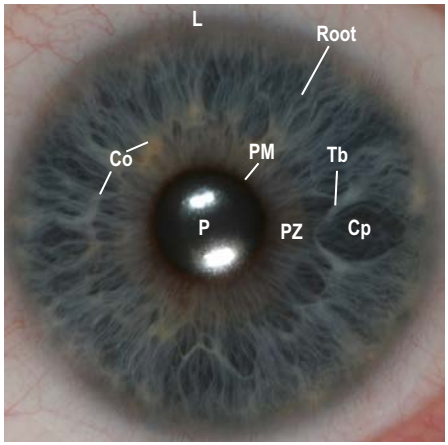
Structure (Fig. 1.20A,B)

The pupil margin and iris root are thin and hence more susceptible to tearing in contusion injuries (iridodialysis). The anterior surface is divided into two zones, the *ciliary zone* and *pupil zone,* by the thickened region known as the *collarette.* The anterior surface is characterized by *radial streaks* (straight when the pupil is contracted and wavy when dilated) and *contraction furrows* (more noticeable in dilated irides). The surface of the iris appears smooth in dark irides, in which the intrastromal melanocytes are heavily pigmented, and more irregular in blue irides, which have a less heavily pigmented stroma. The blue appearance of some irides is because of absorption of long wavelengths and reflectance of shorter wavelengths, especially by the collagenous stroma.

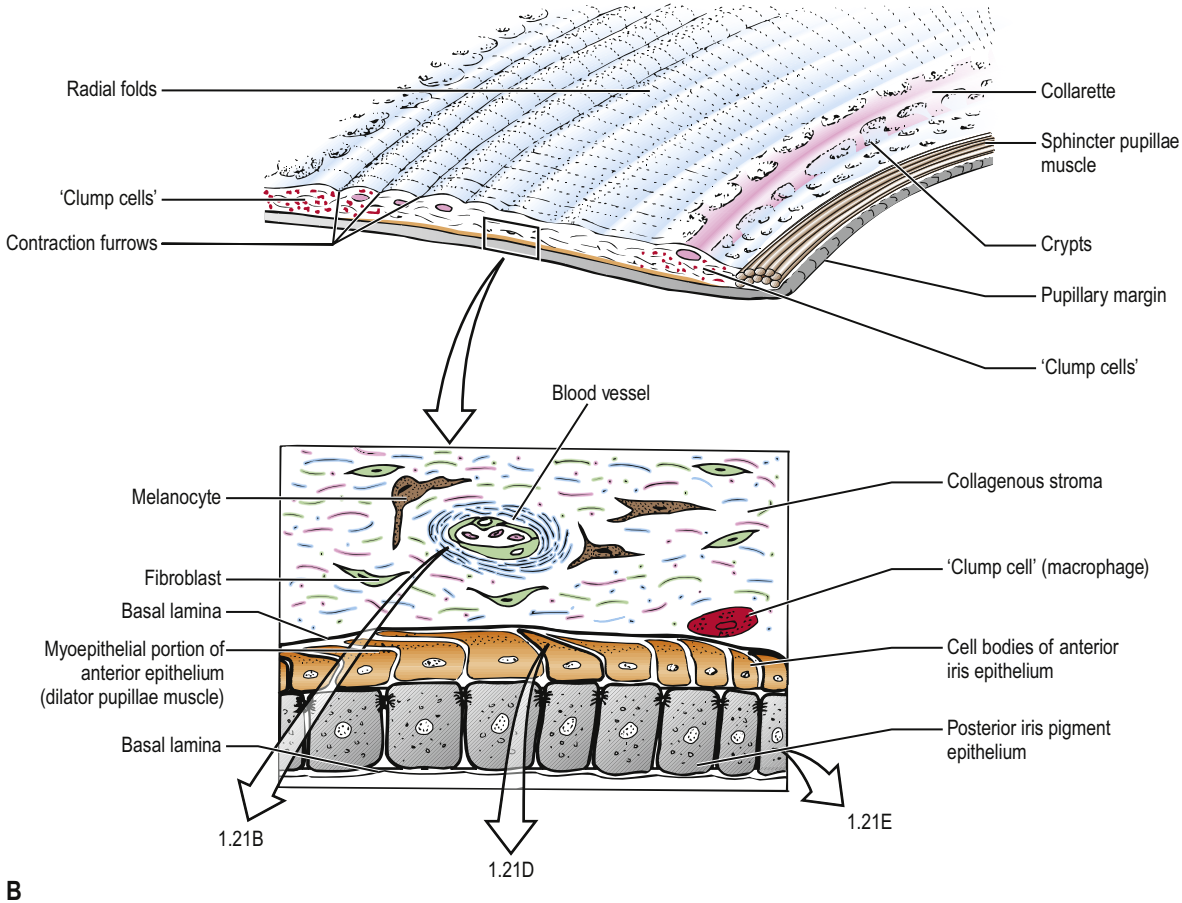
Microscopic anatomy

The iris consists of four layers: anterior border layer, stroma, dilator pupillae muscle and posterior pigment epithelium.

Anterior border layer (Figs 1.20B and 1.21A). This layer is not covered by a layer of epithelial cells, as early anatomists believed, but is in fact made up of modified stroma consisting of a dense collection of fibroblasts, melanocytes and a few interspersed collagen fibres. This layer is deficient in areas; consequently, the iris stroma is in free communication with the aqueous humour in the anterior chamber. Larger deficiencies in the anterior border layer are evident macroscopically as *crypts.* Aggregates of heavily pigmented melanocytes in this layer appear as *naevi. Melanocytes in the uveal tract are unlike skin melanocytes in*

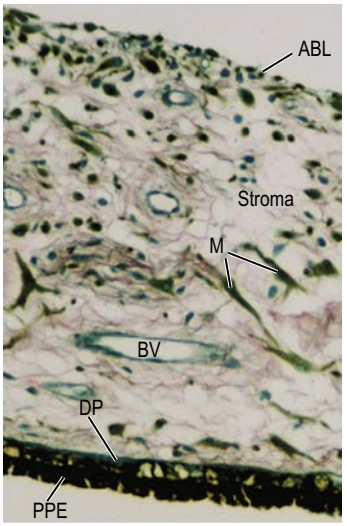


A

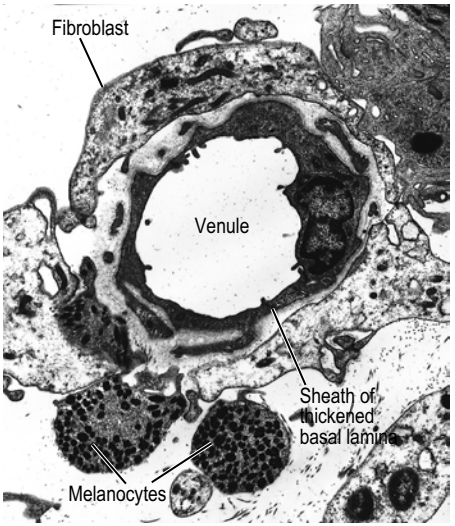


B

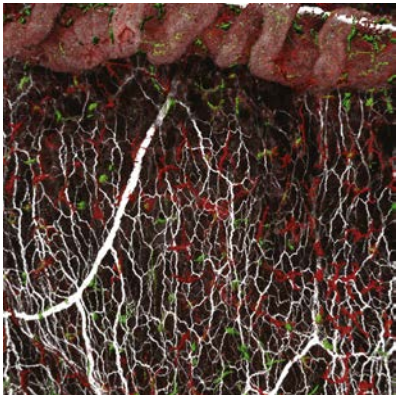
Fig. 1.20 The structure of the iris. (**A**) Clinical macroscopic image of a blue iris. Note the pupil *(P)*, pigmented pupil margin *(PM)* and pupillary zone *(PZ)* separated from the ciliary portion by the collarette *(Co)*. The root is attached at the iridocorneal angle deep to the limbus (*L*). Note the fine collagenous trabecula *(Tb)*, some of which form the boundaries of ovoid crypts *(Cp)*. (**B**) *Top diagram*: surface features of a portion of the iris. *Bottom diagram*: high-power exploded view summarizing the arrangement of the pigmented (posterior) and non-pigmented (anterior) layers.



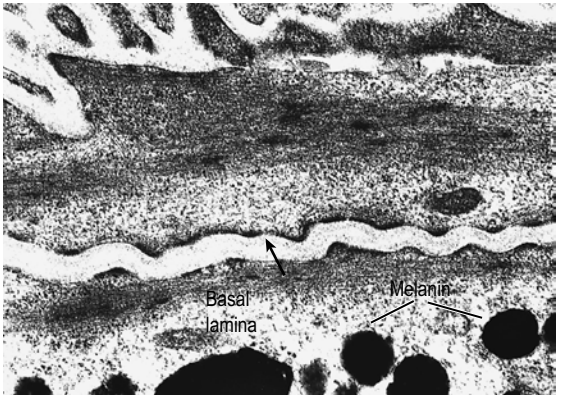
A



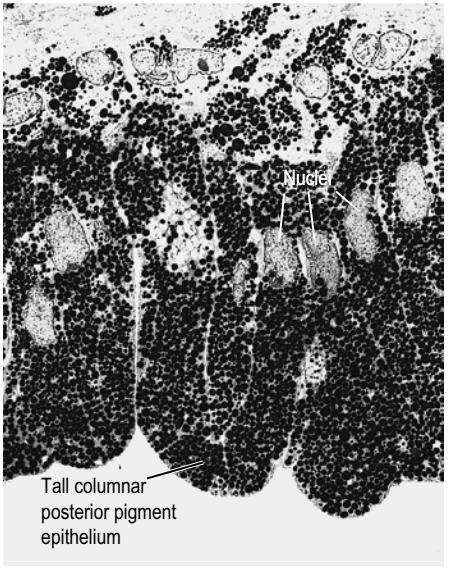
B



C



D



E

Fig. 1.21 (**A**) Histological section of primate iris showing the anterior border layer *(ABL)*, stroma containing melanocytes *(M)*, blood vessels *(BV)* and collagenous matrix (*red*) and posterior portion characterized by dilator pupillae muscle *(DP)* and posterior pigmented epithelium *(PPE)*. (**B**) Electron micrograph of an iris stromal vessel surrounded by fibroblasts and melanocytes. (**C**) Confocal microscopic image of the iris and ciliary body *(top of image)* from a Cx3cr1-GFP transgenic Balb/c (albino) mouse. This iris wholemount has been stained to show the nerves of the iris *(white*: WGA stain*)* and the extensive network of tissue resident macrophages (*red*: Iba1*)*. Cx3cr1^{+} cells of myeloid origin are GFP^{+} *(green)*. Mainly macrophages. (**D**) Ultrastructure of the contractile myoepithelial portion of the anterior iris epithelium, namely the dilator pupillae muscle: note the cytoplasmic microfilaments and membranous densifications, both characteristic features of smooth muscle cells. (**E**) Electron micrograph of the posterior iris epithelium illustrating the size, columnar shape and heavily pigmented nature of these cells. Original magnifications: **A** × 300; **B** × 7100; **D** × 16 000; **E** × 1400.

that they reside in connective tissue and do not constantly shed melanin into adjacent cells (the basis of skin pigmentation) and iris colour is a reflection of both the density of melanocytes, their location in the iris and their melanin content.

Stroma (*Fig. 1.21C–E*). This consists of loose connective tissue containing fibroblasts, melanocytes and collagen fibres (types I and III). The loose nature of this tissue and its free communication via openings in the anterior border layer allow fluid to move in and out of the stroma quickly during dilation and contraction. The iris stroma in humans contains numerous mast cells and macrophages, many of which are perivascular (see Fig. 1.21C). Many of the macrophages are heavily pigmented and a subgroup may form large ovoid 'clump cells' (of Koganei), which tend to accumulate near the iris root and sphincter pupillae muscle (Fig. 1.20B). The origin of this phagocytosed melanin is both from the ciliary/iris epithelium and stromal melanocytes. Lying free within the stroma close to the pupil margin is the *sphincter pupillae* muscle, a circumferential ring of smooth muscle fibres about 1 mm in width. The sphincter muscle consists of muscle bundles, each comprising six to eight smooth muscle cells, which are continuous via gap junctions and surrounded by a basal lamina. This muscle is innervated by parasympathetic nerve fibres derived from the oculomotor nerve (Edinger-Westphal nucleus houses preganglionic cell bodies whose fibres synapse in ciliary ganglion from where postganglionic fibres travel via the short ciliary nerves) although sympathetics also terminate in this muscle. The unusual embryological origin of this muscle from the neuroectoderm is described in Chapter 2. The *dilator pupillae muscle* is a layer of myoepithelial cells derived from the *anterior iris epithelium* (Fig. 1.21D). The iris epithelia are two layers apposed to each other apex to apex with the gap between being theoretical continuous with the subretinal space. The basal processes of the anterior iris epithelium are 4 μm in thickness and extend up to 50–60 μm in a radial direction, while the apices of the myoepithelial cells are lightly pigmented and closely apposed to the apical aspect of the *posterior pigment epithelium* (Figs 1.20B and 1.21A,E). The *dilator pupillae muscle* is innervated by non-myelinated sympathetic fibres whose cell bodies are situated in the superior cervical sympathetic ganglion. Its parasympathetic innervation seems less significant. The dilator pupillae extends only as far centrally as the outer margin of the sphincter pupillae.

Posterior pigment epithelium (*Fig. 1.21E*). This heavily pigmented layer consists of large cuboidal epithelial cells that appear black macroscopically on examination of the posterior surface of the iris. The posterior layer is derived from the inner neuroectodermal layer of the optic cup (see Chapter 2). The posterior epithelium extends for a short distance on to the anterior iris surface at the pupillary margin; this forms the black ruff seen on the pupil margin during slit-lamp examination of the eye (Fig. 1.20A). The posterior pigmented epithelial layer forms a series of radially arranged furrows (most evident near the pupil margin) and circumferential contraction folds (most evident in the periphery).

Pupil movements

Mydriasis (dilation) occurs in conditions of low light intensity and in states of excitement or fear. It is a result of the action of the dilator pupillae muscle.

Miosis (contraction) occurs in more illuminated conditions, during convergence, and while sleeping. It is the result of the action of the sphincter pupillae muscle.

Blood supply of the iris

The iris has a rich blood supply and extensive anastomoses. At the root there is an incomplete major 'circle' of the iris that is derived from anterior rami of the anterior ciliary arteries. Branches from here pass centripetally and form an incomplete minor arterial 'circle' at the level of the collarette.

The arteries have an unusual coiled form to accommodate the variable states of contraction of the iris. Veins lie close to the arteries, with larger veins primarily in the anterior stroma and smaller veins in the deeper layers. Veins drain posteriorly/centrifugally into the ciliary body and eventually the vortex veins (see page 53)

Iris capillaries are characterized by tight junctions and non-fenestrated endothelial cells that have a high density of endocytotic vesicles and tight junctions. This makes them less permeable to a variety of solutes than normal somatic vessels (hence they do not normally leak in fluorescein angiography). These vessels thus constitute an important component of the *blood–ocular barrier* (see Chapter 4, p. 224). The basal lamina of the endothelial cells is thickened (0.5–3 μm) and further strengthened by tube-like arrangements of perivascular collagenous/hyalinized layers (Fig. 1.21A,B) which helps prevent their collapse when the iris is fully dilated. Periarteriolar smooth muscle cells are rare and elastic fibres are absent.

Nerve supply

The iris possesses a rich three-dimensional nerve plexus of myelinated and non-myelinated nerves (Fig. 1.21C). The sensory nerves are branches of the long and short ciliary nerves, themselves branches of the nasociliary nerve (ophthalmic division of the trigeminal). The autonomic innervation of the iris muscles has been discussed previously.

The ciliary body

The ciliary body (Fig. 1.22A–D) is an approximately 5–6 mm wide ring of tissue that extends from the scleral spur anteriorly to the ora serrata posteriorly. Temporally it measures 5.6–6.3 mm and nasally 4.6–5.2 mm. It is divided into two zones, an anterior *pars plicata* (corona ciliaris) and a posterior *pars plana* (Fig. 1.22A). The ciliary body is approximately triangular in cross-section; its base faces the anterior chamber and the apex blends posteriorly with the vascular choroid. The pars plicata is 2 mm wide and consists of 70 radially arranged folds known as *ciliary processes* (Fig. 1.22B), each of which is 0.5–0.8 mm high and 0.5 mm wide. The tips are paler as a result of decreased epithelial pigmentation. Minor ciliary processes may be present in the valleys between the major processes. The pars plana is an approximately 4 mm wide zone stretching from the posterior limits of the ciliary processes to the *ora serrata*, the sharp serrated or dentate junction where non-pigmented ciliary epithelium undergoes a sharp transition to become the neural retina.

The ciliary body can be divided histologically into the ciliary epithelium, ciliary body stroma and ciliary muscle (Fig. 1.22A). Each is described below in the context of the three principal functions of the ciliary body: (1) accommodation; (2) aqueous humour production; and (3) production of lens zonules, vitreal glycosaminoglycans and vitreal

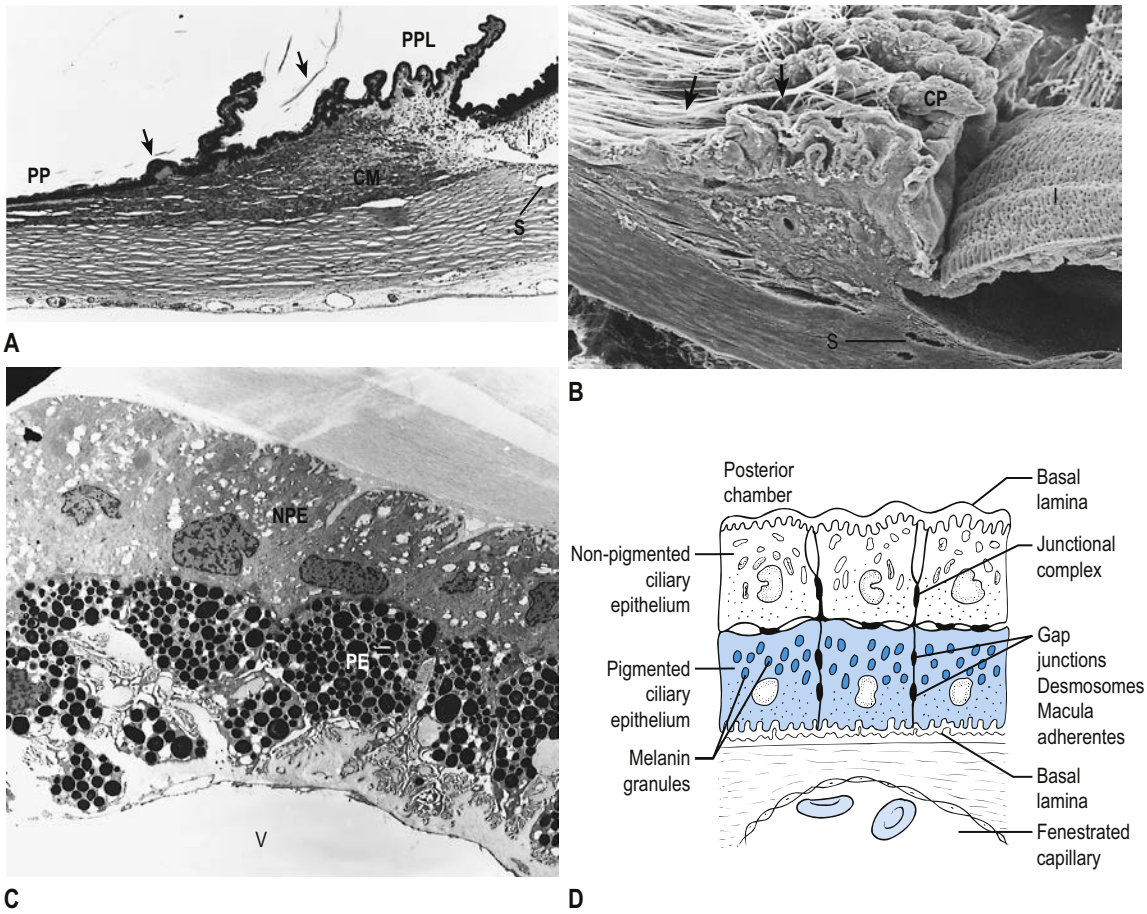


Fig. 1.22 (**A**) Histological section of the ciliary body showing the two major regions, the pars plana (*PP*) and the pars plicata (*PPL*), which includes the ciliary processes. (**B**) Scanning electron micrograph of the inner surface of the ciliary processes (*CP*) and iris (*I*). (**C**) Low-power electron micrograph of the pigmented (*PE*) and non-pigmented ciliary epithelia (*NPE*) of a ciliary process. Note the large fenestrated blood vessel (*V*) and the lens zonules blending with the basal lamina of the non-pigmented epithelium. (**D**) Diagrammatic representation of this double layer of epithelium, which constitutes the major site of the blood–aqueous barrier. *CM*, Ciliary muscle; *I*, iris; *S*, Schlemm's canal. *Arrows* in (**A**) and (**B**), zonular fibres. Original magnifications: **A** × 40; **B** × 45; **C** × 1600. (Fig. 1.22A courtesy W.R. Lee and Springer-Verlag.)

collagen. Chapter 4 (p. 234) deals in more detail on the physiological function of the ciliary body.

Accommodation

The anterior two-thirds of the ciliary body is occupied by the *ciliary muscle* which, in conjunction with the *lens zonules* (suspensory ligament) and the natural elastic nature of the lens fibres and capsule, functions to alter the refractive power of the lens. Histologically, the ciliary muscle in meridional sections consists of three groups of smooth muscle fibre bundles embedded in a vascular connective tissue stroma (Figs 1.18 and 1.22A). The stroma contains melanocytes, fibroblasts and occasional immune cells such as mast cells, macrophages and lymphocytes.

The *outer longitudinal muscle fibres* are attached to the scleral spur and therefore indirectly to the corneoscleral trabeculae anteriorly; posteriorly they are anchored to the inner aspect of the sclera. The *middle oblique* or *radial muscle fibres* are continuous with inner corneoscleral trabeculae. The *inner circular muscle fibres* (one of the three so-called Müller's muscles associated with the eye and adnexa) appear as cross-sectional profiles in conventional meridional sections of the ciliary body (Figs 1.18C and 1.22A). A three-dimensional scheme has been proposed to explain the interrelationship between all three groups of muscle bundles; it appears that the fibres are all components of one interwoven fibre network (see Fig. 1.23 for summary).

Action of the ciliary muscle on aqueous outflow. There is some evidence that contraction of the longitudinal ciliary muscle fibres causes an inwards and posterior movement of the scleral spur as well as an effect via extensions of these tendons to the inner wall of Schlemm's canal, hence distending the inter- and intratrabecular spaces of the trabecular meshwork and preventing collapse of Schlemm's canal during periods of high pressure. This has been proposed as mode of action of miotic drugs such as pilocarpine, which are used to increase aqueous outflow facility in patients with *glaucoma*.

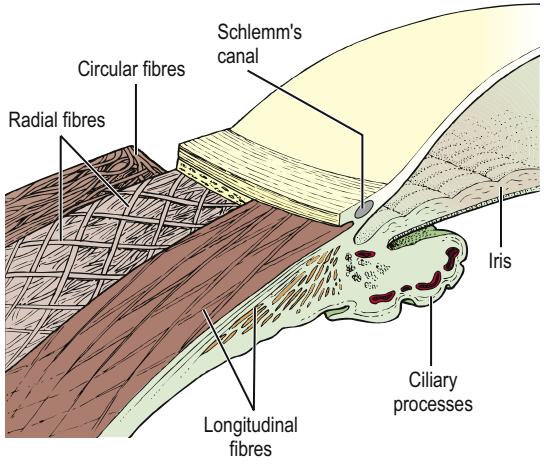


Fig. 1.23 Arrangement of the ciliary muscle fibres as seen on external view with the sclera removed. (After Rohen and Ciliankorpër, 1964.)

Production of aqueous humour

Aqueous humour is a clear colourless fluid actively secreted by the ciliary processes (for details of aqueous constituents see Chapter 4). The ciliary processes consist essentially of delicate finger-like protrusions of loose vascular connective tissue covered by a bilaminar cuboidal/columnar neuroepithelium, the outer pigmented and inner non-pigmented ciliary epithelium, both derived embryologically from the double neuroectoderm of the optic cup (see Chapter 2). The morphological basis of the blood–aqueous barrier is depicted in Fig. 1.22B. Aqueous humour is actively secreted by the inner non-pigmented ciliary epithelium whose apices are connected by junctional complexes including tight junctions. Macromolecules, having filtered through the highly permeable fenestrated stromal capillaries, pass between pigment epithelium cells held together only by punctate adherentes (macula adherentes)-like junctions or permeable band-like junctions (zonula adherentes) and are prevented from passing into the posterior chamber. The non-pigmented ciliary epithelium has morphological features characteristic of secretory epithelium, namely numerous mitochondria, and histochemically can be shown to contain enzymes, such as carbonic anhydrase and ATPase, necessary for active fluid transport (see Chapter 4, p. 229, Box 4.12).

Production of zonules, vitreal collagen and vitreal hyaluronic acid

The non-pigmented inner ciliary epithelial cells of the pars plana are cuboidal, columnar or irregular, depending on age and location. It is likely that these cells play a role in secretion of zonular fibres, the extracellular vitreal components, i.e. collagen and hyaluronic acid, and the inner limiting membrane, especially during embryonic development.

Some inner non-pigmented cells 'tilt' anteriorly, suggesting traction by the zonular fibres. The lens zonules (of Zinn) or suspensory ligament of the lens merge with the fibrous basal lamina material of the non-pigmented ciliary epithelium (Fig. 1.22C). The precise mode of zonule synthesis is still unclear. Tall non-pigmented ciliary epithelial cells in the pars plana have ultrastructural and histochemical features that indicate active hyaluronic acid secretion.

Blood supply of the ciliary body

The blood supply of this region is derived from long posterior ciliary arteries and anterior ciliary arteries. The two long posterior ciliary arteries arise from the ophthalmic artery and after piercing the sclera near the optic nerve head they travel forward in the choroid in the medial and lateral horizontal plane and divide in the ciliary body before anastomosing with anterior ciliary branches, thus forming the major 'circle' of the iris (see p. 54, Fig. 1.39). From this circle arises muscular and recurrent choroidal branches and the numerous branches that form the vascular plexus in the ciliary processes. Venous return occurs predominantly posteriorly through the system of vortex veins and less so through the anterior ciliary veins.

Nerve supply of the ciliary body

The ciliary body has rich parasympathetic, sympathetic and sensory innervations.

Parasympathetic innervation. Preganglionic neurone cell bodies are located in the Edinger-Westphal nucleus, which lies posterior to the main oculomotor nucleus in the rostral midbrain at the level of the superior colliculus (see Fig. 1.52). Their axons travel in the oculomotor (III) nerve and synapse with postganglionic cell bodies located in the ciliary ganglion. The postganglionic fibres travel to the eye in the short ciliary nerves and terminate as an extensive plexus in the ciliary muscle. The action of these nerves is mediated by acetylcholine on muscarinic receptors (see Chapter 5, Box 5.5).

Sympathetic innervation. Preganglionic neurone cell bodies are situated in the lateral grey horn of the first thoracic segment of the spinal cord. Preganglionic fibres relay in the superior cervical ganglion (adjacent to vertebrae C2 and C3, behind the internal carotid artery) with second-order neurones. Postganglionic fibres leave the ganglion as the internal carotid nerve and plexus. These fibres may reach the orbit as either direct branches of the internal carotid plexus or by joining the ophthalmic division of the trigeminal and its main branch in the orbit, the nasociliary nerve. Sympathetic fibres may either pass directly to the retrobulbar plexus behind the eye or through the ciliary ganglion uninterrupted. From the ciliary ganglion, fibres are distributed via the short ciliary nerves to the blood vessels of the eye, including those in the ciliary body. Some terminal filaments of the internal carotid plexus may also be distributed via the ophthalmic artery and its branches. The sympathetic action is mediated by the action of norepinephrine on two subclasses of receptors, α_1 and β_2 adrenoreceptors, both of which are inhibitory (see Chapter 4, p. 223).

Sensory innervation. Sensory innervation is derived from the nasociliary nerve; however, the function of these fibres is unknown.

The lens and zonular apparatus

The lens (Fig. 1.24) is a highly organized system of specialized cells (so-called lens fibres), which constitutes an important component of the optical system of the eye and fulfils the important function of altering the refractive index of light entering the eye to focus on the retina. While it has less refractive power (15 dioptres) than the cornea, the lens has the ability to change shape, under the influence of the ciliary muscle, and thus alter its refractive power. The range of dioptric power diminishes with age (see Chapter 4, p. 242). The transparency of the lens is a result of the shape, arrangement, internal structure and biochemistry of the lens cells or lens fibres. Decreased transparency of the lens is known as cataract (Box 1.13).

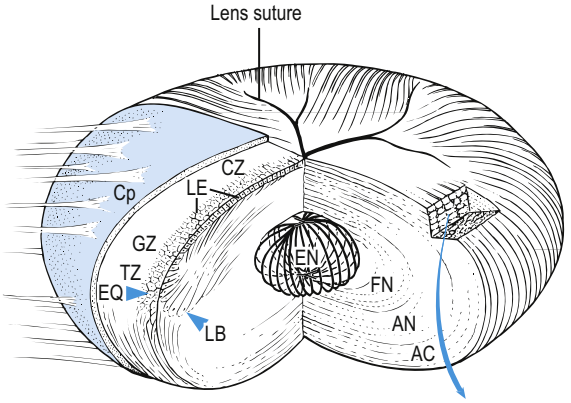
Position, size and shape

The lens, enclosed in its capsule, lies behind the iris and in front of the vitreous body. It is encircled by the ciliary processes and held in position by the zonular fibres laterally, the anterior vitreous face posteriorly (patellar fossa), and the iris anteriorly (see Fig. 1.10). It is normally transparent and avascular following regression of the pupillary membrane and tunica vasculosa lentis late in foetal development (see Chapter 2). It receives its nourishment from the aqueous and vitreous humours. It is a biconvex, ellipsoid structure with differing radius of curvature on the anterior and posterior surfaces. The anterior curvature is approximately 10 mm (range 8–14 mm) and the posterior curvature is approximately 6 mm (range 4.5–7.5 mm). The centre points of these surfaces, described as the *anterior* and *posterior poles*, are connected by an imaginary *axis*. The anterior pole lies 3 mm from the posterior corneal surface. The anterior and posterior surfaces are separated by the *equator*, which has a ridged (indented) appearance caused by the zonular fibres. In the adult eye the lens measures approximately 10 mm in diameter and has an axial length of 4 mm. The lens continues to grow (0.023 mm per year) and alters shape throughout life. It becomes rounder with age, especially after the age of 20 years.

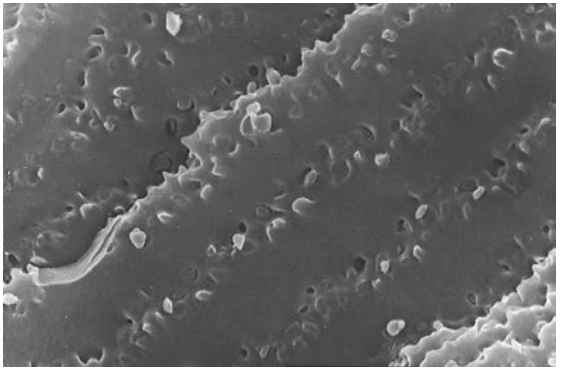
Structure

The lens consists of three parts: (1) the capsule; (2) anterior or lens epithelium; and (3) the lens fibres (Fig. 1.24A,D).

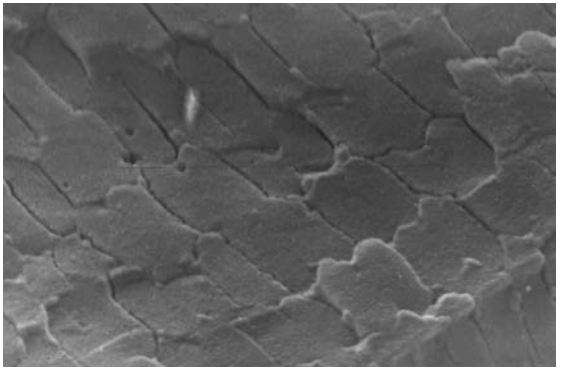
Lens capsule. The *lens capsule* is a thickened, smooth, basement membrane produced by the lens epithelium and lens fibres. It completely envelops the lens and has regions



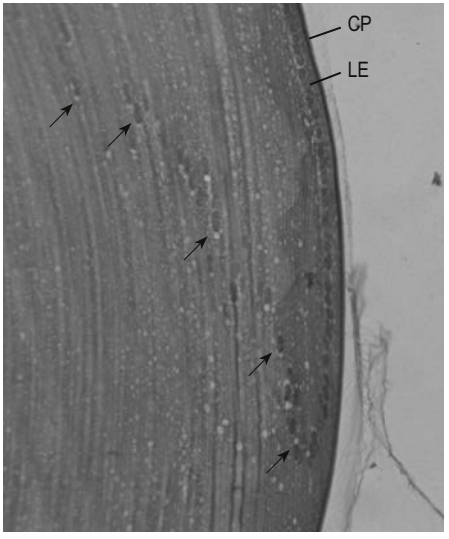
A



B



C



D

Fig. 1.24 (**A**) The structure of the lens. Scanning electron micrographs of lens fibres in (**B**) longitudinal section and in (**C**) cross-section. Note the 'tooth and peg' arrangement interlocking the adjacent lens fibre surfaces. (**D**) Histological section of the equatorial region of the lens showing the orientation of the lens fibres, the lens capsule, lens epithelium and lens bow *(arrows)*. *AC*, Adult cortex; *AN*, adult nucleus; *Cp*, capsule; *CZ*, central zone; *EN*, embryonic nucleus; *EQ*, equator; *FN*, foetal nucleus; *GZ*, germinative zone; *LB*, lens bow; *LE*, lens epithelium; *TZ*, transitional zone. Original magnifications: **B** × 2500; **C** × 5000; **D** × 150.

Box 1.13 **Clinical correlates**

Cataract

Cataract is the loss of normal lens transparency, besides the normal age-related yellowing, and may be caused by cumulative damage by ultraviolet light or perturbations in lens fibre biochemistry. Opacification may be the result of damage or disruption of the capsule, the lens fibre configuration or the lens epithelium (see Chapter 4, p. 211).

of variable thickness, being thickest pre- and post-equatorially (17–28 μm) and thinner at the posterior (2–3 μm) than at the anterior pole (9–14 μm). Ultrastructural examination reveals a fibrillar or lamellar appearance. The interfibrillar matrix consists of basement membrane glycoproteins (type IV collagen) and sulphated GAGs which are responsible for its prominent periodic acid–Schiff-positive staining properties in histological sections. It possesses elastic properties and, when not under tension of the zonules, the capsule together with the cortex causes the lens to assume a more rounded shape.

Lens epithelium. This is a simple cuboidal epithelium (Fig. 1.24A,D) restricted to the anterior surface of the lens. The cells become more columnar at the equator. As they elongate, the apical portion comes to lie deeper to other, more anteriorly positioned, lens cells. These elongated lens cells are known as lens 'fibres' (Fig. 1.24D). The manner in which equatorial lens epithelial cells are transformed into lens fibres is depicted in Fig. 1.24A. The cell nucleus and cell body sink deeper into the lens as further cells are laid down externally. Mitotic activity is maximal in the pre-equatorial and equatorial lens epithelium, known as the germinative zone (Fig. 1.24A).

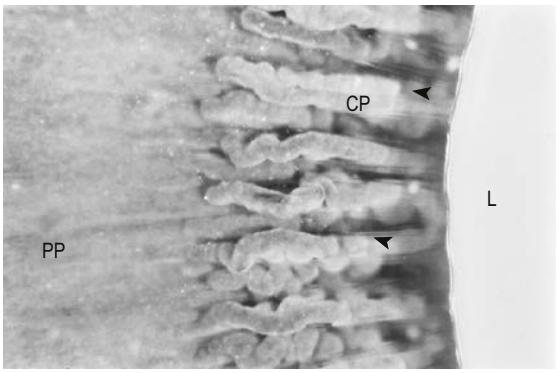
Lens fibres. While each lens fibre is only a 4 × 7 μm wide hexagonal prismatic band in cross-section (Fig. 1.24C), it may be up to 12 mm (12 000 μm) in length (i.e. 1000 times longer than its width). The apical portion of the elongated lens cell (or lens fibre) passes anteriorly, the basal portion posteriorly. The cell nucleus migrates anteriorly as the cell is pushed deeper in the lens, hence creating the anteriorly oriented *lens bow* (Fig. 1.24D). The meridionally oriented lens fibres extend the full length of the lens, meeting at the anterior and posterior sutures (Fig. 1.24A). Deeper (hence older) lens fibres are anucleate. Continual growth of the lens, by addition of superficial strips of new cells, produces a series of concentrically arranged laminae, similar to the layers of an onion (best seen by dissecting a fixed or frozen lens). In life, the outer *cortex* of the lens has a softer consistency than the hard central *nucleus*.

Lens fibres are tightly packed with little intercellular space. Neighbouring cells are linked by ball-and-socket cytoplasmic interdigitations (Fig. 1.24B,C) and numerous gap junctions. The junctions may aid maintenance of centrally positioned cells (via intercellular and molecular coupling or metabolic cooperation) some distance from the source of nutrition (aqueous humour). Superficially located lens fibres are rich in ribosomes, polysomes and rough endoplasmic reticulum, and actively synthesize unique lens proteins, lens crystallins (see Chapter 4, p. 235); however, the cytoplasm of mature lens fibres appears homogeneous. Lens fibres are rich in cytoskeletal elements oriented parallel to the long axis of the cell. Disturbances in the biochemistry of lens fibres or in their arrangement can result in opacification known as cataract (Box 1.13 Clinical correlates).

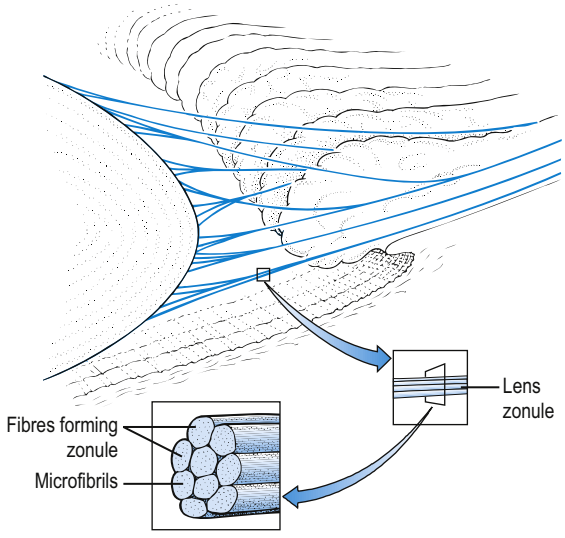
Lens zonules (zonular apparatus)

The lens is held in position by a complex three-dimensional system of radially arranged zonules (zonules of Zinn or the suspensory ligament of the lens) (Fig. 1.25). These delicate but tough fibres are attached to the lens capsule 2 mm anterior and 1 mm posterior to the equator, arise from the pars plana region of the ciliary epithelium and pass forward closely related to the lateral surfaces of the ciliary processes (Fig. 1.25A,B). They are present for the full 360 degrees of the lens circumference and thus resemble a circular trampoline. Understanding their distribution and the tension forces they are able to withstand is important in modern day cataract surgery where the phacoemulsification probe exerts considerable downward force on the zonule. In humans, the zonules normally transmit the radial forces that flatten the lens, allowing the eye to focus on distant objects. However, the zonules are also present in non-accommodating species, where they presumably play a role in lens centration.

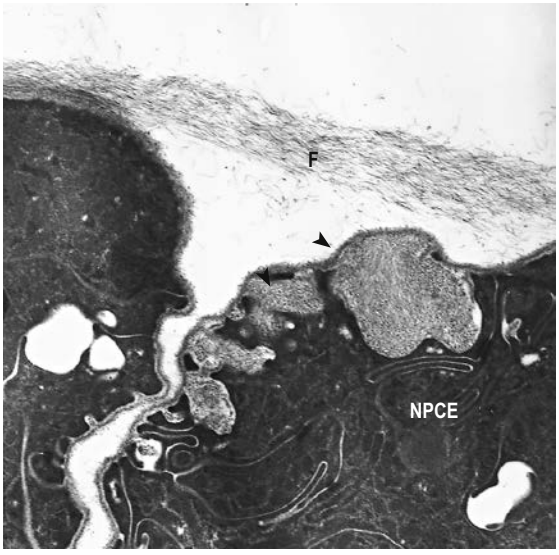
The zonules consist of dense, glassy bundles 5–30 μm in diameter. Each bundle consists of a series of fine fibres (0.35–1 μm in diameter), themselves composed of 10–12 nm diameter microfibrils. Biochemically, the zonules are unique, most closely resembling the periodontal ligament of the teeth. Proteomic analysis has revealed that the zonules are composed principally of fibrillin, a 350 kDa cysteine-rich glycoprotein. Although described as elastic fibres they do not contain significant amounts of elastin. Mutations in the fibrillin-1 gene in Marfan syndrome manifest in the eye (Box 1.14 Clinical correlates). The zonular fibres are synthesized by cells of the non-pigmented ciliary epithelium. At their proximal end they emerge from the basal lamina of the ciliary epithelial cells. The distal portions of the zonular fibres connect to the lens capsule near the lens equator. In humans, the fibres appear to insert directly into the capsule and their tips can be seen



A



B



C

Fig. 1.25 Arrangement of the zonular fibres. (**A**) Macroscopic view of ciliary processes *(CP)* and intervening lens zonules *(arrowheads)* inserting into the lens capsule in a monkey eye. (**B**) Arrangement of the different groups of lens zonules. (**C**) Electron micrograph showing the close association of zonular fibres *(F)* to the non-pigmented epithelial cells *(NPCE)* of the ciliary processes. Note the material similar to zonular fibres beneath the basal lamina of the epithelium *(arrowheads)*. *L*, lens; *PP*, pars plana. Original magnifications: **A** × 24; **C** × 10 500.

Box 1.14 **Clinical correlates**

Ectopis lentis in Marfan syndrome

In humans, mutations in the fibrillin-1 gene cause Marfan syndrome, a condition characterized by disturbances in connective tissue and skeletal elements. In these patients, the zonules are disrupted and the lens dislocated or malpositioned within the eye, a condition known as ectopis lentis.

to terminate below the capsular surface. The lens zonules are synthesized and maintained by cells in the non-pigmented ciliary epithelium. The zonule is synthesized during embryonic and early postnatal development and it is likely that zonular proteins turn over slowly, if at all.

Accommodation. In the non-accommodated state, the ciliary body maintains tension on the zonules. During accommodation, movement of the ciliary body causes slackening of the zonules, whereupon the lens assumes

an increased anterior curvature, with resultant increase in refractive power, owing to elasticity of the lens capsule and the outer cortical layers. Some authorities believe that there are two sorts of zonules: main zonules and 'tension' zonules, the latter being placed under tension during accommodation. Aging changes in the lens compromise its ability to accommodate (Box 1.15 Clinical correlates).

Posterior zonules are closely associated with the collagenous material of the anterior hyaloid membrane. Zonules running perpendicular to the main zonule stream form circumferential bands near the base of the ciliary processes or in the pars plana (posterior zonular girdle) and over the apices of the ciliary processes (anterior ciliary girdle).

Anterior and posterior chambers

The cavity anterior to the lens and lens zonules is divided into two chambers by the iris. These two chambers, the larger anterior and smaller posterior, communicate through the pupil. The boundaries of these two chambers are shown in Fig. 1.26.

Posterior chamber

This is a very small irregularly shaped space whose size varies during accommodation. It is approximately triangular with the apex at the pupil margin; the base is formed by the ciliary processes; the posterior border is the lens and zonular apparatus; and the anterior border is the posterior surface of the iris. Aqueous humour secreted by the ciliary processes continually enters this chamber before passing through the pupil into the anterior chamber.

Box 1.15 **Clinical correlates**

Ageing changes in the eye: presbyopia

This condition may develop around the age of 40–50 years when the elasticity of the lens markedly decreases and there is associated atrophy of the ciliary muscle fibres; consequently, the lens fails to change shape sufficiently during accommodation. This becomes evident as decreasing ability to read, i.e. use near vision.

Anterior chamber

The anterior chamber is bound in front by the cornea and posteriorly by the anterior iris surface and the pupillary portion of the lens. The lateral recess of the anterior chamber is formed by the iridocorneal angle occupied by the trabecular meshwork. The anterior chamber is deepest centrally (3 mm) and contains approximately 250 µL of aqueous humour. Aqueous humour is produced at around 2–4 µL/minute. A full account of aqueous humour dynamics and function can be found in Chapter 4 (see p. 31 and Chapter 4, p. 229).

The vitreous

The vitreous cavity (Fig. 1.27) is the largest cavity of the eye (two-thirds the volume of the eye, weight 3.9 g) and contains the vitreous humour or vitreous. It is bound anteriorly by the lens, posterior lens zonules and ciliary body and posteriorly by the retinal cup. The vitreous is a transparent viscoelastic gel that is more than 98% water, with a refractive index of 1.33. Its viscosity is two to four times that of water. The main constituents of the vitreous, besides water, include hyaluronan (hyaluronic acid), collagens type II, VI and XI, fibronectin, fibrillin and opticin. The gel structure of the vitreous body is dependent on the collagenous constituents and not the hyaluronan. The fine-diameter type II collagen fibres (8–12 nm in diameter) entrap large coiled hyaluronan molecules (see Chapter 4, p. 245).

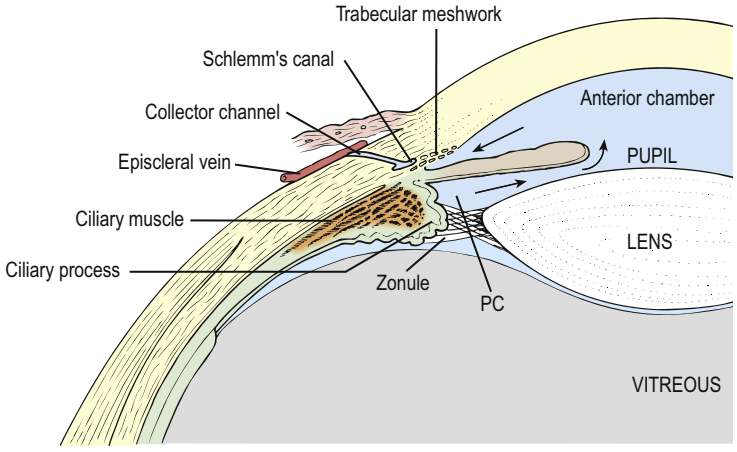


Fig. 1.26 Diagram of the anterior segment of the eye. The major pathway followed by aqueous humour from the posterior chamber *(PC)* to the anterior chamber and outflow pathways is shown by *arrows*.

The vitreous is shaped like a sphere with an anterior depression, the *hyaloid fossa* (also known as the patellar or lenticular fossa) (Fig. 1.27A). The vitreous is traditionally regarded as consisting of two portions: a *cortical zone*, characterized by more densely arranged collagen fibrils, and a more liquid *central vitreous*. The vitreous can be further subdivided for descriptive purposes into three major topographical zones, as shown in Fig. 1.27A.

The cortical vitreous is attached by condensation of fine collagen fibrils at several points around its margin to (Fig. 1.27A,B):

- the peripheral retina and pars plana via the *vitreous base*, a 3–4 mm wide band
- the posterior lens capsule (ligamentum hyaloide capsulare)
- the retina along the margins of the optic disk (base of the hyaloid canal), although this is disputed
- the inner limiting membrane of the retina, especially near retinal vessels (most variable and weakest of attachments).

The *central vitreous* possesses less collagen than cortical vitreous. It is traversed by a central fluid-filled canal (hyaloid or Cloquet's canal), which represents the remnants of the course taken by the hyaloid artery that supplied both the vitreous and lens during embryogenesis and early foetal life (see Chapter 2). The retrolental and intermediate zones in the human eye are semi-liquid. The existence of an ordered, organized, vitreal structure is still controversial owing to the problems of studying a gel that is 98.5%–99.7% water.

The human vitreous begins to degenerate at adolescence, leading to the appearance of liquid-filled cavities and fibrillar strands, such as the retrolental, preretinal and other named tracts of significance only to vitreal specialists. Most of the central tracts (except preretinal) are mobile and change during eye movements. The posterior vitreous cortex may possess zones of reduced density or 'cortical holes' or pockets that, if present, occur close to the fovea, retinal vessels and any developmental anomalies. While these are normal features, secondary pathological holes may develop following various disease processes (Box 1.16 Clinical correlates).

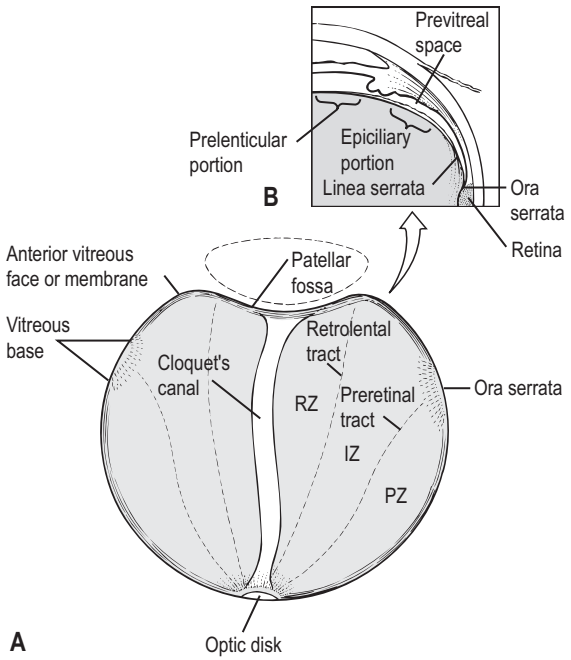


Fig. 1.27 (**A**) Diagram summarizing the anatomical zones of the vitreous and the major vitreal condensations (the retrolental tract, preretinal tract, vitreous base, anterior hyaloid face, vitreous cortex). (**B**) Higher power of the anterior vitreous face and its relation to the ciliary body and iris. *IZ*, Intermediate zone; *PZ*, preretinal zone; *RZ*, retrolental zone.

Vitreous cells (Fig. 1.28A,B)

The vitreous is essentially acellular; however, occasional isolated cells may occur in the cortex, particularly near the vitreous base, optic disk and retinal vessels. Cells known as hyalocytes, which have the morphological, ultrastructural, immunophenotypic and functional characteristics of macrophages, are the main cell type in the vitreous. During development they likely originate from yolk sac precursors such as the microglia but are replenished at a low rate in the adult eye from bone marrow derived monocytes. A marked vitreal cellular infiltrate is indicative of pathological or inflammatory processes in adjacent tissues, e.g. uveoretinitis.

Retina and retinal pigment epithelium (Fig. 1.29)

The retina is the innermost of the three coats of the eye. It consists of two primary layers: an *inner neurosensory retina* and an outer simple epithelium, the *retinal pigment epithelium* (RPE). These two layers can be traced embryologically

Box 1.16 **Clinical correlates**

Posterior vitreous detachment

The thin potential (subhyaloid or sublaminar) space between the surface of the cortical vitreous and the retina may fill with fluid in cases of vitreous detachment. The vitreous might detach relatively easily in the posterior segment where it is less weakly bound to the retina. The fluid may accumulate rapidly in the case of rhegmatogenous vitreous detachment and more slowly where the cortex is not ruptured (arrhegmatogenous vitreous detachment). The former is an age-related change in the vitreous. Vitreous detachment may predispose to retinal detachment (see Chapter 9, p. 498).

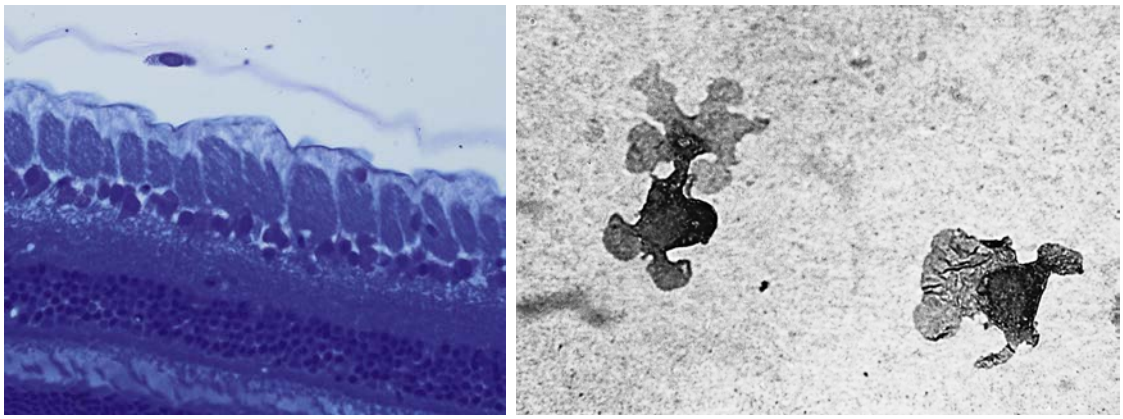


Fig. 1.28 (**A**) Section of the inner retina in primate eye showing a hyalocyte. (**B**) CD169+ hyalocytes in the rat subhyaloid space (plan view, retinal wholemount).

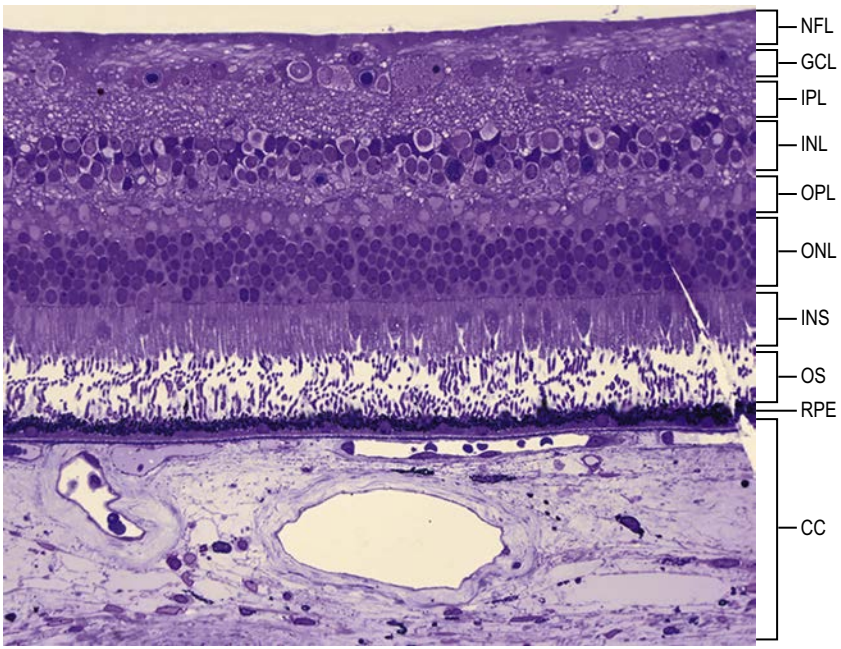


Fig. 1.29 Low-power micrograph of the human retina (resin histology). *CC*, Choriocapillaris; *GCL*, ganglion cell layer; *INS*, inner segments; *INL*, inner nuclear layer; *IPL* inner plexiform layer; *NFL*, nerve fibre layer; *ONL*, outer nuclear layer; *OPL*, outer plexiform layer; *OS*, outer segment; *RPE*, retinal pigment epithelium. Original magnification: ×150.

to the inner and outer layers of the invaginated optic cup (see Chapter 2). In the adult they are continuous anteriorly with the epithelial layers over the ciliary processes and posterior iris surface. Between the neural retina and RPE is a potential space, the *subretinal space*, across which the two layers must adhere. The neural retina is firmly attached only at its anterior termination, the ora serrata, and at the margins of the optic nerve head. Detachment of the retina occurs between the neural retina and RPE (Box 1.17 Clinical correlates).

The retina is bound externally by Bruch's membrane and on its internal aspect by the vitreous (Fig. 1.26). It is continuous with the optic nerve posteriorly, the site of exit of ganglion cell axons from the eye.

Box 1.17 **Clinical correlates**

Retinal detachment

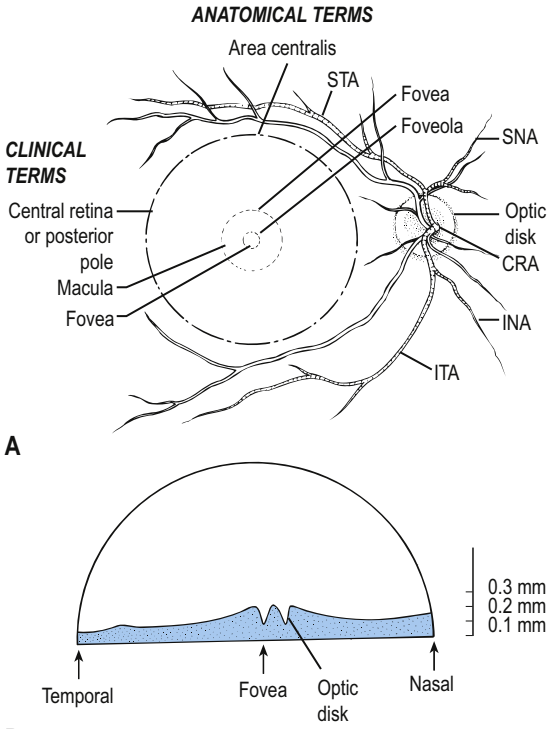
In this condition the neural retina separates from the retinal pigment epithelium (RPE), thus reopening the embryonic intraretinal space or optic ventricle (analogous to the ventricles of the brain), known in the adult as the *subretinal space*. Proteinaceous exudate tends to accumulate in the newly formed space (see Chapter 9, p. 498). Adhesion of the neural layer and RPE is normally maintained by negative pressure, viscous proteoglycans in the subretinal space and electrostatic forces.

Regions of the retina

Before considering the histological structure of the individual retinal layers (Fig. 1.29) and their constituent cell types, it is important that the reader appreciates the regional or topographical variations of the human retina (Fig. 1.30A,B).

There is often confusion regarding the terminology of the regions of the retina owing to the use of differing terms by clinicians and anatomists. Fig. 1.30A gives a summary of both terminologies.

1. The *posterior pole* or *central retina* (anatomically, area centralis) is a 5–6 mm diameter circular zone of retina situated between the superior and inferior temporal arteries. This region is cone-dominated and is characterized histologically by the presence of more than a single layer of ganglion cell bodies.
2. The *macula lutea* (anatomically, fovea) is a 1.5 mm diameter area in the posterior pole, 3 mm lateral to the optic disk. It is partly yellow as a result of xanthophyll carotenoid pigments (zeaxanthin and lutein) in the cone axons. This may serve to act as a short wavelength filter protecting against UV irradiation.
3. The *fovea centralis* (anatomically, foveola) is a central 0.35 mm wide zone in the macula, consisting of a depression surrounded by slightly thickened margins. Cone photoreceptors are concentrated here at maximum density to the exclusion of rods. The inner retinal layers in the margins of the pit (clivus) are displaced laterally (Fig. 1.30C). The foveal retina is avascular and relies wholly on the choriocapillaris for nutritional support.
4. The *optic disk* lies 3 mm medial to the centre of the macula (fovea). There are no normal retinal layers in this zone (blind spot) as ganglion cell axons from the retina pierce the sclera to enter the optic nerve. This pale pink/whitish area is 1.8 mm in diameter with a slightly raised rim. The central retinal vessels emerge at the centre of the optic disk, pass over the rim, and radiate out to supply the retina (Fig. 1.30A). The vein usually lies lateral to the artery.
5. The *peripheral retina* is the remainder of the retina outside the posterior pole. The distance from the optic disk to the ora serrata is 23–24 mm on the temporal aspect and



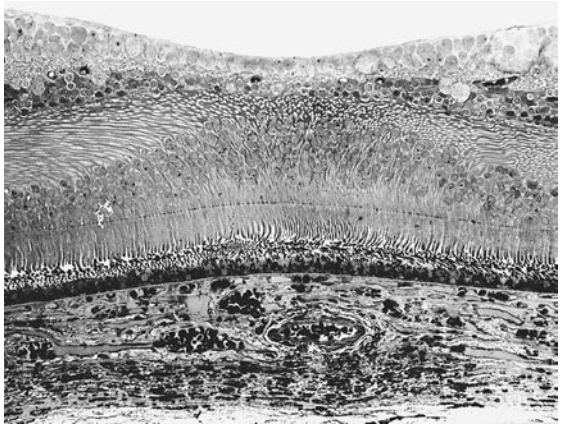


Fig. 1.30 (**A**) Anatomical and clinical terminology used to describe the regions of the retina: *CRA*, central retinal artery; *INA*, inferior nasal artery; *ITA*, inferior temporal artery; *SNA*, superior nasal artery; *STA*, superior temporal artery. (**B**) Regional variations in retinal thickness. (**C**) Section of the retina at the fovea. Original magnification: × 150. (Courtesy D. Aitken and W.R. Lee.)

approximately 18.5 mm on the nasal aspect. The peripheral retina is 110–140 μm in thickness, rich in rods and possesses only one layer of ganglion cell bodies.

6. The *ora serrata* is the scalloped or dentate anterior margin of the sensory retina. At this transition zone, the neuroretina is continuous with the columnar non-pigmented epithelial cells of the pars plana. The ora serrata is around 1 mm closer to the limbus on the nasal than on the temporal side.

For descriptive purposes, the retina is divided into nasal and temporal halves by a vertical line through the fovea. The optic nerve head is often used as a central point to describe the retina as having supero- and inferonasal and supero- and inferotemporal quadrants. The area of the retina is approximately 1250 mm^2 and varies in thickness from 100 μm (periphery) to 230 μm (near the optic nerve head) (Fig. 1.30B).

Retinal pigment epithelium (Fig. 1.31)

The RPE is a continuous monolayer of cuboidal/columnar epithelial cells, which extends from the margins of the optic nerve head to the ora serrata, where it is continuous with the pigment epithelium of the pars plana. This cell layer has many physical, optical, metabolic/biochemical and transport functions, which play a critical role in the normal visual process (Fig. 1.31A). These include: maintaining adhesion of the neurosensory retina; providing a selectively permeable barrier between the choroid and neurosensory retina; phagocytosis of rod and, to a lesser extent, cone outer segments; synthesis of interphotoreceptor matrix; absorption of light and reduction of light scatter within the eye, hence improving image resolution; and transport plus storage of metabolites and vitamins (especially vitamin A). The complex morphology of this neuroectoderm-derived epithelium reflects these multiple functions.

Cell size, shape and structure (*Fig. 1.31*). The RPE cells vary in size and shape depending on age and location, being more columnar in the central retina (14 μm tall, 10 μm wide) and more flattened (10–14 μm tall, 60 μm wide) in the peripheral retina. The basal aspect of the cells lies on *Bruch's membrane* and their apical surface is intimately associated with the photoreceptor outer segments (Fig. 1.31A,B). When examined *en face* they form a highly organized hexagonal pattern of homogeneously sized cells (Fig. 1.31C). The number of RPE cells per eye varies from 4.2 million to 6.1 million. The correlation between structure and function of this monolayer is summarized diagrammatically in Fig. 1.31A. Ageing changes in the RPE are important to understanding the pathogenesis of age-related macular degeneration (Box 1.18 Clinical correlates).

Neurosensory retina (Figs 1.29 and 1.32)

The neurosensory retina is a thin transparent layer of neural tissue that in life has a red/purple tinge caused by the presence of visual pigments; however, after death and in fixed specimens, it is white or opaque and often detached from the underlying RPE. Light stimuli are converted into neural impulses in the retina. These impulses are then partially integrated locally before transmission to the brain via the ganglion cell axons in the optic nerve. An appreciation of the anatomy of the neurosensory retina is crucial to an understanding of the physiology of vision (see Chapter 5).

The retina consists of several cell types among which neural cells predominate; other cell types include glial cells, vascular endothelium, pericytes and microglia. The three principal neurone cell types that relay impulses generated by light are photoreceptors, bipolar cells and ganglion cells, and their activity is modulated by other cell types such as horizontal cells, amacrine cells, and possibly by non-neuronal elements. It is the culmination of this neural processing concerning the visual image that is eventually transmitted to the brain along the optic nerve. Retinal cells are arranged in a highly organized manner and in histological sections appear as eight distinct layers that include three layers of nerve cell bodies and two layers of synapses. The arrangement as seen in conventional histological section is shown in Fig. 1.29 and the ultrastructural appearance is shown in Fig. 1.32A. A simplified schematic diagram of retinal circuitry is shown in Fig. 1.32B. Each of the principal cell types is described briefly later.

Photoreceptors (Fig. 1.33)

There are two types of photoreceptor in the human eye: rods and cones. They are situated on the outer or 'sclerad' aspect of the retina. There are approximately 115 million rods and 6.5 million cones in the human eye. Rods are responsible for sensing contrast, brightness and motion, while cones subserve fine resolution, spatial resolution and colour vision (see Chapter 5). The density of rods and cones varies in different regions of the retina, the periphery being rod-dominated (30,000/mm^2) while cone density increases nearer the macula (150,000/mm^2 at the fovea), the fovea being exclusively cones.

Each photoreceptor consists of a long narrow cell with an inner and outer segment joined by a connecting stalk consisting of a modified cilium (Fig. 1.33). These inner and outer segments are 'separated' from the cell body by the outer limiting membrane. The nucleus is situated in the outer nuclear layer of the retina and axons pass into the outer plexiform layer where they form synaptic terminals (cone pedicle or rod spherule) with bipolar cells and interneurones (horizontal cells) (Fig. 1.32B). The outer segments of rods and cones are shaped precisely as their name implies. They contain the visual pigments that are responsible for absorption of light and initiation of the neuroelectrical impulse.

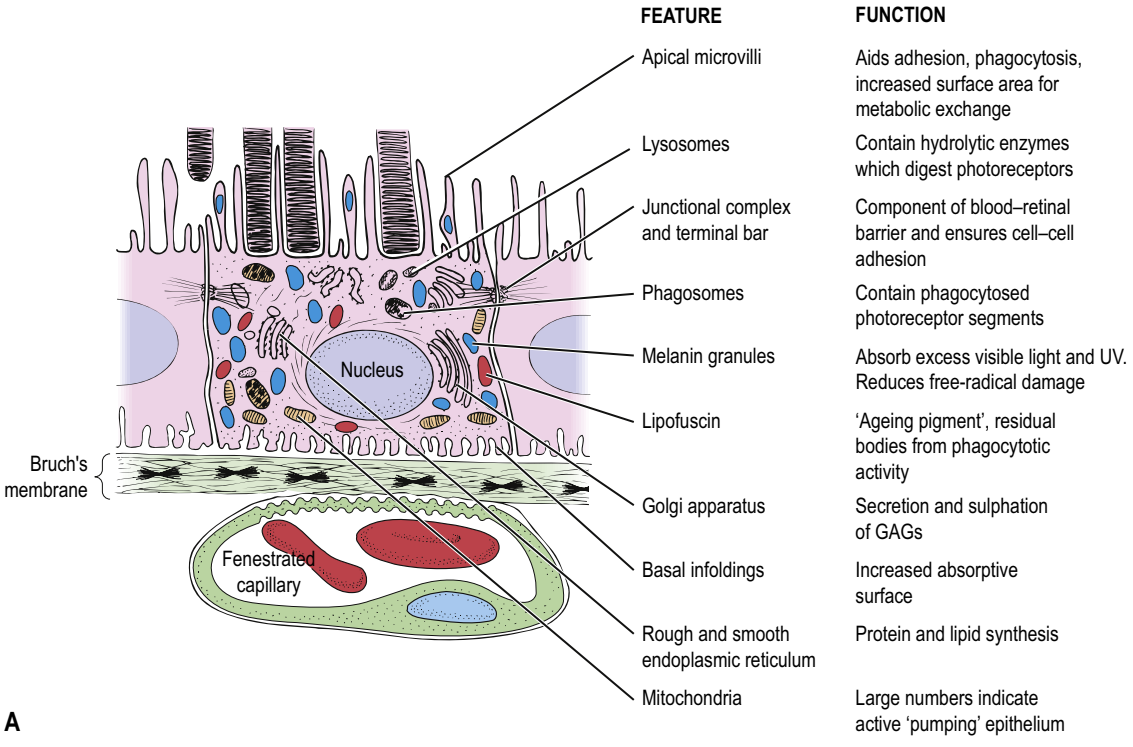


Fig. 1.31 (**A**) Diagram summarizing the main ultrastructural features of the retinal pigment epithelium (RPE). (**B**) Transmission electron micrograph of human RPE layer. (**C**) Scanning electron micrograph of the apical surface of the retinal pigment epithelium. Note the hexagonal shape and the ovoid melanin granules, only visible because of post-mortem-induced disruption of the apical cell membrane. *CC*, Choriocapillaris; *BM*, Bruch's membrane. Original magnifications: **B** × 2600; **C** × 3600. (Fig. 1.31B and C courtesy D. Aitken and W.R. Lee.)

Box 1.18 **Clinical correlates**

Ageing and the retinal pigment epithelium

The RPE is fundamental to photoreceptor function. It is this layer, present since birth, that degenerates with age and is the main site of age-related macular degeneration (AMD). Degenerative changes are variable depending on the site (periphery versus central). In addition, as the RPE becomes atrophic and loses cellularity, it attempts to cover the defects by enlarging and spreading. In the process, multinuclear RPE cells form, with numbers of nuclei ranging from two to several. How this occurs has been assumed to be a result of cell–cell fusion, similar to the process that occurs in other cells such as macrophages. However, recent studies have shown that failure of cytokinesis might be the main mechanism of formation of multinuclear giant RPE cells, i.e. although the nuclei can replicate in the initial stages of cell division and repair, the aged RPE cells cannot complete the process and cytoplasmic separation of two new cells fails.

Also evident with age is an increase in lipofuscin within the RPE, which displays autofluorescent properties.

Rod cells. Rods are long (100–120 μm) slender cells whose outer segment contains the visual pigment rhodopsin sensitive to blue-green light (maximal spectral sensitivity 496 nm). Rods are highly sensitive photoreceptors and are used for vision in dark-dim conditions. The rhodopsin is contained within the membrane-bound lamellae or disks (up to 1000 per cell, 10–15 nm thick) that are enclosed by a single cell membrane (a conceptually useful analogy is to liken them to coins stacked inside a stocking). Each outer segment is only 1–1.5 μm in width (Fig. 1.33A–C,G) and 25 μm in length. The disks are produced at the base of the outer segment (the ciliary connection) and, over the course of 10 days, travel to the tips, which are enclosed by the apical microvilli of the RPE. Here they are phagocytosed by the RPE cells in a circadian manner (predominantly shed in the early morning). The rods are separated by a modified extracellular ground substance known as interphotoreceptor matrix, which contains a 135 kDa glycolipoprotein, interphotoreceptor binding protein (IRBP). The inner half of the inner segment is known as the *myoid*, the outer half the *ellipsoid* (3 μm in length). The ellipsoid is connected to the outer segment by a modified cilium (nine doublet microtubules without a central pair) (Fig. 1.33B) whose basal body is situated in the ellipsoid. The cilium represents the embryological vestige of the ciliated neuroepithelial cells that line the primitive retinal or optic ventricle (see Chapter 2). The cilium acts as a conduit for metabolites and lipids between the inner and outer segments. The remainder of the ellipsoid contains numerous mitochondria, indicative of the high metabolic activity of these cells. The myoid region contains numerous organelles including Golgi apparatus, smooth endoplasmic reticulum, microtubules and glycogen, evidence of a metabolically and synthetically active cell.

Cones (*Fig. 1.33D–F*). In most diurnal animals two spectrally distinct cone types exist (one maximally sensitive to short wavelengths and one to long wavelengths; known as dichromatic retina); however, in diurnal Old World primates, apes and humans, a third type exists (trichromatic retina). The three types are generally referred to as blue, green and red (or the short, medium and long wavelength) sensitive cones. Cone outer segments are generally shorter than rods and are so called because they are generally conical (6 μm at the base, 1.5 μm at the tip); however, in the fovea they are long, slender and tightly packed (see Fig. 1.30C). The lamellae or disks in cones are not surrounded by a plasma membrane in the same manner as rods but are in free communication with the interphotoreceptor space. Cone disks have a greater lifespan than rods and are not produced in the same manner; in addition, they do not undergo circadian phagocytosis by the RPE cells. They are surrounded by the long villous melanin-containing apical processes of the RPE.

Cones are about 60–75 μm in length. The outer segment is connected to the mitochondria-rich ellipsoid region (containing around 600 mitochondria per cell) of the inner segment by a cilium similar to that described in rods (Fig. 1.33F). The cell body of the cone can be easily identified histologically in the sclerad aspect of the outer nuclear layer because of its large pale-staining nucleus and perinuclear cytoplasm (Figs 1.29 and 1.32A).

The cell bodies of rods and cones are connected by an *inner fibre* to specialized expanded synaptic terminals, known as *spherules* and *pedicles*, respectively. These synapse with bipolar and horizontal cells and contain many highly specialized presynaptic vesicles.

Rod spherules lie more sclerad than the cone pedicles and are deeply indented by bipolar and horizontal cell processes (telodendria). A specialized region known as a *synaptic ribbon* is present between two adjacent nerve fibres. The horizontal cell telodendria penetrate deeply into the spherule; the bipolar cell dendrites (from one to four cells) have a shallower penetration. Up to five processes may be embedded in one spherule (Fig. 1.33A). There is no apparent contact between rod spherules; however, cone pedicles may be connected by gap junctions.

Cone pedicles are broader than rod spherules (7–8 μm) and have a pyramidal shape. In the cone pedicle there are up to 12 indentations, each of which contains three neuronal terminals (*triad*) (Fig. 1.33D–E). The central process in each triad is a midget bipolar cell dendrite (each may have multiple contacts with the same pedicle; Fig. 1.32B). The

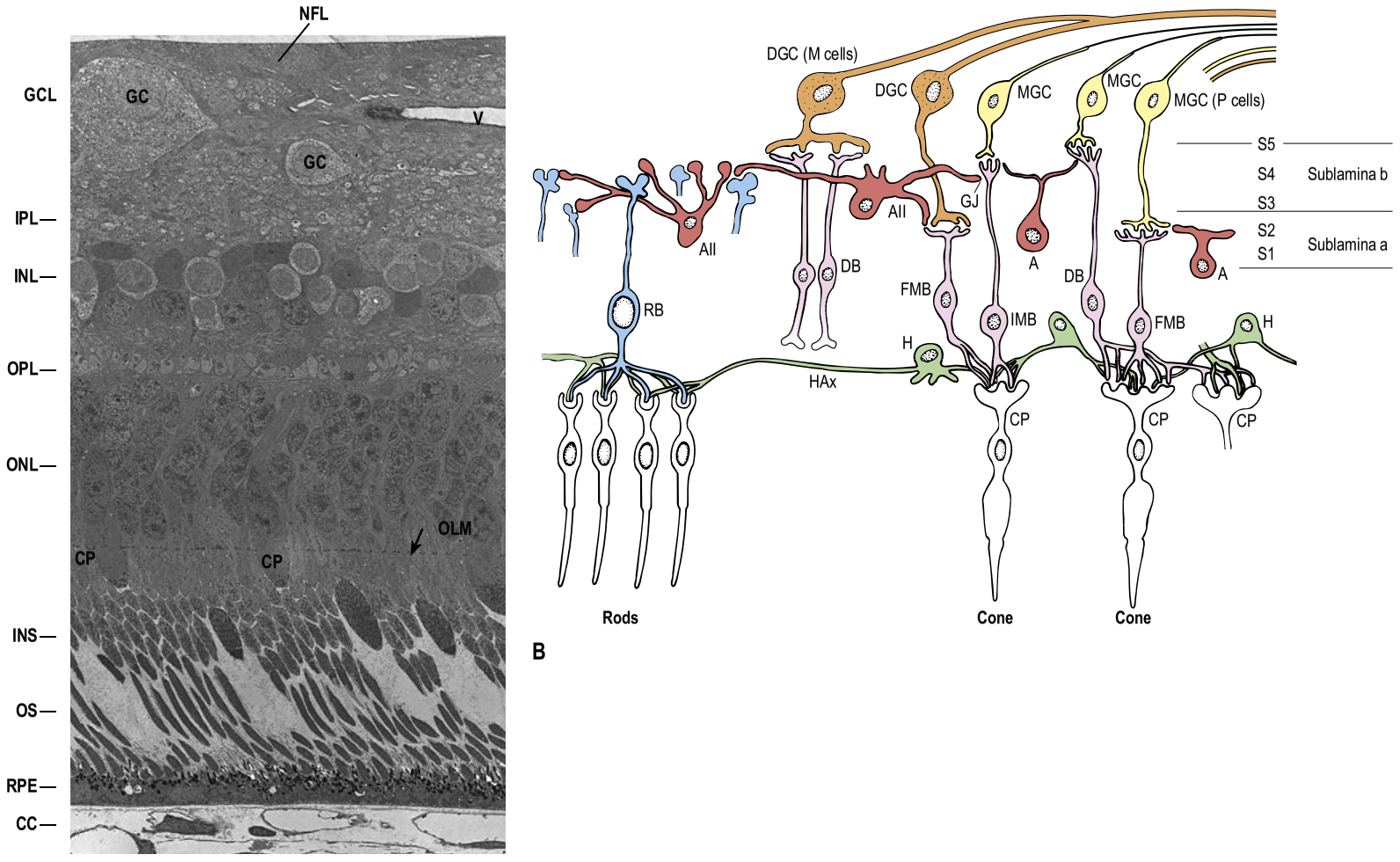


Fig. 1.32 (**A**) Low-power transmission electron micrograph of the primate retina demonstrating the layered arrangement. (**B**) Schematic diagram showing the arrangement and relations of the major cell types in the retina. *A*, Cone amacrine cell; *All*, All (rod) amacrine cell; *CC*, choriocapillaris; *CP*, cone pedicle; *DB*, diffuse bipolar cell; *DGC*, diffuse ganglion cell; *FMB*, flat midget bipolar cell; *GC*, ganglion cell; *GCL*, ganglion cell layer; *GJ*, gap junction; *H*, horizontal cell; *HAx*, horizontal cell axon; *IMB*, invaginating midget bipolar cell; *INL*, inner nuclear layer; *INS*, inner segments; *IPL*, inner plexiform layer; *MGC*, midget ganglion cell; *NFL*, nerve fibre layer; *OLM*, outer limiting membrane; *ONL*, outer nuclear layer; *OPL*, outer plexiform layer; *OS*, outer segment; *RB*, rod bipolar cell; *RPE*, retinal pigment epithelium; *V*, retinal vessel. Original magnification: **A** × 930.

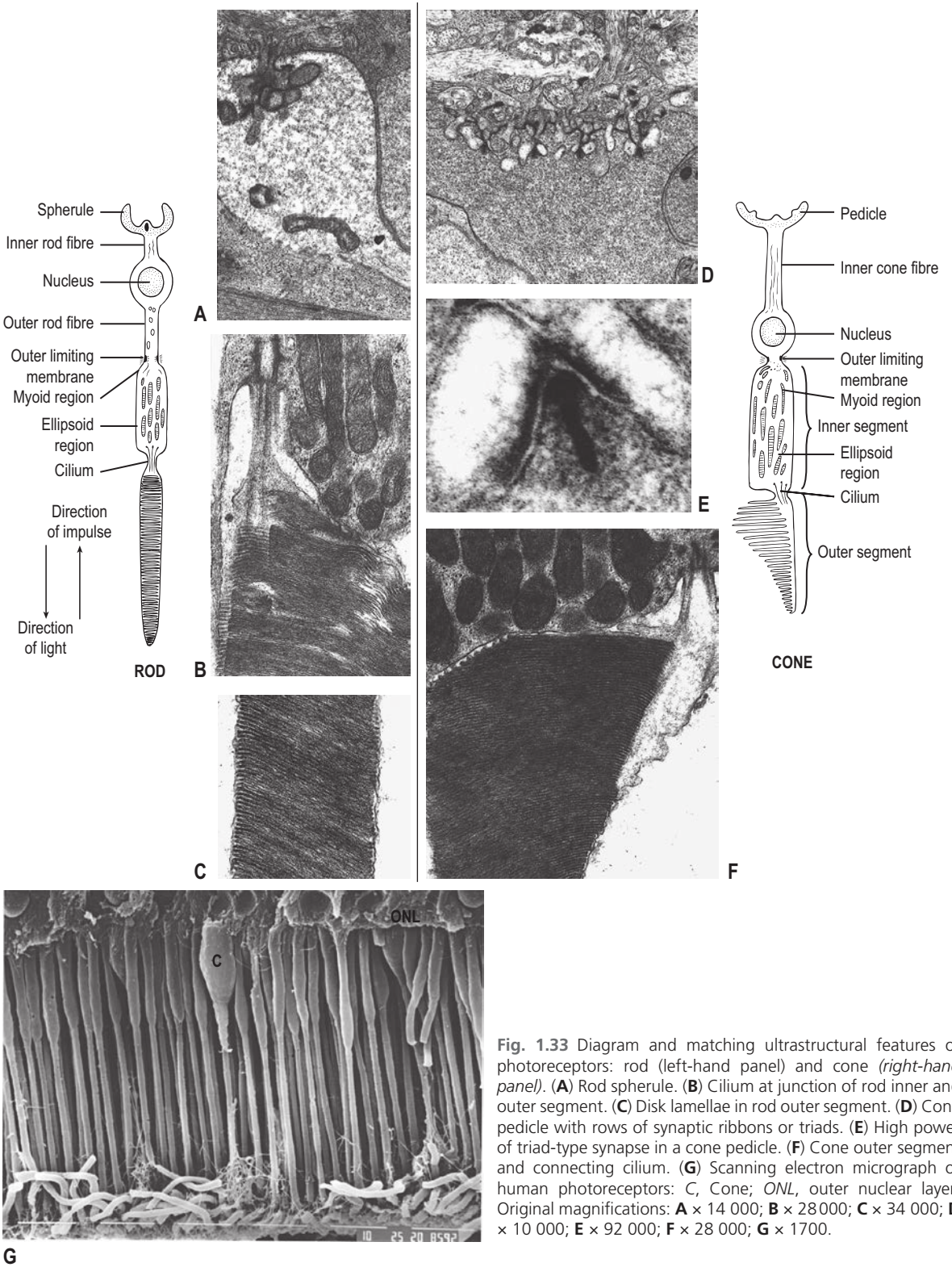


Fig. 1.33 Diagram and matching ultrastructural features of photoreceptors: rod (left-hand panel) and cone *(right-hand panel)*. (**A**) Rod spherule. (**B**) Cilium at junction of rod inner and outer segment. (**C**) Disk lamellae in rod outer segment. (**D**) Cone pedicle with rows of synaptic ribbons or triads. (**E**) High power of triad-type synapse in a cone pedicle. (**F**) Cone outer segment and connecting cilium. (**G**) Scanning electron micrograph of human photoreceptors: *C*, Cone; *ONL*, outer nuclear layer. Original magnifications: **A** × 14 000; **B** × 28 000; **C** × 34 000; **D** × 10 000; **E** × 92 000; **F** × 28 000; **G** × 1700.

laterally disposed processes in the triad are horizontal cell processes that may also be involved in several triads on the one pedicle. Thus, there may be up to 25 synaptic ribbons in each pedicle (Figs 1.32B and 1.33D,E). Each cone is usually contacted by all the horizontal cells (four to six) in the immediate area or field. Each pedicle also has numerous shallow indentations or synapses with flat diffuse bipolar cells (Fig. 1.32B).

Bipolar cells (Figs 1.32B and 1.34A)

The retina contains approximately 35.7 million bipolar cells, which comprise several functional and morphological subtypes. They are primarily responsible for transmitting signals from photoreceptors to ganglion cells, between which they are interposed. Their cell bodies lie in the *inner nuclear layer* and are oriented in a radial fashion parallel to the photoreceptors. Their single or multiple dendrites pass outwards to synapse principally with photoreceptors (but also with horizontal cells), while their single axon passes inwards and synapses with ganglion and amacrine cells. In the foveal region of the central retina the ratio of cones : bipolar cells : ganglion cells can be 1 : 1 : 1, whereas in the peripheral retina one bipolar cell receives stimuli from up to 50–100 rods. In intervening regions, the ratio corresponds to the decreasing visual acuity present in the peripheral retina. This summation of stimuli is a crucial factor in the sensitivity of the rod system to low levels of illumination. Bipolar cells have been subdivided in humans into nine morphological subtypes (Figs 1.32B and 1.34A): one rod bipolar type and eight types of cone bipolars. The latter group can be subdivided into five types of diffuse cone bipolars and three types of midget bipolars (see also Chapter 5 p. 299).

Rod bipolar cells. *Rod bipolar cells* have a receptive field or dendritic tree, which is small in the central retina (15 μm wide, 10–20 rods) and larger in the peripheral retina (30 μm and 30–50 rods). These represent 20% of all bipolar cells and are most dense around the fovea. In the periphery they contact up to 50 rods and synapse with AII amacrine cells; only rarely do they synapse directly with diffuse ganglion cells.

Diffuse cone bipolar cells. *Diffuse cone bipolar cells* are concerned with converging information from many cones. Their dendrites fan out (up to 70–100 μm) to end in clusters of between five and seven cone pedicles (can be as high as 15–20). The overlap of adjacent cells of this type is extensive in the perifoveal region.

Midget bipolar cells. *Invaginating midget bipolar cells* are the smallest of the bipolar cells whose dendrites penetrate the base of a single cone pedicle (occasionally two) to form a central element in triads. The near 1 : 1 ratio of midget bipolar cells to cones decreases peripherally. Their dendrites may synapse with amacrine cells and midget ganglion cells.

Flat midget bipolar cells connect single cones with single midget ganglion cells or more rarely may connect several cones. They are similar to invaginating midget bipolar cells except their dendrites do not invaginate deeply into the cone pedicle. Potentially, therefore, most cones can be in contact with these two types of midget bipolar cells (flat and invaginating midget) as well as the diffuse type.

Blue cone-specific bipolar cells (blue S-cone) are present in primates and humans and appear to make invaginating contact with only a limited number of cones in their territory, the suggestion being that these are specifically blue cones.

Ganglion cells

The cell bodies of most ganglion cells are located in the innermost nucleated layer of the retina (ganglion cell layer) situated between the nerve fibre layer and the inner plexiform layer (Figs 1.29 and 1.32); however, 'displaced' ganglion cells have been identified in the inner nuclear layer. Ganglion cells are the last neuronal link in the retinal component of the visual pathway (Fig. 1.32B). Their axons form the nerve fibre layer on the innermost surface of the retina and synapse with cells in the lateral geniculate nucleus of the thalamus. The axons form bundles separated and ensheathed by glial cells (Fig. 1.34E). The bundles leave the eye to form the optic nerve. Upon exiting through the *lamina cribrosa*, the axons become myelinated with oligodendrocytes. There are up to seven layers of ganglion cell bodies in the central retina or fovea (ganglion cell layer is 60–80 μm thick) and as few as one cell layer in the peripheral retina (10–20 μm thick). There are approximately 1.2 million ganglion cells per retina; thus, theoretically, there are approximately 100 rods and four to six cones per ganglion cell. While they are functionally diverse, ganglion cells are characterized morphologically by a large cell body, abundant Nissl substance (arrays of rough endoplasmic reticulum) and a large Golgi apparatus (Fig. 1.34E). They are classified into different types on the basis of cell body size, dendritic tree spread, branching pattern and branching level in the five strata of the inner plexiform layer (Fig. 1.32B,C) (see also Chapter 5 p. 299) . Some ganglion cells in the macular area may contain yellow (xanthophyll carotenoid) pigment in the cytoplasm, although the cone axons and Müller cells are thought to also contain these pigments in the macular region. Impulses are received primarily from bipolar cells and amacrine cells via axodendritic and axosomatic synapses, the former occurring predominantly in the inner plexiform layer (Fig. 1.32B) where their dendrites repeatedly branch to form the 'dendritic tree', whose form and size vary considerably and may be correlated with

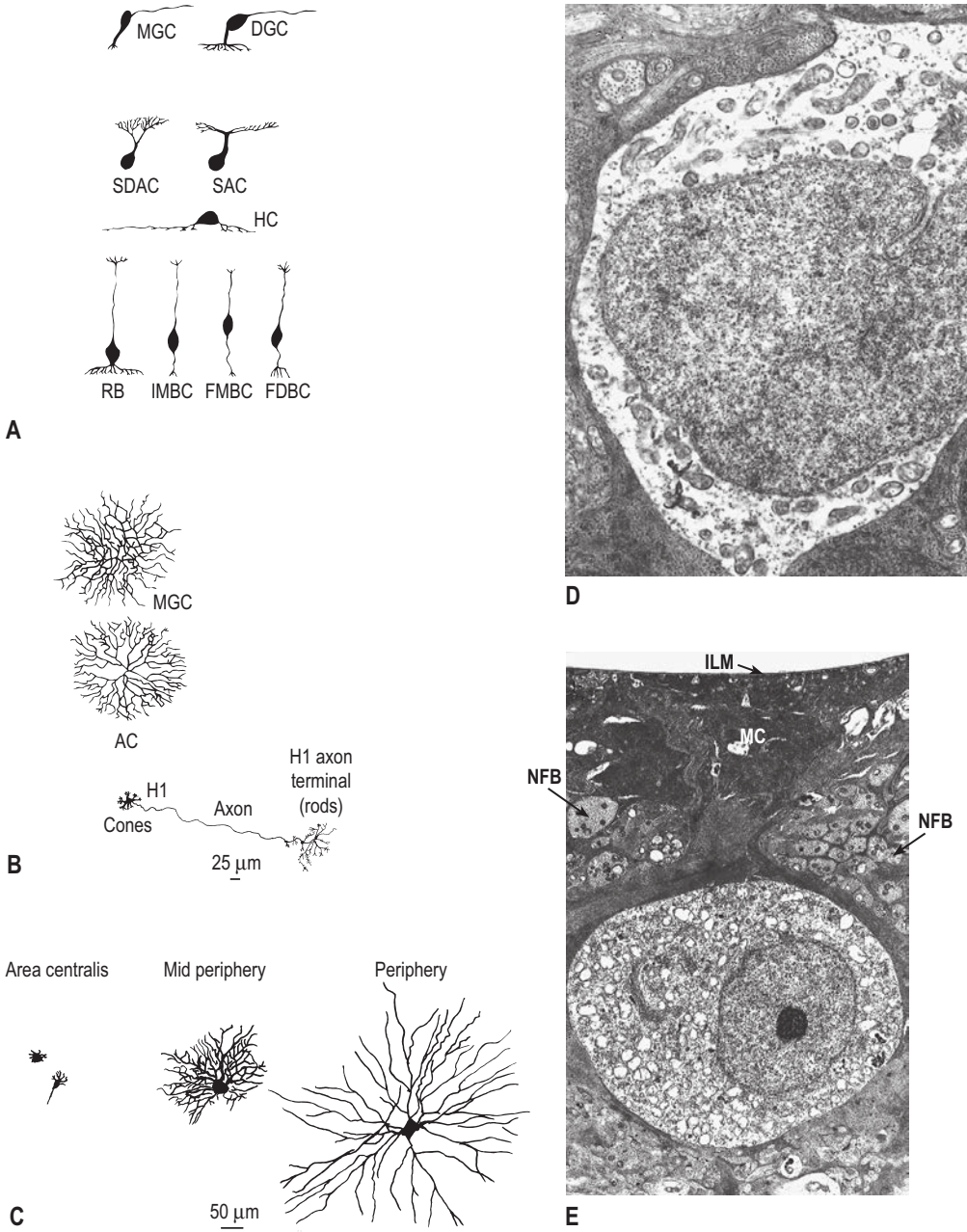


Fig. 1.34 (**A**) Diagram showing the arrangement and location of major neuronal cell types as seen in Golgi-stained preparations in relation to their position in the layers of the retina. (**B**) Shape of dendritic fields of midget ganglion cell *(MGC)*, amacrine cell *(AC)* and H1 horizontal cell *(HC)* as seen in Golgi preparations or single cell injections in retinal whole mounts. (**C**) Cat ganglion cells increasing in dendritic tree span with increasing eccentricity from the fovea. (**D**) Electron micrograph of amacrine cell. (**E**) Ganglion cell body and adjacent nerve fibre bundles (*NFB*). *DGC*, Diffuse ganglion cell; *FDBC*, flat diffuse bipolar cell; *FMBC*, flat midget bipolar cell: *HC*, horizontal cell; *ILM*, inner limiting membrane; *IMBC*, invaginating midget bipolar cell; *MC*, Müller cell; *MGC*,
midget ganglion cell; *RB*, rod bipolar; *SAC*, stratified amacrine cell; *SDAC*, small diffuse amacrine cell. Original magnifications: **D** × 7100; **E** × 4200. (Fig. 1.34B and C, after Kolb et al., 1992, 1994.)

location in the retina and therefore function (receptive field size) (Fig. 1.34B,C). The morphological diversity of ganglion cells (up to 25 types in mammalian and 18 types in human retinas) has classically categorized ganglion cells into α, β and γ, or X, Y and W types, predominantly based on research in cats.

Recently a non-rod, non-cone photoreceptive pathway, arising from a population of retinal ganglion cells, was discovered first in nocturnal rodents and then in primates. These ganglion cells express the photopigment melanopsin and by signalling gross changes in light intensity serve the subconscious, 'non-image-forming' functions of circadian photoentrainment and pupil constriction via the suprachiasmatic nucleus in the brain (see Chapter 5, p. 282). The primate retina, in addition to being intrinsically photosensitive, is strongly activated by rods and cones to signal irradiance over the full dynamic range of human vision. Thus, in the diurnal trichromatic primate, 'non-image-forming' and conventional 'image-forming' retinal pathways are merged, and the melanopsin-based signal might contribute to conscious visual perception.

Midget ganglion cells. These cells synapse exclusively with amacrine cells and one midget bipolar cell (and thus usually one cone) (Fig. 1.32B). Dendritic spread is around 5–10 μm in diameter in the central retina; however, this increases 10-fold in a zone of 2–6 mm eccentricity and attains a maximum of over 100 μm (Fig. 1.34C). Neighbouring midget ganglion cell dendritic fields do not overlap but form mosaics. In humans they are also known as P-cells because they project to the parvocellular layer of the lateral geniculate nucleus (LGN).

Diffuse (parasol) ganglion cells. These comprise a large synaptic field with all types of bipolar cells except midget bipolar cells. They occur in the central retina and their cell bodies (soma) are 8–16 μm in diameter with 30–70 μm dendritic fields, these being smaller nearer the fovea than the periphery. They are also known as M-cells because they project to the magnocellular layer of the LGN.

The finding that midget ganglion cells synapse exclusively with the midget bipolar cells, and that both are common near the fovea, provides the anatomical basis for the observation of small receptive fields and high visual acuity in this region. There are five types of diffuse ganglion cell, classified on the basis of morphology. The anatomical basis of antagonistic fields surrounding receptive fields is complex, although they do not appear to vary much in size from within an 8 mm radius of the fovea. The basis of the antagonist field may be the lateral extensions of the amacrine cell, with its extensive interconnections with ganglion cell dendrites and bipolar cells as well as fellow amacrine cells.

Association neurones (amacrine and horizontal cells) (Figs 1.32B and 1.34A,B)

Horizontal cells. These cells derive their name from the extensive horizontal extensions of their cell processes. There are two distinct morphological varieties in the retina of most species, of which the cat is the most extensively studied: type A is a large sturdy axonless cell with stout dendrites that contact only cones; type B has a smaller bushier dendritic tree that contacts cones exclusively but, in addition, has an axon up to 300 μm in length that ends in extensive arborization that is postsynaptic only to rods (Figs 1.32B and 1.34B). Type A cells have much larger receptive fields than type B. In primates it appears that the two types of horizontal cell, HI (approximates to type B) and HII (approximates to type A), both possess axons. A third type (HIII) has been described in the human retina. Each rod has connections with at least two horizontal cells and each cone with three or four horizontal cells of each type. In primates the stout dendrites of HI cell soma processes contact around seven cones near the fovea (dendritic tree covering 15 μm); this number increases to as many as 18 further from the fovea (dendritic tree covering 80–100 μm). The axon from HI cells passes laterally and terminates up to 1 mm away in a thickened axon terminal bearing a fan-shaped protrusion of lollipop-like endings in rod spherules (up to 100) (Fig. 1.34B). HII dendritic trees are more spidery and contact about twice as many cones. Their axons are generally shorter (100–200 μm) and contact cone pedicles by small wispy terminals. The manner of their insertion is depicted in Fig. 1.32B. Their cell bodies are located primarily in the outer part of the inner nuclear layer. They have few distinctive cytoplasmic organelles except the crystalloids, a series of densely stacked tubules with associated ribosomes. Their processes ramify in the outer plexiform layer close to the cone pedicles. The overlap between horizontal cells is considerable and any one area of retina may be served by up to 20 horizontal cells. Horizontal cells have an integrative role in retinal processing and release inhibitory neurotransmitters, mainly γ-aminobutyric acid (GABA). Recent evidence suggests that there is some colour-specific wiring for the three types of horizontal cells in the human retina.

Amacrine cells (Figs 1.32B and 1.34D). These association neurones were thought to lack axons; however, recent studies have shown that some do indeed possess an axon. They are located in the vitread or inner aspect of the inner nuclear layer (bipolar cell layer) and are distinguishable as a result of their larger size (12 μm) and oval shape. They display a remarkable degree of diversity. There are at least 25 different types in the monkey and human retina. Their cell body is usually flask-shaped and the numerous dendritic processes of these cells ramify and terminate predominantly in the synaptic complexes formed by the bipolar

and ganglion cell processes, namely the inner plexiform layer. The shape of their dendritic fields is highly variable and a few examples are shown in Fig. 1.34A,B. They can be divided into subtypes on several criteria such as the stratification of their dendrites in the inner plexiform layer or their shape, for example diffuse, starburst and stratified. Diffuse types can cover narrow fields (approximately 25 μm wide), their fibres being cone-shaped. Other types may spread their axon-like processes several millimetres. They may also be classified on the basis of their neurotransmitters. Amacrine cells may be GABAergic and dopaminergic or can release acetylcholine indicating, together with their morphology, that these cells play a role in modulation (most probably inhibitory) of signals reaching ganglion cells. A subclass of amacrine cells are also thought to be the principal source of the peptide somatostatin, an important neuroactive peptide, in the retina. It may function as a neurotransmitter, neuromodulator or trophic factor.

Retinal neuroglia

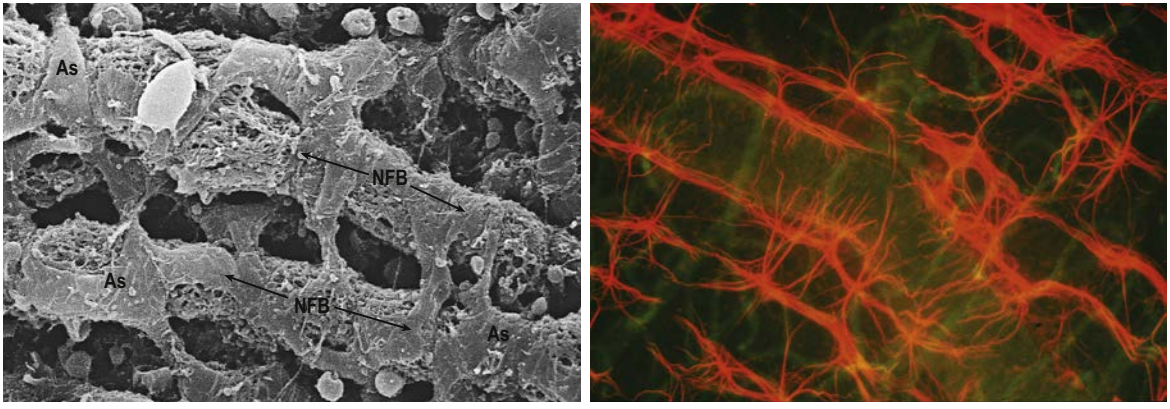
Astrocytes. Astrocytes are not the principal or predominant glial cell in the retina. This role is fulfilled by Müller cells, which are analogous to central nervous system oligodendrocytes. Astrocytes are predominantly located in the nerve fibre layer, ganglion cell layer, inner plexiform layer (site of cell bodies) and their outer limit is the vitread aspect of the inner nuclear layer in humans. They form an irregular honeycomb scaffold between vessels and neurones perpendicular to the Müller cells. They may occur as fibrous (elongated) or protoplasmic (rounded) astrocytes. They both contain abundant cytoplasmic structural fibrils (10 nm in diameter) consisting of glial fibrillary acid protein (GFAP) (Fig. 1.35B) which becomes highly expressed where the cells are stressed (Box 1.19 Clinical correlates). Astrocytes are often oriented perpendicular to the direction of the neurone cell bodies or processes, such as in the nerve fibre layer (Fig. 1.35). It seems their role may be to isolate the receptive surfaces of neurones in the retina, thus preventing unwanted signals or effects in neighbouring neurones. They have abundant intracytoplasmic glycogen and form 'gap junctions' with neighbouring astrocytes. Their pedicles or foot processes were once believed to constitute an important functional component of the blood–retinal barrier (see later).

Müller cells. Müller cells (Fig. 1.36A,B,) are the principal supporting glial cells of the retina and are considered analogous to central nervous system radial glial or ependymal cells. They have a radial orientation and extend through the depths of the retina from the inner surface, where their expanded 'foot process' lies adjacent to the inner limiting membrane (ILM), to their outer limit where they have

Box 1.19 **Clinical correlates**

Proliferative vitreal retinopathy

When injured, the retina frequently responds by forming astroglial scars. Indeed, normal age-related degenerative processes in the peripheral retina (microcystoid degeneration) are accompanied by astrocyte proliferation. Disruption of the inner limiting membrane can lead to astrocyte proliferation in the subhyaloid space and in the vitreous itself.



A **B**

Fig. 1.35 (**A**) Scanning electron micrograph (viewed from the vitreous aspect) of astrocytes *(As)* surrounding nerve fibre bundles *(NFB)* in the inner retina (the inner limiting membrane has been removed to expose the underlying nerve fibre layer). (**B**) Double-colour immunofluorescence illustrating the relations of astrocytes shown with an antibody to glial acidic fibrillary protein (GFAP) *(red)* and lectin-stained vessels *(pale green)*. Original magnifications: **A** × 1500; **B** × 150. (Fig. 1.35B courtesy Dr T. Chan-Ling.)

ALGrawany

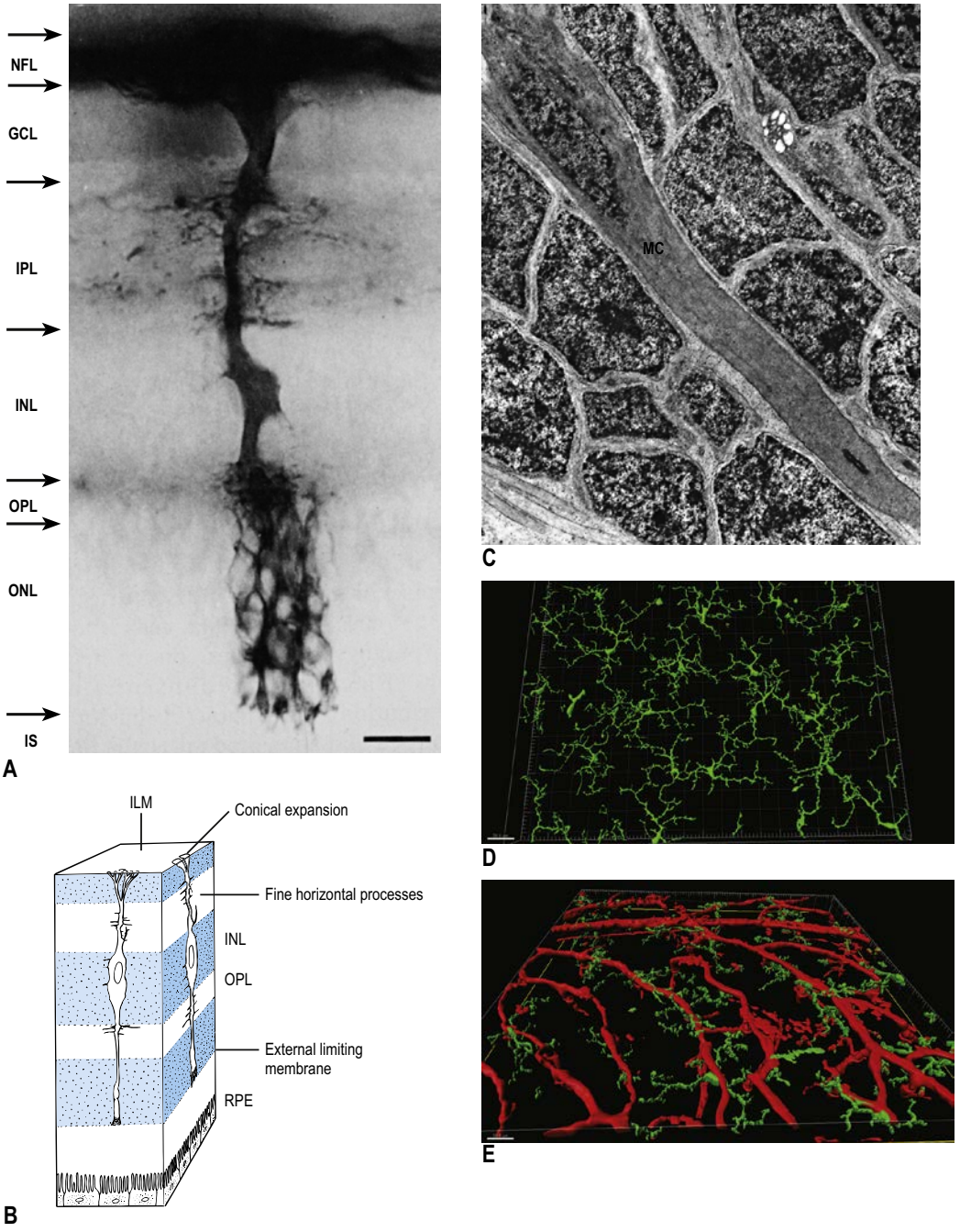


Fig. 1.36 (**A**) Micrograph of a horseradish peroxidase (HRP)-filled Müller cell in the rabbit retina. The dark band at the top of the micrograph is composed of Müller cell endfeet and the labelled axons of ganglion cells in the nerve fibre layer. The Müller cells possess side-processes that form different strata in the inner plexiform and they send numerous processes to wrap around the somata of photoreceptors. (**B**) Diagram of the shape and position of a Müller cell. (**C**) Morphology of a Müller cell *(MC)* within the outer nuclear layer. Note the intracytoplasmic microfilaments. (**D**) Microglial cells (specialized macrophages) in a retinal whole mount from a transgenic mouse in which eGFP is expressed alongside the locus for the chemokine receptor CX_3CR1. All microglia in these animals express CX_3CR1 and thus appear fluorescent *green* in confocal microscopy. (**E**) 3D-rendered retinal microglial network in a retinal wholemount in which retinal vessels have been highlighted *(red)* by perfusion of a vascular dye. *GCL*, Ganglion cell layer; *INL*, inner nuclear layer; *IPL*, inner plexiform layer; *IS*, inner segments; *NFL*, nerve fibre layer; *ONL*, outer nuclear layer; *OPL*, outer plexiform layer; *RPE*, retinal pigment epithelium. Original magnifications: **A** bar 10 mm; **B** × 4400; **C** × 200. (Fig. 1.36A courtesy Dr S. Robinson.)

adherens junctions with photoreceptor inner segments to form the external limiting membrane (ELM). Thus, Müller cells provide a trellis-basket-like support for the retina, with the photoreceptors extending the 'business end' of segments through the gaps in the ELM (Fig. 1.36C). They envelop blood vessels, neuronal cell bodies and processes, creating glial 'tunnels' via a series of cytoplasmic processes, as shown in Fig. 1.36A,B. Müller cells in humans contain little glycogen, in contrast to species with avascular retinae. Their cytoplasm contains abundant endoplasmic reticulum and microtubules, reflecting their role in protein synthesis, intracellular transport and secretion (Fig. 1.36C). These cells may help to nourish and maintain the outer retina, which lacks a direct blood supply.

While there is extensive coupling between astrocytes and Müller cells, which allows the exchange of tracer molecules, recent studies have demonstrated an absence of significant spread of spatial buffer current between retinal glial cells. Both astrocytes and Müller cells have high K^+ membrane conductance and most spatial buffer current will flow out through these conductances rather than spreading into neighbouring glial cells through gap junctions. In contrast to electrical coupling, chemical coupling between astrocytes is sufficiently strong to mediate propagation of intercellular signals such as the spread of metabolites and ions between glial cells. Coupling between glial cells, therefore, could serve to enhance the transport of key metabolites, such as glutamate, glutamine and lactate, both into and out of glial cells, by allowing them to diffuse between neighbouring cells in the glial syncytium. Clearly the underlying arrangement of both astrocytes and Müller cells reflects this function. Müller cells have also been described as acting like fibre-optic cables permitting the rapid transmission of photons of light to the photoreceptor outer segments, particular in the fovea/macular area where they increase the efficiency in light capture.

Microglia. Microglia (Fig. 1.36D,E) are a highly specialized subpopulation of tissue macrophages that serve as immune sentinels in the parenchyma of the central nervous system parenchyma. They are becoming more widely appreciated as playing a role in synapse remodelling (synaptic pruning). These cells most likely arise from yolk sac erythromyeloid precursors during early development and may be self-maintained or replenished in adulthood by bone marrow-derived circulating monocytes and it is often difficult to distinguish the differences in the two populations. They are characterized by an extremely arborized morphology and an immunophenotype of resting macrophages. In the retina, their cell bodies are located largely in three strata, one at the nerve fibre layer–ganglion cell layer interface, one in the inner nuclear layer and another in the outer plexiform layer, although the latter is more obvious in the rodent than human retina. Their processes form a lateral and vertical three-dimensional network within the retina extending only as far as the outer limiting membrane in the normal eye. Less arborized subtypes, sometimes referred to as perivascular macrophages, which closely resemble homologous cells in the parenchyma of the brain, are associated with the perivascular space of retinal capillaries (see Fig. 1.38C), although they are less numerous than brain perivascular macrophages. Retinal microglia share many properties with brain microglia including tissue homeostasis and host defence. Their highly arborized processes are constantly on the move, sampling their immediate microenvironment. Activated microglia play a role as immune effectors, via the release of chemokines and cytokines. Upon injury to the retina these cells become activated and assume the role of wandering phagocytes and in murine models of disease such as photoreceptor degeneration, retinitis pigmentosa, diabetes, systemic viral infection, or indeed ageing, microglia will migrate into the subretinal space where they may play a role in protecting RPE integrity (Box 1.20 Clinical correlates). Recent data suggest that there is some heterogeneity in the retinal microglia populations with those in the inner plexiform layers being IL-34 dependent with a role in synaptic function relating to colour vision whereas those in the outer retina are IL-34 independent.

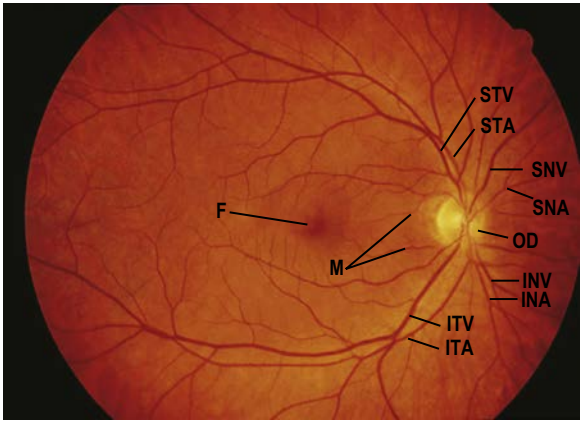
Box 1.20 **Clinical correlates**

Ageing changes in macrophage populations in the retina

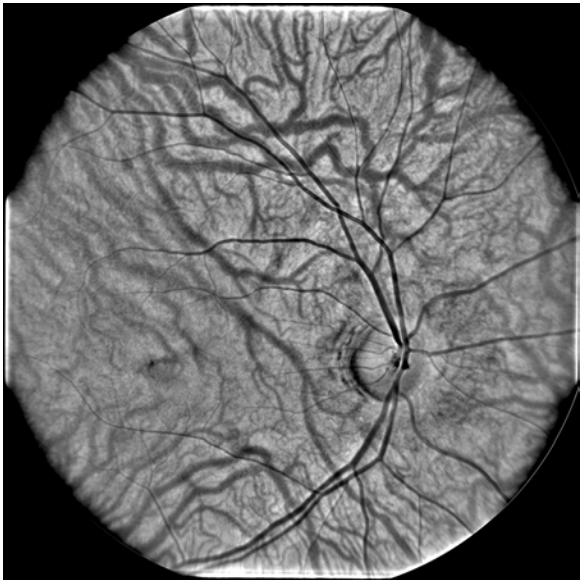
With age there is an increasing tendency in animals for microglia to assume a more amoeboid form and activated phenotype and migrate towards the subretinal space. In humans this is evident especially in the peripheral retina close to age-related cystoid retinal degeneration.

Blood supply of the retina (Figs 1.30A and 1.37)

The retina is an extremely metabolically active sheet of neural tissue with the highest oxygen consumption (per weight) of any human tissue. Like the brain, the retina has a highly selective blood–tissue barrier, which serves primarily to regulate the optimal extracellular environment to facilitate neural transmission. It also regulates the passage of pathogens and intravascular leucocytes, thus partly protecting the neural environment from 'surveillance' by immune cells. In humans, the retina has a dual blood supply (holangiotic), the inner two-thirds being nourished by branches from the central retinal vessels, while the outer one-third is nourished



A



B

Fig. 1.37 (**A**) Wide-field photograph of the normal human fundus. (**B**) Multispectral digital ophthalmoscopic image of retinal and choroidal circulations. This new technique captures high-resolution image data through the retinal and subretinal layers and hence shows the larger vessels of the choroidal circulation. It both expands the examination wavelength range to include image data from invisible wavelengths of light and generates the probe wavelengths to separate specific spectral regions for enhanced visibility and discrimination. No intravascular contrast is used in this method of visualizing the fundus. *F*, Fovea; *INV* and *INA*, inferior nasal vein and artery; *ITV* and *ITA*, inferior temporal vein and artery; *M*, macular vessels; *OD*, optic disk; *SNV* and *SNA*, superior nasal vein and artery; *STV* and *STA*, superior temporal vein and artery. (Fig. 1.37A courtesy C. Barry; Fig. 1.37B, courtesy Annedis, Canada.)

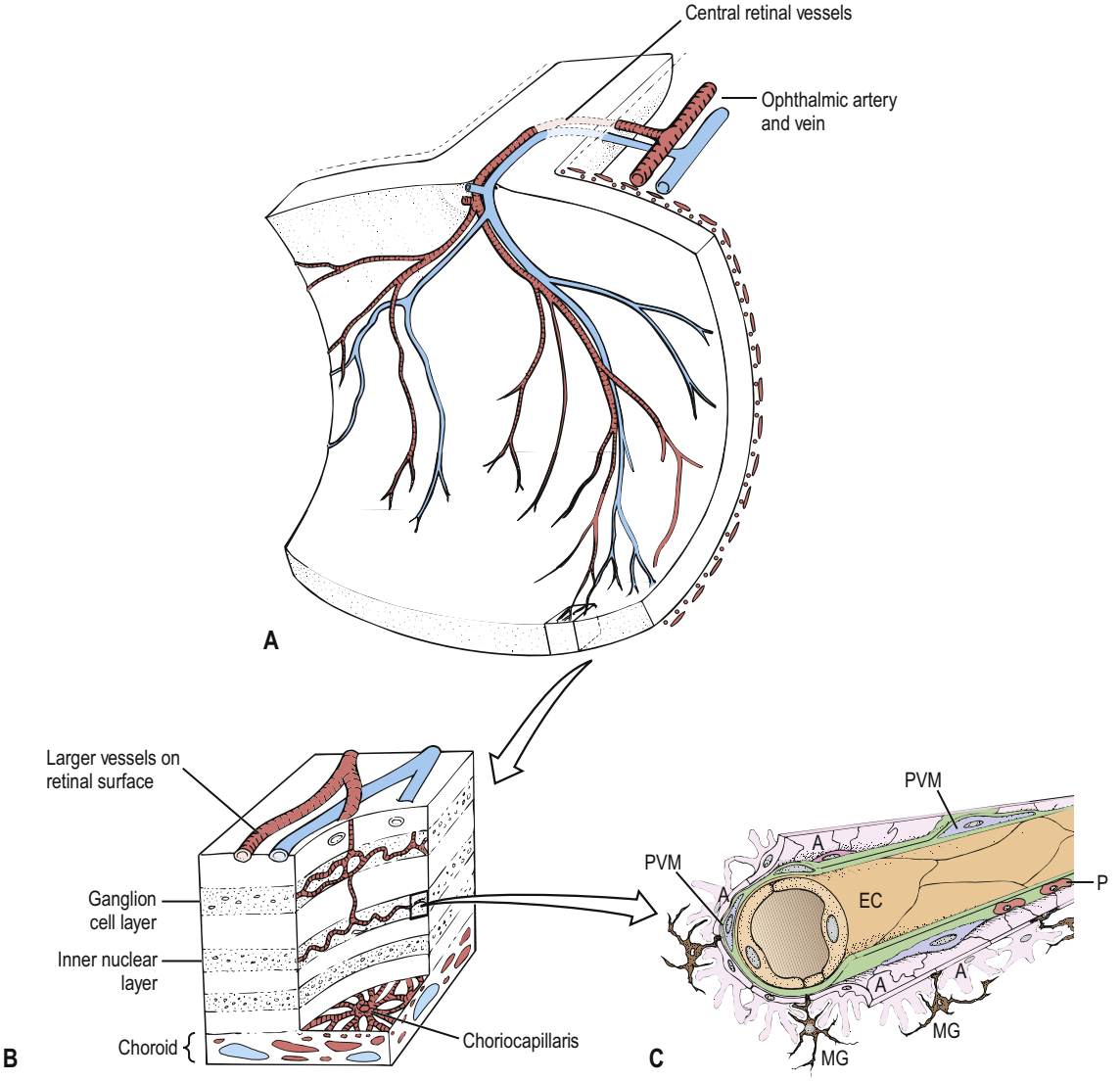
Box 1.21 **Clinical correlates**

Central retinal artery occlusion

This condition is a vivid reminder of the functional end-artery status of the retinal blood supply. In complete central retinal artery occlusion, irreversible changes occur after 1–2 hours and the inner retina becomes white and oedematous except at the fovea, which survives owing to the underlying choroidal circulation, which shows through as a round red patch.

by the choroidal circulation. The choroidal circulation has a high flow rate (150 mm/s), low oxygen exchange and a fenestrated capillary bed; the retinal circulation has a low flow rate (25 mm/s) and high oxygen exchange. The blood–retinal barrier is defined by two sets of characteristics. The first is the structural character of the endothelial and RPE cells located at the endothelial and RPE tight junctions, and the second is the membrane-associated transport characteristics.

Central retinal artery and branches (Figs 1.30A, 1.37 and 1.38). This vessel (0.3 mm in diameter) arises from the ophthalmic artery either in the optic canal or close to the optic foramen where the ophthalmic artery lies bound to the dural covering of the nerve (see p. 94). The central retinal artery then travels forward on the undersurface of the nerve within its dural covering. About 1–1.5 cm behind the eye it pierces the inferomedial aspect of the remainder of meningeal coverings to pass through the subarachnoid and then pierces the nerve. As it passes forward in the centre of the optic nerve, the artery is accompanied by the central retinal vein and a few sympathetic fibres. It resembles other muscular arteries and indeed is affected by conditions such as atheroma and giant cell arteritis. It pierces the papilla centrally, having passed through a constriction or gap in the lamina cribrosa (see, page 59 Fig. 1.42E). This is a potential site for partial or complete occlusive disease (Box 1.21 Clinical correlates). It branches into superior and inferior branches, which subdivide into nasal and temporal arteries, a pattern best appreciated and investigated clinically by fluorescein angiography. A small vessel, the cilioretinal artery, may be present near the optic nerve head and provide a small anastomotic connection between the choroidal and retinal circulations. The central retinal artery diameter decreases to 100 μm upon emerging from the disk. The large retinal arterial branches travel in the nerve fibre layer beneath the inner limiting membrane. These vessels have no internal elastic lamina (which is lost at the optic disk) and are thus not affected in temporal arteritis. They possess a well-developed muscularis, and numerous pericytes lie within the endothelial basal lamina. Each of its four major branches (Figs 1.30A and 1.37) supplies a sector of the retina between which there



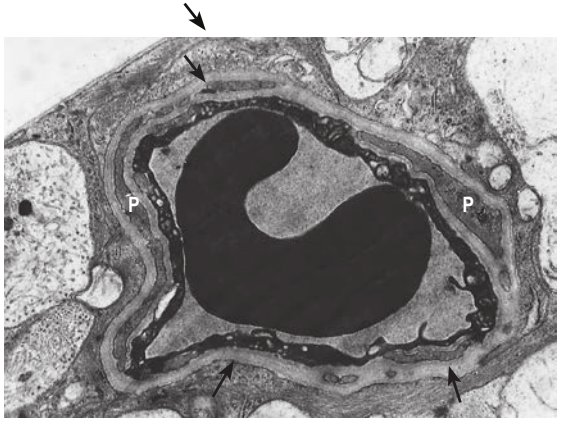


Fig. 1.38 Diagrams of the retinal blood supply. (**A**) The diagram illustrates the manner in which the central retinal artery obtains access to the optic nerve after branching from the ophthalmic artery. (**B**) The levels of the retinal capillary networks. (**C**) High-power diagram illustrating the components of a retinal (or brain) vessel wall that contribute to the blood–retinal (or blood–brain) barrier. From the lumen outwards these are the vascular endothelial cells *(*emissary canal [*EC*] *pale orange)*, basal lamina *(green)* and the glia limitans composed of astrocyte foot processes (**A**). Note the position of the perivascular microglia *(MG)*, perivascular macrophages *(PVM)* and pericytes *(P)*. (**D**) Electron micrograph of a retinal capillary. Note the thickened basal lamina *(arrows)* and pericyte processes *(P)* surrounding the endothelial cell. Original magnification: × 6000. (Fig. 1.38C from Forrester JV, McMenamin PG, Dando SJ. (2018). CNS infection and immune privilege. Nat Rev Neurosci. 2018;11:655–671.)

Box 1.22 **Clinical correlates**

Peripapillary capillaries

Besides the capillaries described in the main text, a further lamina fans out over the nerve fibre layer in the peripapillary region. This unique radial capillary network may be more vulnerable to raised intraocular pressure in glaucoma because of the long course of these vessels (over 1000 μm), infrequent arterial input and lack of anastomoses. Flame-shaped haemorrhages (caused by hypertension or papilloedema) or cotton-wool spots (in ischaemic disease) occur predominantly in this unusual capillary network.

is no overlap, i.e. they are *functional end-arteries*. The superior and inferior temporal arteries curve above and below the macula and foveal region. Arteries pass over veins and may in some pathological situations cause 'nipping' or narrowing of the veins. There are two main levels of capillary networks, which spread like a vast cobweb throughout the retina (Fig. 1.38B). The inner plexus is situated at the level of the ganglion cell layer and the outer plexus at the level of the inner nuclear layer. The concept of these two laminae is not universally accepted, although patterns of vascular disease support the concept (see Chapter 9). There may be up to four layers of capillaries in the peripapillary zone and single layers in the perifoveal region and at the ora serrata.

In humans, retinal capillaries pass only as far as the sclerad margin of the inner nuclear layer, the outer retina being normally avascular. Capillaries are most dense in the macula but are absent from the fovea itself (capillary-free zone 500 μm in diameter), which is thus dependent on the choriocapillaris for nutritional support. Larger arterioles are surrounded by a capillary-free zone. Capillary network density decreases towards the peripheral retina. A small group of vessels in the peripapillary region may be more important than others in understanding glaucoma (Box 1.22 Clinical correlates).

Retinal capillaries are characterized by complete circumferentially oriented endothelial cells joined by non-leaky tight junctions (zonulae occludentes); however, the high number of endocytotic vesicles suggests that they are more permeable than brain capillaries (Fig. 1.38C,D). They are surrounded by a thick basal lamina, pericytes and astrocyte foot processes (Fig. 1.38C), which are four times more numerous around retinal vessels than brain capillaries and may act as a second front in the blood–retinal barrier and thus compensate for the more permeable nature of retinal vascular endothelium. The numbers of these supportive cells decrease in diabetes, macroglobulinaemia and other ischaemic diseases (see Chapter 9, p. 509)

The luminal diameter of retinal capillaries (3.5–6 μm) is somewhat smaller than that of conventional capillaries and is just about large enough to allow white cells to squeeze through. There is very little extravascular connective tissue around retinal vessels. Mast cells, a common perivascular element in other tissues, including the choroid, are absent in the retina, which has a high threshold of tolerance to histamine. There are *no lymphatic* vessels in the retina.

The choroid

The choroid (Fig. 1.39A–G) is the posterior portion of the middle vascular coat of the eye, the *uveal tract*. It is homologous to the pia–arachnoid of the brain. The choroid is a thin, highly pigmented, vascular, loose connective tissue situated between the sclera and the retina, whose principal function is to nourish the outer layers of the retina. It also acts as a conduit for vessels travelling to other parts of the eye and may also have a thermoregulatory role. Furthermore, absorption of light by choroidal pigment aids vision by preventing unwanted light from reflecting back through the retina. In some nocturnal mammalian species, a specialized reflective layer known as a tapetum helps increase the capture of photons in dark conditions (Box 1.23 Comparative anatomy). The regulation of blood flow in the choroid may also influence intraocular pressure by affecting perfusion rates of the ciliary processes.

The choroid extends from the optic nerve margins to the ciliary body and, although its thickness is probably dependent on blood flow dynamics and has a diurnal variation, it is quoted as being approximately 220 μm at the posterior pole and 100 μm anteriorly. Its inner surface is smooth and forms part of Bruch's membrane beneath the RPE. The outer surface, the suprachoroid, is irregular and firmly attached to the lamina fusca of the sclera. Histologically the human choroid consists of Bruch's membrane, the choriocapillaris, a vascular layer and the suprachoroid (Fig. 1.39C).

Bruch's membrane (lamina vitrea)

This modified connective tissue layer is 2–4 μm thick and histologically appears as an acellular glassy membrane beneath the RPE (Figs 1.31A,B and 1.39F,G). Bruch's membrane consists of five layers: the RPE basal lamina (0.3 μm thick) (not truly part of the choroid); an inner collagenous zone; a middle elastic layer (incomplete interwoven bands or perforated sheets of elastic 'fibres'); an outer collagenous zone (which blends with the stroma between the choriocapillaris); and the basement membrane of the endothelial cells in the choriocapillaris. Age-related changes in Bruch's membrane lead to areas of diffuse or discrete thickening known as drusen (see Chapter 9, p. 518).

Choriocapillaris

This is an extraordinarily rich bed of wide-bore fenestrated capillaries that extends only as far anteriorly as the ora

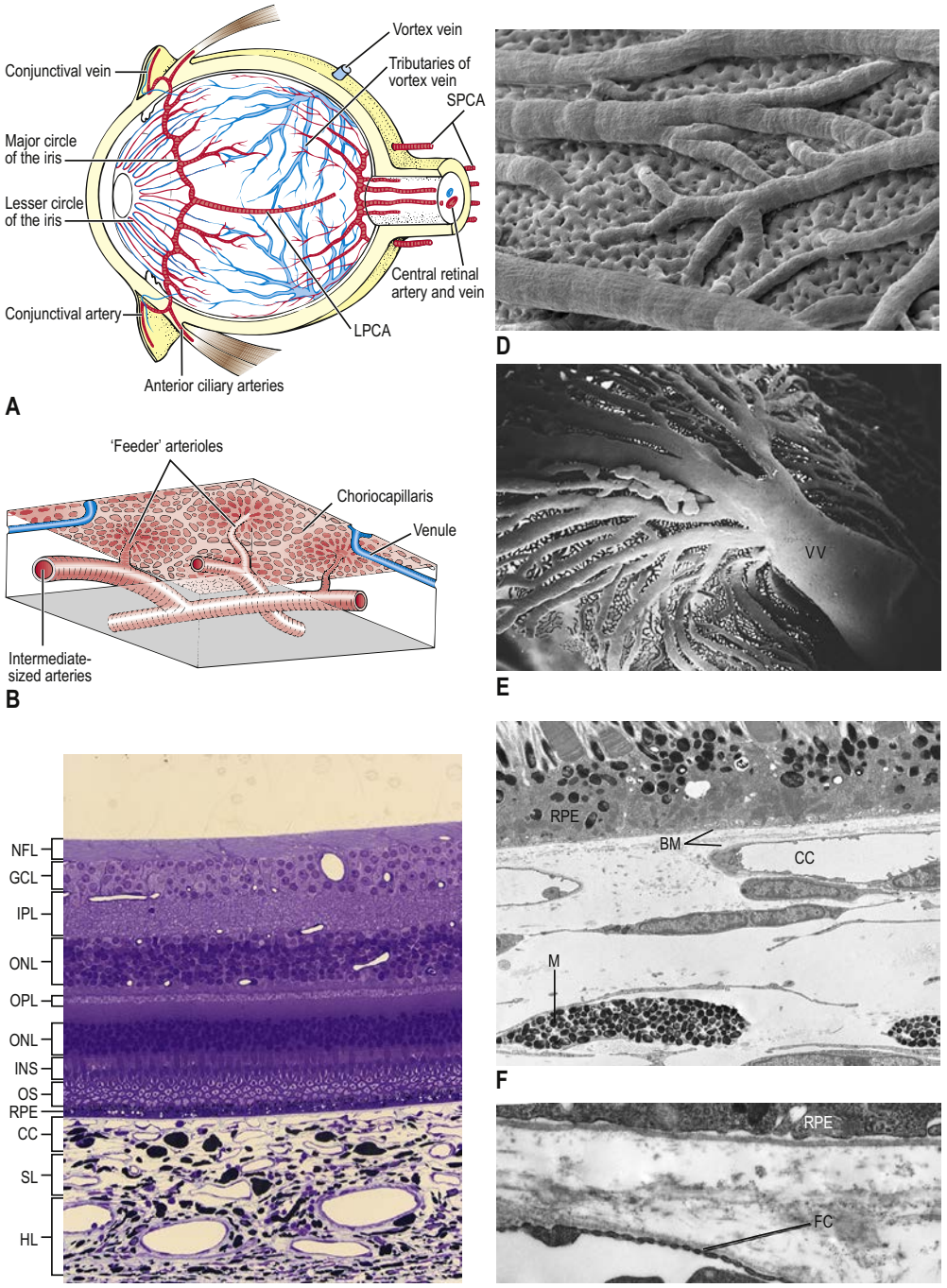


Fig. 1.39 (**A**) Diagram of the uveal tract blood supply. (**B**) Schematic representation of the hexagonal units in the choricapillaris fed by small arterioles. (**C**) Semi-thin resin section of the outer retina and choroid in the primate eye (baboon). Note the heavy degree of pigmentation and the layers of the choroid. (**D**) Resin vascular cast of the choriocapillaris viewed from the choroidal aspect. (**E**) Resin vascular cast of a quadrant of the choroid viewed from the external aspect showing a large vortex vein *(VV)*. (**F**) Low-power electron micrograph of the retinal pigment epithelium, Bruch's membrane *(BM)* and the closely related choriocapillaris in a primate eye. (**G**) Higher-power electron micrograph of the basal aspect of the RPE with its basal lamina forming part of Bruch's membrane. Note the fenestrated capillaries *(FC)* in the endothelial lining of the choriocapillaris. Note the capillary endothelial cells are characterized by fenestrae *(FC)* on the retinal aspect adjacent to Bruch's membrane. *CC*, Choriocapillaris; *GCL*, ganglion cell layer; *HL*, Haller's layer; *IPL,* inner plexiform layer; *LPCA*, long posterior ciliary arteries; *M*, melanocyte; *NFL*, nerve fibre layer; *ONL*, outer nuclear layer; *OPL*, outer plexiform layer; *OS*, outer segment; *RPE*, retinal pigment epithelium; *SL*, Sattler's layer; *SPCA*, short posterior ciliary arteries. Original magnifications: **C** × 350; **D** × 35; **E** × 120; **F** × 2400; **G** × 14 000. (Fig. 1.39D courtesy Dr Valantin

Box 1.23 **Comparative anatomy**

Tapetum is the cause of 'eye shine'

Many mammalian (carnivores, ruminants, cetaceans, seals) and non-mammalian (fish, crocodiles) species possess a reflective tapetum that may serve to increase photoreception in low light conditions. This may be located in the choroid or the retinal pigment epithelium (RPE). The choroidal tapetum may be cellular or fibrous and may occupy only part, usually the upper portion, of the globe. In carnivores it generally consists of several layers of flattened cells containing reflective material (e.g. guanine, zinc cysteine). In ruminants (e.g. cows, sheep) and cetaceans (dolphins) it is fibrous (tapetum fibrosum) and consists of fine regularly arranged collagen bundles that cause diffractive patterns depending on their orientation. A retinal tapetum generally consists of lipid (e.g. opossum) or guanine (fish, crocodiles) deposits within the RPE.



serrata and functions to provide nutritional support for the outer retina, especially the photoreceptors. The capillary 'network' is more akin to a perforated vascular 'net' than a network of capillaries (Fig. 1.39B,D,E).

The bore of the capillaries (20–40 μm) and the density of the 'net' (Fig. 1.39D) are greatest near the macula. The capillaries are fenestrated (75–85 nm diameter) on their retinal aspect (Fig. 1.39F,G) and these fenestrae occur at a density of approximately 46 per μm^2. Smooth muscle cells are not usually present in this layer. This sheet or net of capillaries is fed from arterioles, from the layer composed of arterioles and venules (Sattler's) in the manner depicted in Fig. 1.39B, i.e. hexagonal patches or 'lobules' of choriocapillaris are fed by a central precapillary arteriole that runs perpendicular to the flat choriocapillaris. This lobular pattern is clinically significant because choroidal ischaemia often occurs as pale hexagonal patches (mosaic pattern). The venous channels drain the periphery of these lobules (Fig. 1.39B).

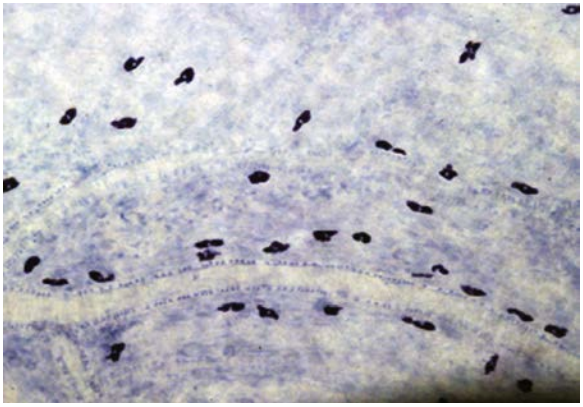
Vascular layer

This layer (Fig. 1.39C) lies beneath the choriocapillaris and can be subdivided into an inner layer of intermediate-sized vessels (arterioles and venules, Sattler's layer) and an outer component (major arteries and veins, Haller's layer). The blood supply of the choroid is chiefly from the long and short posterior ciliary arteries, although recurrent branches from the anterior ciliary arteries anastomose with anterior choroidal vessels (Fig. 1.39A). Venous drainage occurs via a series of large vortex veins (venae vorticosae), of which there are usually four (but may be up to six), each draining a sector of the choroid (Fig. 1.38B). These large veins pierce the sclera through emissary canals (Figs 1.17A and 1.39A,E) and drain into the superior and inferior ophthalmic veins in the orbit.

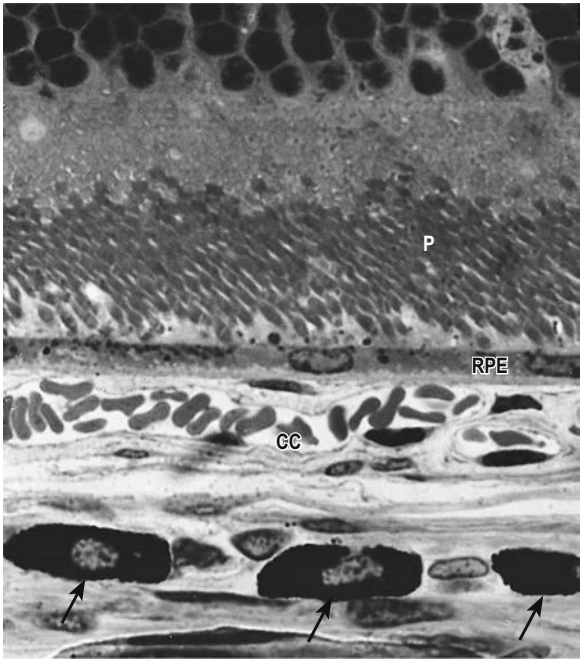
The choroidal stroma consists of randomly arranged collagen fibres (type I), flattened ribbon-like elastic fibres, fibrocytes and numerous melanocytes (Fig. 1.39C,F). The extent of choroidal pigmentation influences the appearance of the fundus, with the highly pigmented choroid of darker-skinned people showing through more than in the fundus of a person with less skin pigmentation, in which the red–orange reflex is primarily the result of the choroidal vasculature. The choroid, being a connective tissue, contains resident populations of immunocompetent cells including occasional plasma cells and lymphocytes, numerous perivascular mast cells (Fig. 1.40A,B,) and networks of resident tissue macrophages and dendritic cells whose cell bodies are located on the external (sclerad) aspect of the choriocapillaris and deeper in the choroid but whose processes extend close to the basal aspect of the RPE (Fig. 1.41). Their role in age-related diseases is only now being explored (Box 1.24 Clinical correlates). The clinical significance of the pattern of anastomosis between the long and posterior ciliary arteries is discussed in Box 1.25 Clinical correlates.

Suprachoroid

This is a 30 μm thick transition zone between the choroid and the sclera. It consists of thin interconnected lamellae of melanocytes, fibroblasts and connective tissue fibres separated by a thin 'potential' (supra- or perichoroidal) space, which in pathological conditions may become separated by fluid and blood. It is frequently artefactually enlarged in histological preparations. It is an avascular layer, the only vessels being those that traverse the suprachoroid entering or leaving the choroid. The lamellae blend with the choroid



A



B

Fig. 1.40 Mast cells in the choroid. (**A**) Low-power view of rat choroidal whole mount stained with toluidine blue which demonstrates the perivascular arrangement of mast cells. (**B**) Semi-thin resin section of rat outer retina and choroid stained with toluidine blue showing three mast cells *(arrows). A*, Artery; *CC*, choriocapillaris; *P*, photoreceptors; *RPE*, retinal pigment epithelium. Original magnifications: **A** × 75; **B** × 900.

and the lamina fusca of the sclera. The suprachoroidal space is continuous with the supraciliary space anteriorly. A previously unrecognized, highly organized network of non-vascular smooth muscle cells is present in the suprachoroid. These networks were particularly evident behind the fovea, around the entry points of the posterior ciliary arteries and nerves and in bundles running parallel to vessels travelling anteriorly from the posterior pole as far as the exit points of the vortex veins. The function of this network of smooth muscle cells in the human choroid remains speculative but may include regulation of blood flow.

Nerve supply of the choroid

The choroid is innervated by the *long and short ciliary nerves.* The long ciliary nerves (from the nasociliary branch of V_1) pass through the choroid and transmit sensory fibres to the cornea, iris and ciliary body. Sympathetic fibres are also carried in these nerves to the dilator pupillae (see p. 29). The *short ciliary nerves* arise from the ciliary ganglion and carry sensory (from nasociliary), sympathetic and parasympathetic fibres (derived predominantly from nerve III, but also from VII). The latter have already synapsed in the pterygopalatine ganglion. Both long and short ciliary nerves pierce the sclera in the form of a ring 2–3 mm anterior to the optic nerve sheath, along with the long and short posterior ciliary arteries. The nerve terminals branch extensively and form plexi of unmyelinated fibres in the choroid and suprachoroid adjacent to vascular smooth muscle cells; however, they do not extend into the choriocapillaris. Fibres containing vasoactive intestinal peptide (VIP) and neuropeptide Y have been identified in the choroid and probably act as vasodilator and vasoconstrictor agents, respectively. Multipolar and bipolar ganglion cell bodies immunoreactive for nitric oxide synthase (NOS) and VIP have been recently identified in the choroid and their axons may supply the choroidal vasculature (vasodilatory) or non-vascular smooth muscle cells. Their structure and immunohistochemical characteristics suggest that they may have a mechanosensory role.

Optic Nerve

The optic nerve (Fig. 1.42) is unique anatomically as it is the only tract in the central nervous system to leave the cranial cavity. Furthermore, it is subdivided into fascicles by connective tissue and glial septae and is surrounded by cerebrospinal fluid. It is also unique in that it is the only central nervous system tract that can be visualized clinically.

The optic nerve is formed by convergence of ganglion cell axons at the optic disk, the commencement of the nerve. Foveal/macula fibres constitute around 90% of all axons leaving the eye and form the distinct *maculopapillary bundle.* From the disk, the axons extend along the nerve through the orbit to traverse the optic canal in the sphenoid bone.

The optic nerve can be divided into four main portions: *intraocular* (1 mm in length), *orbital* (25–30 mm), *intracanalicular* (4–10 mm) and *intracranial* (10 mm). The latter portion is discussed on p. 94 in the context of the visual pathways.

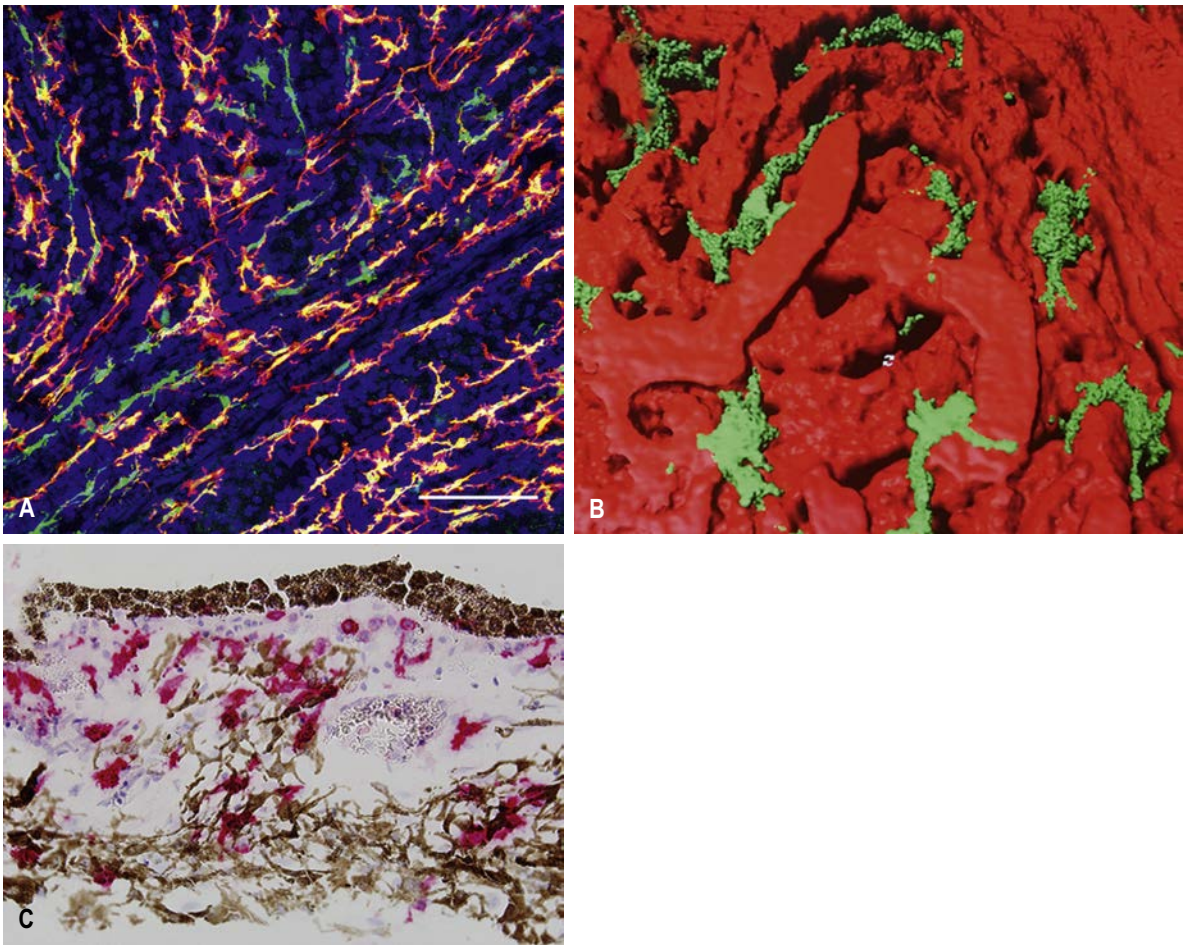


Fig. 1.41 (**A**) Confocal microscopy image of a large region of a choroidal wholemount from Balb/c Cx3cr1-$^{gfp/gfp}$ mouse stained with MHC class II *(red)*. GFP/*green* fluorescence indicates that all myeloid cells are Cx3cr1^{+} with a significant majority co-expressing MHC class II *(yellow/orange)*. (**B**) The sclerad aspect of the choriocapillaris *(red)* in Balb/c Cx3cr1-$^{gfp/gfp}$ mice showing the location of macrophage cell bodies. (**C**) Normal human choroid (53 year old) obliquely sectioned (hence the retinal pigment epithelium [RPE] appearing multilayered and the choroid being apparently thicker in profile) and stained with Iba-1 *(red)* to illustrate large numbers of pleomorphic Iba-1^{+} macrophages in the choroid beneath the RPE and interspersed between the melanocytes *(brown)*.

Box 1.24 **Ageing changes in the eye**

Choroidal immune cells and age-related macular degeneration (AMD)

The deposition of lipid, complement and immunoglobulin G that accompanies the senescent changes in Bruch's membrane may be augmented by impaired macrophage recruitment and/or decreased homeostatic scavenging by the resident macrophages and dendritic cells in the choroid and thus they may contribute to the accumulation of debris and the formation of drusen in 'dry AMD' and to the eventual choroidal neovascularization in 'wet AMD'.

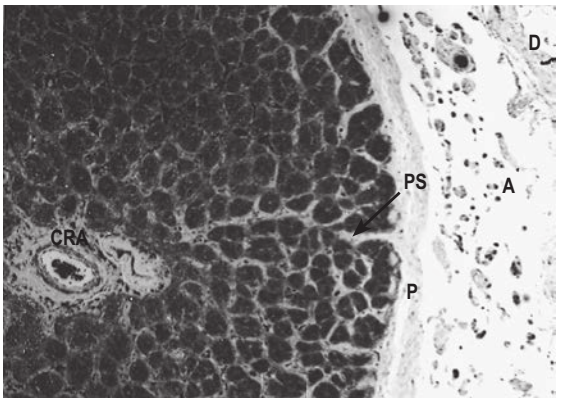
Box 1.25 **Clinical correlates**

Choroidal infarctions

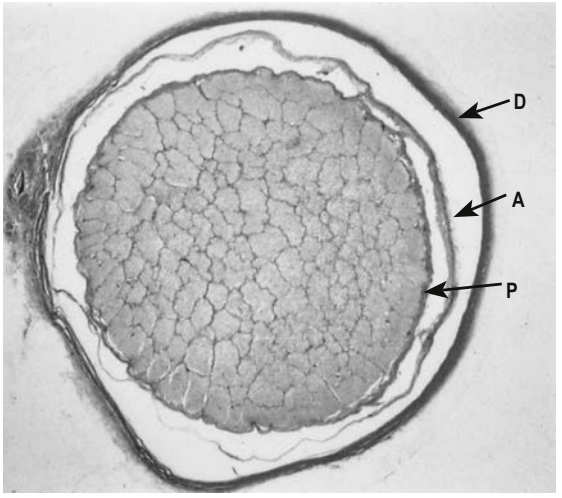
These appear on angiograms to take the form of triangular areas near the equator, with the apex pointing towards the optic disk. There is probably less functional anastomosis between choriocapillaris lobules than was once suspected; however, some degree of anastomosis in the subcapillary arterioles exists. Peripheral retinal cobble or paving-stone degeneration represents chronic focal ischaemic changes in the anterior choroid.



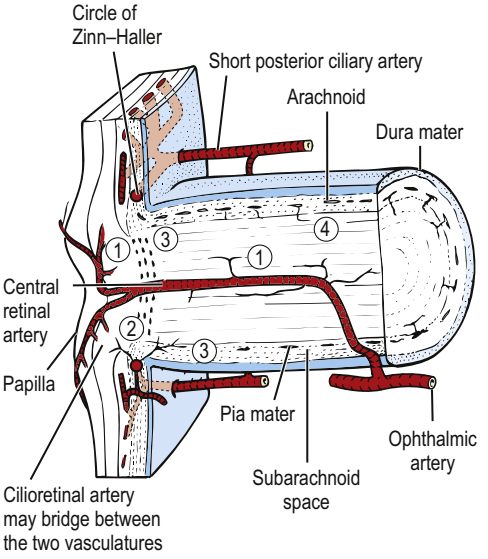
A



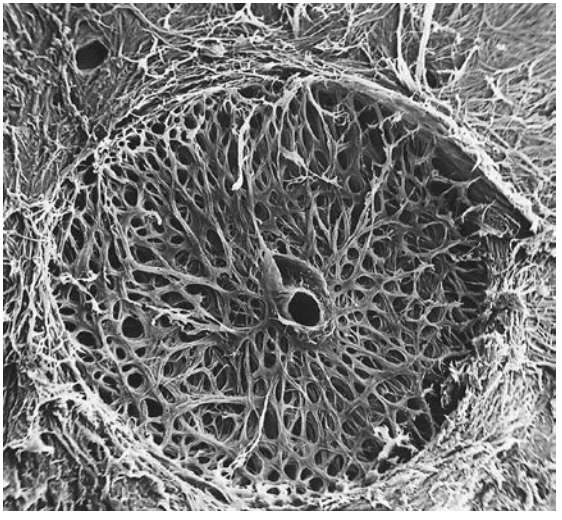
B



C



D



E

Fig. 1.42 (**A**) Histological section of the optic nerve head: *LC*, lamina cribrosa; *A* and *V*, central retinal artery and vein; *SAS*, subarachnoid space. (**B**) Transverse section (Loyez stain) of the orbital portion of the optic nerve revealing the arrangement of the myelinated nerve fascicles *(darkly stained)* separated by pial septae *(PS)* which penetrate as far as the central retinal artery *(CRA)* in the middle of the nerve. The three layers of meninges surrounding the nerve, dura *(D)*, arachnoid *(A)* and pia mater *(P)* are clearly visible here and in (**C**). (**C**) Cross-section (trichrome stain) of an entire optic nerve and surrounding meninges posterior to the entry of the central retinal artery. (**D**) Blood supply of the optic nerve. The four sources of vessels supplying the optic nerve include: *1*, branches from the central retinal artery or its branches; *2*, branches from the circle of Zinn-Haller; *3*, choroidal branches; *4*, pial branches. (**E**) Scanning electron microscopy of the lamina cribrosa *(LC)*. *A* and *Vn*, Apertures for the central artery and vein. Original magnifications: **A** × 60; **B** × 290; **C** × 40. **E** × 75. (Fig. 1.42E, courtesy Dr A Thale.)

ALGrawany

Box 1.26 **Clinical correlates**

Glaucoma

The intraocular portion is the part of the optic nerve damaged in glaucoma. Axonal damage may be a consequence of either interference with blood flow or interruption of axonal transport and raised intraocular pressure is a significant risk factor. No single hypothesis has been proposed that adequately explains why specific regions of the nerve are more likely to be damaged than others, resulting in the characteristic visual field defects or *scotomas*. Recent research has highlighted the differential pressure gradient across the lamina cribrosa between the cerebrospinal fluid pressure and the intraocular pressure (IOP) (the translaminar pressure gradient (TLPG)) and how it may influence central retinal venous pressure as it traverses the lamina cribrosa, including the 'arteriolization' of the vessel wall and endothelial lining. It appears that other factors such as the buffering effects of orbital tissue, pia mater and the conformation of the lamina itself may further influence the TLPG and this previously unsuspected pressure gradient may in part be responsible for the outward bowing of the lamina in the optic cup. Increased TLPG and its relationship to IOP may have a role in the progression of optic nerve damage and glaucoma.

Box 1.27 **Clinical correlates**

Papilloedema

Papilloedema is the swelling of the papillary fibres, which appears as a raised white disk margin, and is partly the result of the lack of Müller cells in this region. These cells serve to bind the nerve fibres together in the remainder of the retina. Papilloedema may be a cardinal sign of raised intracranial pressure transmitted to the subarachnoid space, which envelopes the optic nerve as far anteriorly as the sclera surrounding the optic nerve.

Meningitis

The continuation of the subarachnoid space from the cranial cavity along the nerve may facilitate the spread of infection or tumours from the orbit to the cranium and vice versa.

Intraocular portion

The *intraocular portion* (Fig. 1.42A) extends from the surface of the optic disk to the posterior margins of the sclera. The nerve fibres are not myelinated in this portion. It can be further subdivided into three regions: the retinal (pars retinalis), choroidal (pars choroidalis) and scleral (pars scleralis) portions. Myelination commences approximately level with the termination of the subarachnoid space at the posterior limits of the lamina cribrosa. As the fascicles of nerve fibres pass posteriorly from the optic disk into the intraocular portion, the glial cells become more common; columns of glial cell nuclei are especially prominent in the scleral portion, where they account for up to 40% of the tissue mass (Fig. 1.42A). The commencement of the optic nerve, the *optic disk*, varies depending on the method of measurement but is approximately between 1.7 and 2.8 mm in diameter, although variations both within a population and between races are observed and this variation has been linked to susceptibility to glaucoma. The layers of the retina and the choroid terminate at the edge of the disk as specialized regions of glial tissue, the intermediary tissue (of Kuhnt) and marginal border tissue (of Elschnig). The absence of retinal tissue in this region explains the *'blind spot'* phenomenon in the visual field. As the 1.2 million ganglion cell axons in the nerve fibre layer become crowded towards the disk, they create a raised area or *papilla*, which is thickest on the lateral aspect owing to the large number of fibres in the *maculopapillary bundle*. The raised margin of the optic disk surrounds an indentation, the *physiological cup*. As the fibres pass posteriorly, they pierce the sieve-like connective tissue mesh, the *lamina cribrosa*, which fills the posterior scleral foramen. The lamina cribrosa is formed by irregular collagen fibre bundles continuous with the sclera. These bundles are arranged in the form of circles or a figure of eight (Fig. 1.42E). Elastic tissue from the choroid and Bruch's membrane is continuous with and 'anchored' to the adventitia surrounding the central retinal artery and vein. The collagenous bundles in the lamina cribrosa are separated from the axons by a covering of glial tissue, which may protect the nerve fibres as they pierce the irregular openings. The *scleral canal* is about 0.5 mm long and may vary in shape from cone-like (narrowest portion nearest the disk) to double cone or funnel-like. Posterior to the pars scleralis, the nerve fibres become myelinated by *oligodendrocytes* (Fig. 1.42B), causing a doubling of the thickness of the optic nerve. The clinical significance of the anatomy of the optic disk and the optic nerve with its meningeal coverings enclosing an extension of the subarachnoid space is highlighted by its central role in understanding the pathogenesis of glaucoma (Box 1.26 Clinical correlates), as an indicator of raised intracranial pressure (Box 1.27 Clinical correlates) and as a conduit between the orbit and cranial cavity (Box 1.27 Clinical correlates).

Orbital portion

The *orbital portion* of the optic nerve (Fig. 1.42C,D) extends backwards and medially from the back of the eye to the optic canal in the sphenoid at the apex of the orbit. It is covered by three layers of meninges: pia, arachnoid and dura (see page 8). The dura and arachnoid blend with the

sclera, and the subarachnoid space (SAS) around the nerve terminates at the posterior surface of the sclera in the form of a fluid-filled ring (Fig. 1.42A). The central retinal vessels must cross the SAS and are therefore vulnerable, particularly the vein, in cases of increased intracranial pressure.

The majority of the axons in the nerve are 1 μm in diameter and approximately 10% are between 2 and 10 μm. The glial septae between fascicles present in the intraocular portions extend into the orbital portion but become less distinct as the orbital apex is approached. The orbital portion of the optic nerve has a slight S-shaped bend, which allows a full range of ocular movement without stretching the nerve. As the optic nerve approaches the orbital apex it is surrounded by the tendinous annulus, which is the origin of the rectus muscles.

Intracanalicular portion

The *intracanalicular portion* of the optic nerve passes through the optic canal (foramen) accompanied by the ophthalmic artery and sympathetic nerves. The dura surrounding the nerve splits at the orbital opening, the majority continuing as the dural sheath of the nerve inside the canal and a thinner portion blending with the periorbital fascia (see next section) (Fig. 1.43).

Blood supply

The optic nerve has a complex blood supply, which has been extensively investigated because of its importance in the pathogenesis of glaucoma (see Chapter 9). The intraocular portion is supplied by branches from four sources: central retinal vessels and their branches, scleral vessels (the circle of Zinn-Haller), choroidal vessels and pial vessels (see Fig. 1.42D). The first three are derived from the ophthalmic or central retinal artery and pial vessels from the adjacent branches of the internal carotid artery. The majority of capillaries pierce the nerve and course longitudinally within the nerve via the glial septae.

Orbital contents

General arrangement

The orbits are a pair of bony sockets, with each orbital cavity having a volume of about 30 cm^3.

Periorbita and orbital fibroadipose tissue (Figs 1.43 and 1.44)

The orbital contents are bound together and supported by fibroadipose tissue. This connective tissue is classically divided into separate components.

- *Periorbita* or periosteum of the orbit. This layer of connective tissue is frequently described as having a dense outer layer and a looser inner layer, which invests orbital nerves and the lacrimal gland. It is tightly bound to the bones only at the sutures, fissures and foraminae in the orbital walls, and also to the posterior lacrimal crest

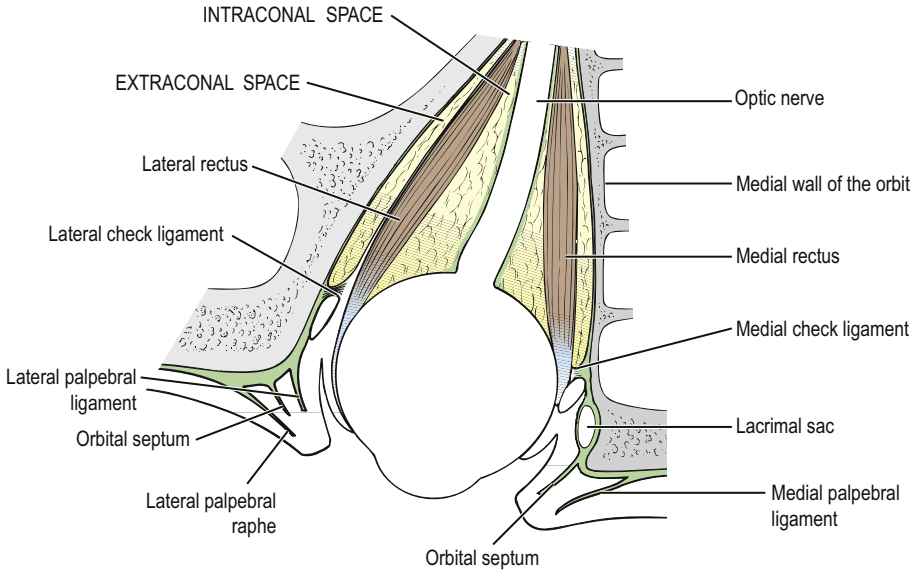


Fig. 1.43 Diagram of a horizontal section through the orbit illustrating the formation of an intra- and extraconal space by the four rectus muscles (only medial and lateral rectus shown in this section).

where it covers the lacrimal sac and is continuous with the fibrous lining of the nasolacrimal duct. It forms a dense membrane over the inferior and superior orbital fissures, with sufficient gaps for transmission of nerves and vessels. It is continuous with the periosteum lining the optic foramen and with the sheath of the optic nerve, itself an extension of dura mater of the brain. The periorbita is firmly attached at the orbital margins anteriorly where it becomes continuous with the orbital septum (palpebral fascia) in the eyelids.

- *Bulbar fascia (Tenon's capsule)*, a thick fibrous sheath enclosing the globe but separated from it by a layer of loose connective tissue.
- *Muscular fascial sheaths* that surround the extraocular muscles and blend with the bulbar fascia.
- *Medial and lateral check ligaments.*
- *Suspensory ligament (of Lockwood).*

The fibrous intermuscular membrane connecting the four rectus muscles helps create the intraconal space (best developed in the anterior part of the orbit; incomplete behind the globe) (Fig. 1.44B,C). Besides the 'check' ligaments there are other specific attachments via fibrous bands to the orbital walls throughout their course (Fig. 1.44A–C). A theoretical framework known as the 'active pulley hypothesis' postulates a crucial role for these connective tissue bands, known as 'pulley suspensions', in understanding the kinematics of extraocular muscle action. These suspensions pass between the orbital wall and the 'pulley sleeve' of each muscle, which is described as a ring-like extension of the connective tissue from Tenon's capsule posteriorly around the muscle. The tone of the pulleys is possibly under neuronal control because of the presence of smooth muscle fibres. The active pulley hypothesis proposes that the rectus muscles have an orbital layer of fibres that are continuous with (or 'blend with' or 'insert into') these sleeves (and thus also into the pulley suspensions), in essence one part of a bifid insertion. The inner half or global layer of the rectus muscle continues through the sleeve and bulbar fascia to insert directly into the sclera. The newly postulated function for orbital connective

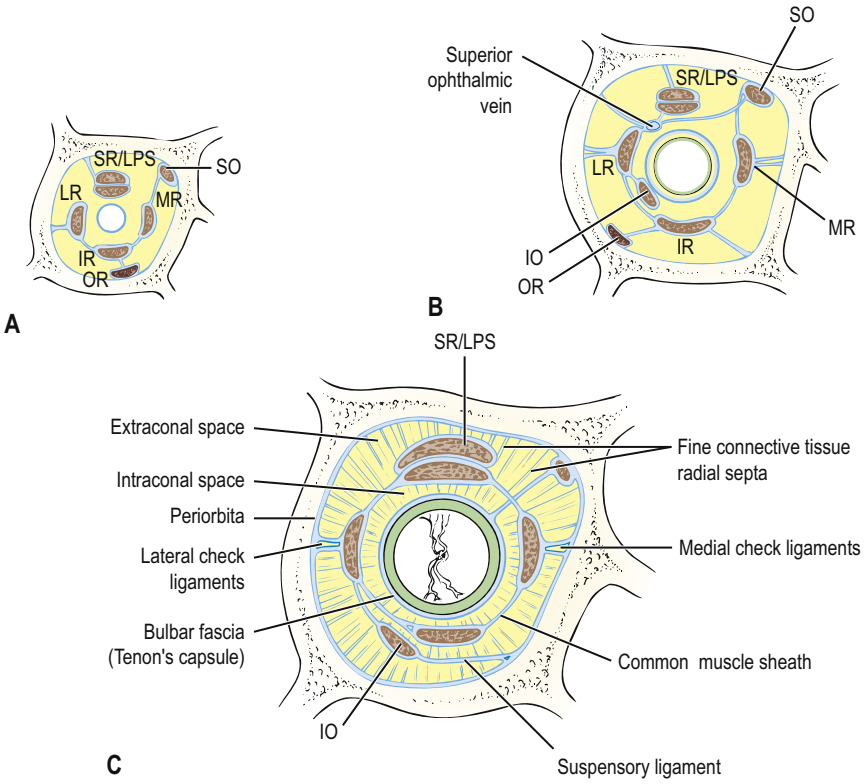


Fig. 1.44 Schematic diagrams of the connective tissue septae associated with each extraocular muscle at three levels in the orbit: (**A**) near orbital apex; (**B**) posterior part of the globe; (**C**) close to the equator of the globe. *IO*, Inferior oblique; *IR*, inferior rectus; *LPS*, levator palpebrae superioris; *LR*, lateral rectus; *MR*, medial rectus; *OR*, orbitalis; *SO*, superior oblique; *SR*, superior rectus. (Modified from Koornneef, 1982.)

tissue in the active pulley hypothesis, which states in effect that this dual insertion allows the pulleys to act as a second origin and thus influence the direction of pull of the extraocular muscles, has gained wide acceptance but some investigators have questioned the anatomical evidence of orbital muscle fibres terminating in connective tissue other than the sclera.

There are well-recognized but variable amounts of smooth muscle within the orbital connective tissue, including the sleeves of some of the recti muscles, whose functions (besides the superior and inferior palpebral muscles) are currently unclear. It has been suggested that like the smooth muscle covering the inferior orbital fissure (orbitalis or Müller's muscle) they may represent a redundant evolutionary remnant.

Veins passing through the orbit are supported by connective tissue septae (Fig. 1.44B); arteries, instead, travel among the fat locules and frequently pierce the septae. A thickened band of orbital fibrous tissue connects the superior rectus and the levator palpebrae superioris. This aids in coordinating lifting of the eyelid when the eye is directed upwards by the superior rectus.

The complex and interlinked nature of the fibroadipose system of connective tissue septae might explain why patients with orbital floor 'blow-out' fractures display vertical ocular mobility problems. It is not necessary to invoke the incarceration of the inferior rectus and inferior oblique muscles in the fracture to explain the symptoms.

Extraocular muscles

There are six true extraocular muscles responsible for movements of the globe. In addition, there is one further orbital muscle, the levator palpebrae superioris, which originates at the orbital apex and inserts into the tarsal plate and upper eyelid (see p. 87).

The true extraocular muscles are *four rectus muscles*, which arise from the tendinous ring at the apex of the orbit and insert into the sclera about 4–8 mm behind the limbus, and *two oblique muscles* (superior and inferior), whose tendons approach the globe from in front and insert into the posterior aspect of the sclera (Figs 1.45 and 1.48). Details of the six true extraocular muscles, including their innervation, origin, insertion, tendon length (important in the surgical management of strabismus), length of muscle belly, the angle subtended by the muscle axis to the vertical, and the size of the motor units, are provided in Table 1.2. The origins of the muscles are shown in Fig. 1.47 and the pattern of insertion into the sclera is shown in Fig. 1.48. The collagen bundles of the tendons blend with the scleral collagen as shown in Fig. 1.17C. The relations of the orbital muscles to each other and to the orbital nerves are summarized diagrammatically in Figs 1.45 and 1.46. Movements are discussed in Chapter 5 (p. 339).

Microscopic anatomy of extraocular muscle

Histologically, extraocular muscle differs from skeletal muscle in the following respects (compare Fig. 1.49A with Fig. 1.49B).

- The epimysium or muscle sheath of extraocular muscle is generally very thin by comparison with other muscles.
- The fibres are not tightly packed but are separated by unusually large amounts of connective tissue (perimysium) rich in reticulin and elastic fibres.
- The muscle fibres are rounded or oval in shape with small fibres (5–15 μm) around the periphery of the muscle and larger fibres (10–40 μm) in the centre.
- Extraocular muscle is the most vascular in the body, after myocardium. The most vascularized region is the orbital aspect.
- In normal extraocular muscle there often appear to be histopathological or ultrastructural changes normally associated with myopathy, i.e. mild mononuclear cellular infiltrate, centrally placed nuclei, disorganization of the sarcolemma, disruption of the Z lines and mitochondrial clumping.
- Extraocular muscle contains large numbers of specialized sensory or proprioceptive endings, including large muscle spindles up to 1 mm long (nuclear bag fibres, nuclear chain fibres and annular nerve terminals). Golgi tendon organs are also numerous and are generally found within the tendons of extraocular muscles in greater numbers than in skeletal muscle (Fig. 1.49). The afferent fibres from extraocular muscles are transmitted initially for part of their course in the respective cranial nerve innervating the muscle (either III, IV or VI); however, they leave these nerves and join the ophthalmic division of the trigeminal, either in the cavernous sinus or in the brainstem. Their cell bodies are situated in the mesencephalic nucleus, although some muscle afferents have been traced to Purkinje cells in the cerebellum, and play an important role in positional sense and control of ocular movements (both saccadic and tracking).

The structural differences between extraocular and skeletal muscle outlined above are not surprising in light of the fundamental differences in function, namely the constancy of activity (even during sleeping) and the rapidity and fine gradation of contraction of extraocular muscle required to fixate subjects of interest on the fovea. Since both eyes must move together, both sets of six muscles must be highly coordinated and move simultaneously (see Chapter 5). Up to six types of muscle fibre have been identified morphologically, but functionally there appear to be three main types (Table 1.3).

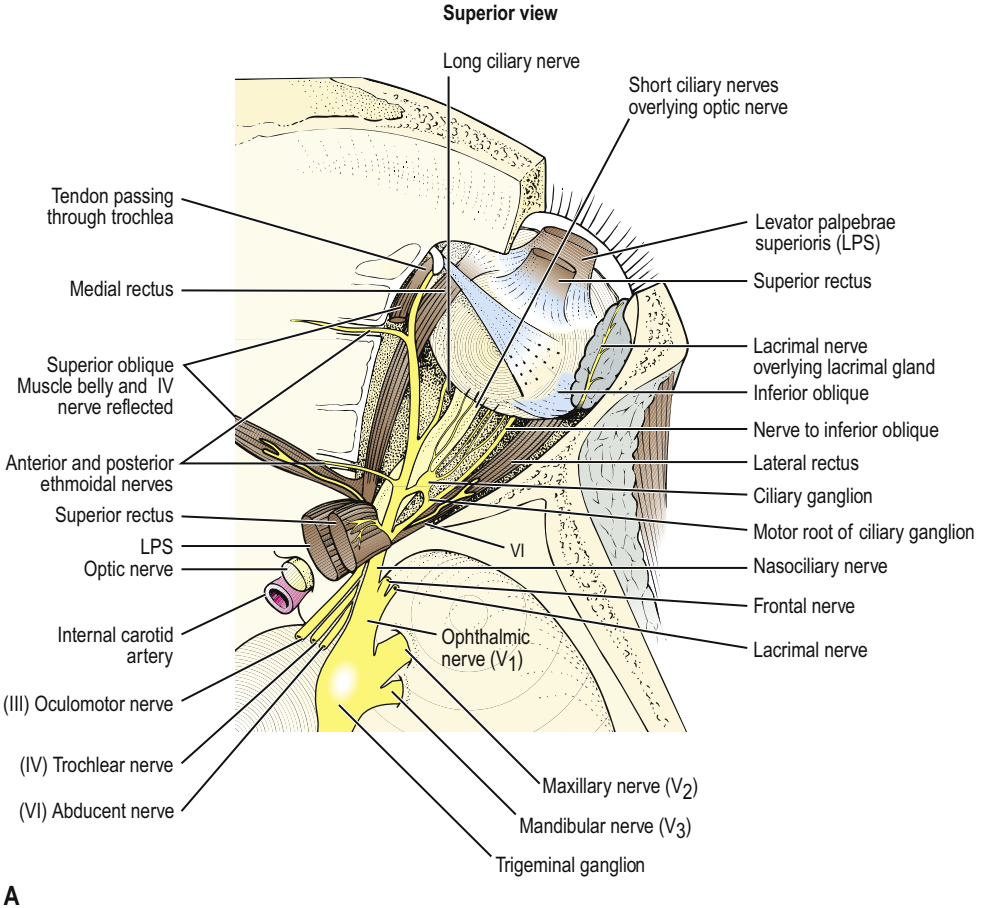


Fig. 1.45 (**A**) Diagram and (**B**) prosected specimen of the orbit viewed from above, revealing the relations of the orbital nerves and extraocular muscles. Orbital fat and vessels have been excluded for the purposes of clarity.

- The roof of the orbit and superior orbital fissure have been removed and the periorbita divided.
- In (**A**) the lacrimal *(L)*, frontal *(F)* and trochlear *(IV)* nerves have been cut. In (**B**) these nerves are intact as they pass external to the tendinous ring or annulus.
- In (**A**) only one long ciliary nerve is shown arising from the nasociliary nerve.
- The sensory root of the ciliary ganglion emerges from the nasociliary. The motor root (parasympathetic fibres) arises from the branch of the oculomotor supplying inferior oblique.
- In (**A**) and (**B**) the lateral dural covering of the cavernous sinus has been removed to expose the cranial nerves before they pass through the superior orbital fissure.
- The optic canal has not been opened.

ICA, Internal carotid artery; *LG*, lacrimal gland; *LPS*, levator palpebrae superioris; *LR*, lateral rectus; *OB*, olfactory bulb; *ON*, optic nerve; *SO*, superior oblique; *SR*, superior rectus; *TG*, trigeminal ganglion; V_1, V_2, V_3, divisions of the trigeminal nerve; *III*, oculomotor nerve; *VI*, abducens nerve.

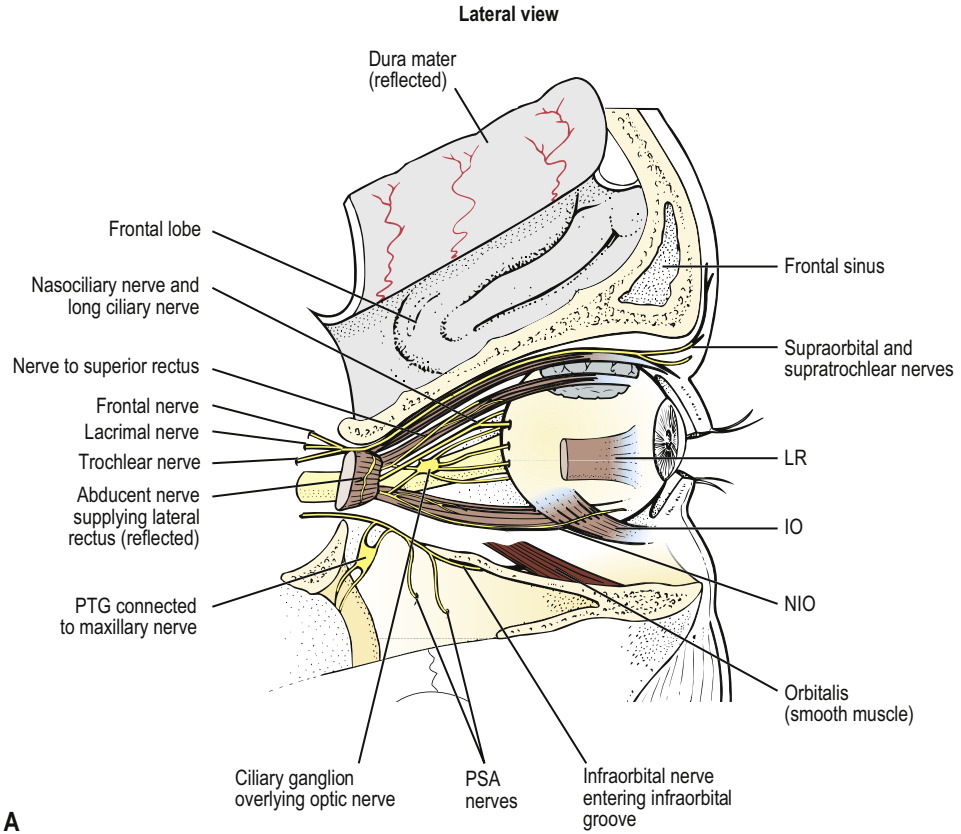


Fig. 1.46 (**A**) Diagram illustrating a lateral view of a dissected orbit revealing the relations of the orbital nerves and extraocular muscles (vessels have been excluded for the purposes of clarity). (**B**) Dissection of the orbit similar to the previous diagram (Fig.1.45) except that lateral rectus has not been cut and the course of the orbital nerves within the cavernous sinus is also shown (by removal of the lateral dural wall).

- The lateral wall has been removed and the infratemporal fossa has been dissected to expose the pterygomaxillary fissure and pterygopalatine fossa.
- The cranial cavity has been opened to reveal the dura (reflected) covering the frontal lobe.
- The lateral rectus has been divided and reflected to expose the optic nerve and other cranial nerves entering the orbit through the tendinous ring. Note the abducent nerve entering its bulbar surface.
- The ciliary ganglion lies between the lateral rectus and the optic nerve. Note the motor root and sensory root as seen in the superior view. The nerve to inferior oblique *(NIO)* is a useful landmark for finding the ciliary ganglion. The short ciliary nerves emerge from the ganglion and enter the globe around the optic nerve.
- Note the three nerves that enter the orbit outside the tendinous ring: lacrimal, frontal and trochlear.
- The nerve to superior rectus (branch of superior division of III nerve) pierces the muscle and enters the levator palpebrae superioris, which it supplies, from below.
- Branches of the pterygopalatine ganglion *(PTG)* enter the orbit through the inferior orbital fissure and contribute to the formation of the retrobulbar plexus (not shown).
- Inferior oblique passes backwards, laterally and superiorly beneath the inferior rectus.
- Orbitalis (Müller's muscle), a band of smooth muscle, covers the inferior orbital fissure.

ICA, Internal carotid artery; *IO*, inferior oblique; *ION*, infraorbital nerve; *LR*, lateral rectus; *MA*, maxillary artery; *MS*, maxillary sinus; *PCA*, posterior communicating artery; *PC*, posterior cerebral artery; *PSA*, posterior superior alveolar nerves; *SR*, superior rectus; *TG*, trigeminal ganglion; *V_1, V_2, V_3*, divisions of the trigeminal nerve; *IV*, trochlear nerve. *Arrow*, nerve to superior oblique branch of oculomotor nerve (III).

Table 1.2 **Summary of anatomical features of human extraocular muscles**

Muscle	Innervation	Origin	Insertion (mm from cornea)	Tendon length (mm)	Length of muscle belly (mm)	Size of motor unit	Comment
Medial rectus	III (inf)	Tendinous ring	5.6	3.6	40	1 : 1.7–1 : 4	Largest of the ocular muscles
Inferior rectus	III (inf)	Tendinous ring	6.6	5	40	1 : 2–1 : 6	
Lateral rectus	VI	Tendinous ring via two heads	7.0	8.4	40	1 : 3–1 : 6	Opening between the two heads bridging the medial end of superior orbital fissure
Superior rectus	III (sup)	Tendinous ring	7.8	5.4	41	1 : 4	Lies beneath levator palpebrae superioris
Superior oblique	IV	Superomedial to optic canal	Lateral aspect of posterosuperior quadrant	Tendon forms 10 mm before winding around trochlea	32	1 : 5–1 : 6	The only extraocular muscle with a fusiform shape. IV nerve enters the muscle on its upper border
Inferior oblique	III (inf)	Behind orbital margin lateral to nasolacrimal canal	Posterolateral quadrant, mostly below horizontal	Very short tendon; muscle fibres almost reach the sclera	34	1 : 7	The only extraocular muscle not to originate at apex of orbit. Passes between the eye and lateral rectus

III (inf), Inferior division of oculomotor nerve; *III (sup)*, superior division of oculomotor nerve.
Values are approximate means based on several studies; for full details and ranges see Eggers (1982).

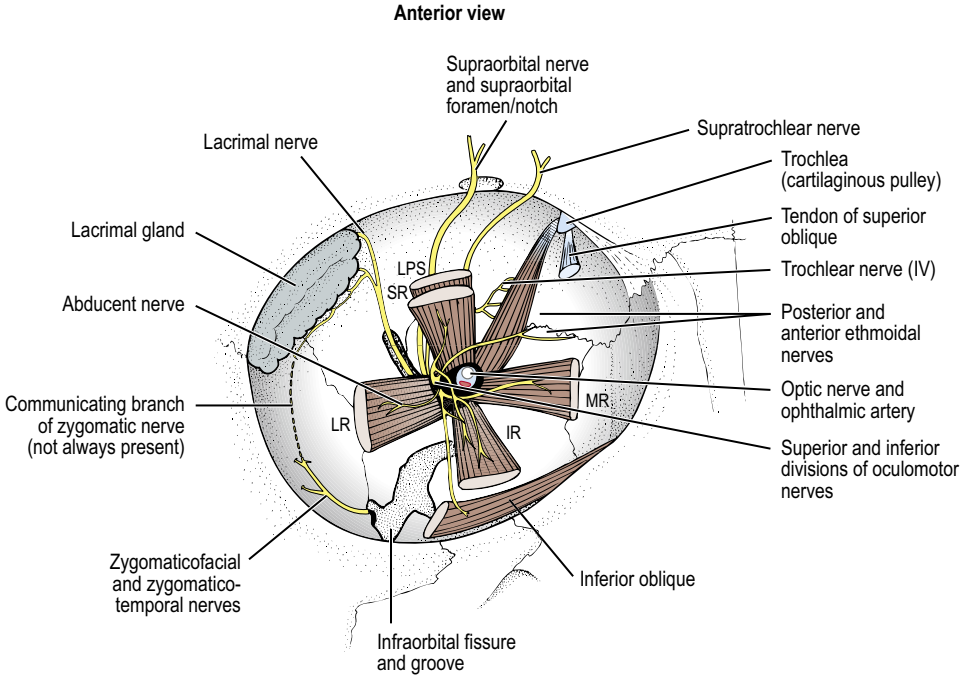


Fig. 1.47 The orbit from in front with the globe removed to show the origins of the extraocular muscles and the orbital nerves (vessels and fat are not included in this diagram).

Blood vessels of the orbit (Fig. 1.50)

The orbital contents are supplied chiefly by the ophthalmic artery, which usually arises from the internal carotid artery shortly after it emerges from the roof of the cavernous sinus. It commences its course beneath the optic nerve, closely bound to the dura while in the optic canal, then winds around its lateral aspect and finally passes above the nerve. It then proceeds forwards above the medial rectus and under the superior oblique. It ends its tortuous course by dividing into dorsal nasal and supratrochlear branches. The branches are summarized in Fig. 1.50. There are several important points of anastomosis between arteries derived from the internal carotid and the external carotid arteries (Table 1.4).

These anastomoses may be important during occlusive vascular disease of the ophthalmic artery by serving as alternative routes of blood supply to the eye and orbit. The veins that accompany the above arteries, in common with most veins of the head and neck, lack valves and thus there are several sites of communication between veins on the upper face and lids with intraorbital veins (superior and inferior ophthalmic veins), which drain posteriorly into the cavernous sinuses. The inferior ophthalmic vein may drain via the inferior orbital fissure into the pterygoid venous plexus. These communications are clinically important because they act as potential routes for the spread of infection from the face around the nose and eye to the cavernous sinuses and the cranial cavity.

Cranial nerves associated with the eye and orbit

General functional arrangement

Cranial nerves contain a diversity of functional components. Besides those found in spinal nerves (somatic efferents, somatic afferents, general visceral efferents, general visceral afferents), cranial nerves also contain additional functional categories including special visceral efferents (branchiomotor), special somatic afferents (special senses, hearing and balance) and special visceral afferents (taste and smell).

A table listing the cranial nerves and their target tissues/organs is available online at https//expertconsult.inkling.com

The sites of origin of the cranial nerves from the brainstem are illustrated in Figs 1.51–1.58.

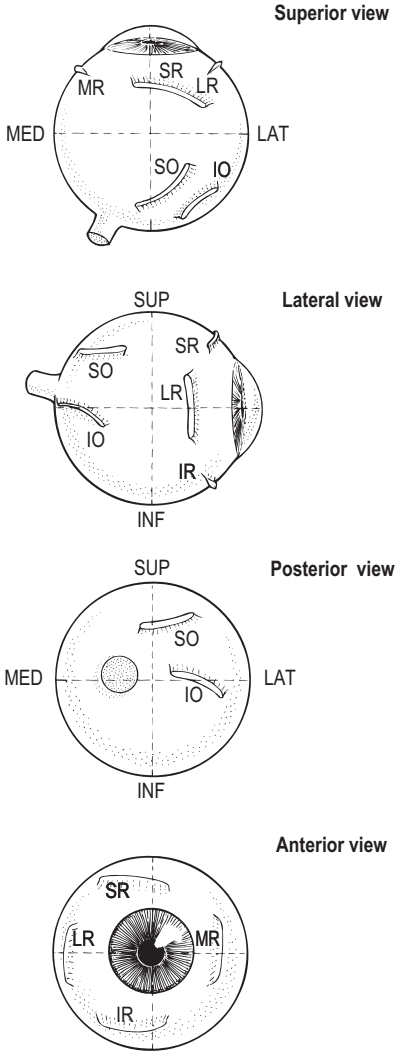


Fig. 1.48 Four views of the right globe to demonstrate the insertions of the extraocular muscles. *IR*, inferior rectus; *SR*, superior rectus; *MR*, medial rectus; *LR*, lateral rectus; *SO*, superior oblique; *IO*, inferior oblique.

Oculomotor nerve (cranial nerve III) (Figs 1.52 and 1.53)

This is the largest of the extraocular nerves and supplies all the extraocular muscles except the lateral rectus and superior oblique.

Origin (Fig. 1.52)

The oculomotor nerve nuclei comprise two main types:

- a complex of five individual motor (*somatic efferent*) nuclei containing the cell bodies of the multipolar motor neurones whose axons directly innervate their respective extraocular muscles (Fig. 1.52E).
- Dorsal nucleus (inferior rectus)
- Intermediate nucleus (inferior oblique)
- Medial nuclei (superior rectus)
- Central caudal nucleus (Levator palpebrae superioris)
- Ventral nucleus (medial rectus)
- A *general visceral efferent* nucleus, the Edinger-Westphal nucleus, containing small spindle-shaped preganglionic (first-order) *parasympathetic* neurones.

The oculomotor nuclei lie at the level of the superior colliculus in the midbrain in the ventral to the cerebral aqueduct in the periaqueductal grey matter and extends cranially for a short distance into the floor of the third ventricle (Fig. 1.52A–C). The *medial longitudinal fasciculus* lies ventrolateral to the nucleus and contains the axons of internuclear neurones that pass vertically between the brainstem nuclei of the III, IV and VI nerves (Fig. 1.52D). The fibres emerge from the oculomotor nuclei, pass anteriorly through the tegmentum of the midbrain and red nucleus and medial to the substantia nigra and emerge on the medial aspect of the cerebral peduncle at the upper border of the pons in the interpeduncular cistern (Fig. 1.52A).

Intracranial and intracavernous course

The nerve passes forwards, laterally and slightly downwards in the interpeduncular fossa (one of the enlargements of the subarachnoid space or cisterns) lateral to the posterior communicating artery (Fig. 1.51). It passes between the posterior cerebral artery (dorsal) and the superior cerebellar artery (ventral). It grooves the posterior clinoid process and courses forwards before it passes through the dural roof of the cavernous sinus (Figs 1.46B and 1.53B). The nerve runs forwards in the upper part of the lateral wall of the cavernous sinus (Fig. 1.9C) and enters the intraconal space of the orbit through the superior orbital fissure within the tendinous ring (Figs 1.5 and 1.6), where it divides into superior and inferior divisions. The clinical relevance of its relations before entering and within the cavernous sinus are discussed in Box 1.28 Clinical correlates.

Intraorbital course

In the orbit the nasociliary nerve is interposed between the two divisions of the oculomotor nerve. The superior division supplies the superior rectus, which it pierces to reach levator palpebrae superioris. The inferior division splits into several branches that supply the medial rectus and inferior rectus and a long branch passes forwards on the lateral aspect of inferior rectus to reach inferior oblique (Figs 1.46B and 1.53A). It is from this latter branch that the stout motor root (preganglionic parasympathetic fibres) passes to the

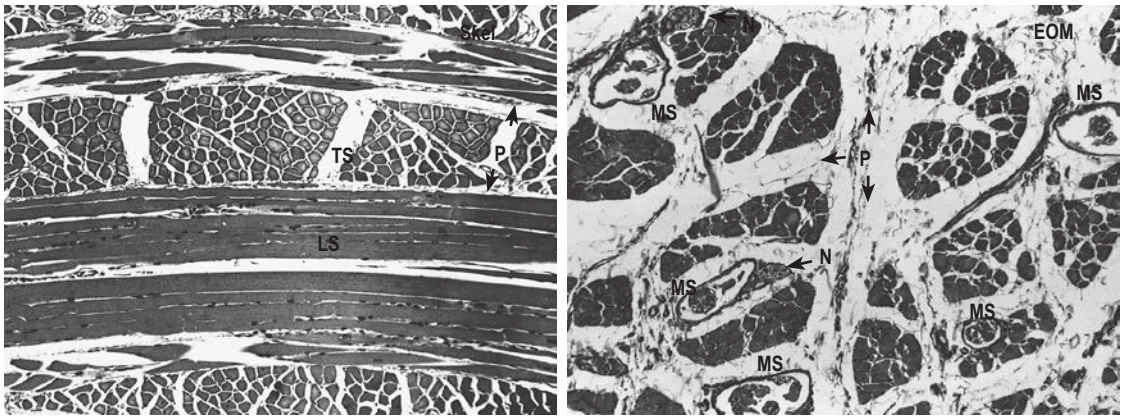


Fig. 1.49 Histological section of (**A**) normal human skeletal muscle fibres *(SKel)* and (**B**) extraocular muscle fibres *(EOM)*. *LS*, Longitudinal section of fibres; *MS*, muscle spindles; *N*, nerve; *P*, perimysium; *TS*, fibres in transverse section. Original magnifications: **A** and **B** × 100.

Table 1.3 **Classification of mammalian extraocular muscle fibres**

Type A	Type B	Type C
Large diameter	Intermediate diameter	Small diameter
Single end-plate	Multiple end-plates	Small *en grappe* plates
Fast twitch	Slow twitch	Tonic contractions
Required for saccadic movements	Needed for smooth pursuit movements	Function to align both visual axes, i.e. fine local contractions

ciliary ganglion, the site of postganglionic parasympathetic (second-order) neurones (Fig. 1.46A). Axons from the postganglionic neurones travel in the short ciliary nerves to supply the choroid, sphincter pupillae of the iris and the ciliary muscle.

The nerves that supply extraocular muscles generally pierce the muscle one-third of the way along the muscle belly on the bulbar aspect.

The complex lesions that can affect the oculomotor nerve and have functional consequences are outlined in Box 1.29 Clinical correlates.

Trochlear nerve (cranial nerve IV)

This is the only somatic efferent nerve to emerge from the posterior aspect of the central nervous system. It is also unusual in that it decussates before leaving the brainstem. It supplies only one extraocular muscle, the *superior oblique.*

Origin (Fig. 1.54A–C)

The nucleus lies in the anterior part of the periaqueductal grey matter at the level of the inferior colliculus (caudal to III nerve nucleus) in line with the other oculomotor nuclei. The fibres first pass anteriorly and laterally towards the tegmentum before turning and passing posteriorly around the periaqueductal grey matter and into the superior medullary velum (part of the roof of the fourth ventricle) where they decussate before emerging from the posterior surface of the brainstem in the posterior cranial fossa (Fig. 1.55). It is the only cranial nerve to decussate completely. The slender nerves emerge caudal to the inferior colliculi on the dorsal surface of the brainstem.

Intracranial and intracavernous course (Fig. 1.54 and 1.55)

Once the trochlear nerves emerge below the inferior colliculi, each of these slender nerves winds around the crura of the midbrain (cerebral peduncles) above the superior cerebellar artery and the pons and below the posterior cerebral artery. It continues anteriorly immediately beneath the free edge of the tentorium cerebelli (Fig. 1.53B). It pierces this dura and enters the lateral wall of the cavernous sinus beneath the oculomotor nerve (Figs 1.9C, 1.45B and 1.46B). The trochlear nerve then passes upwards, thus coming to lie above the oculomotor nerve before entering the orbit outside the tendinous ring in the lateral part of the superior orbital fissure.

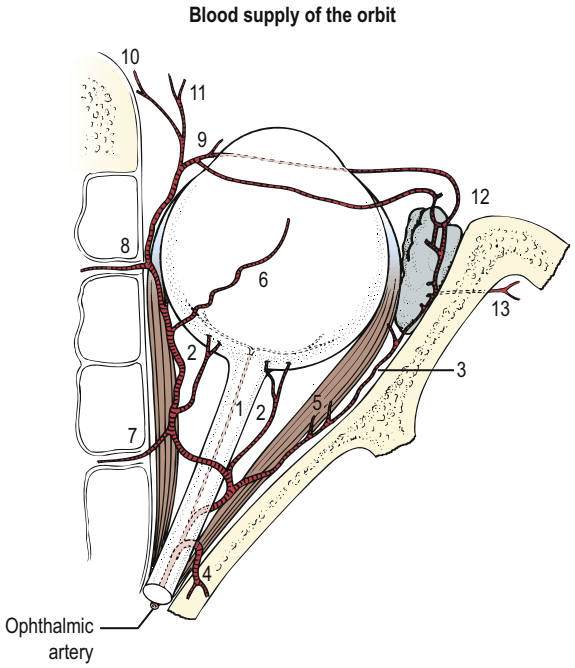


Fig. 1.50 Diagram summarizing the blood supply of the orbit as seen in a superior view: *1*, central retinal artery; *2*, posterior ciliary arteries that usually emerge as two trunks dividing into the short posterior ciliary arteries (seven or more) and the long posterior ciliary arteries (usually two, medial and lateral); *3*, lacrimal artery; *4*, recurrent branches (to meninges); *5*, muscular branches (give rise to anterior ciliary arteries); *6*, supraorbital artery; *7*, posterior ethmoidal artery; *8*, anterior ethmoidal artery; *9*, superior and inferior medial palpebral arteries; *10*, dorsalis nasi; *11*, supratrochlear; *12*, superior and inferior lateral palpebral arteries; *13*, zygomatic branches of the lacrimal artery.

Intraorbital course

It passes forwards above the origin of levator palpebrae superioris close to the bone, to enter the upper free edge of the superior oblique (Figs 1.45A,B, 1.47 and 1.55).

The intracranial lesions that can affect the trochlear nerve and their functional consequences are outlined in Box 1.30 Clinical correlates.

Abducent nerve (cranial nerve VI)

Origin (Figs 1.56 and 1.57)

The abducens nucleus lies in the mid-pons beneath the floor of the upper part of the fourth ventricle, close to the midline beneath the facial colliculus. The fibres pass ventrally (anteriorly) and caudally to emerge on the lower border of the pons in the pontomedullary junction near the midline (Figs 1.51 and 1.56).

Table 1.4 Sites of anastomosis between branches of the internal and external carotid arteries

External carotid branch	Internal carotid branch	Region of anastomosis
Angular artery (facial)	Dorsalis nasi (ophthalmic)	Medial palpebral margin
Transverse facial artery (superficial temporal)	Lacrimal artery (ophthalmic)	Lateral palpebral margin
Middle meningeal artery and deep temporal artery	Lacrimal artery (ophthalmic)	Orbit

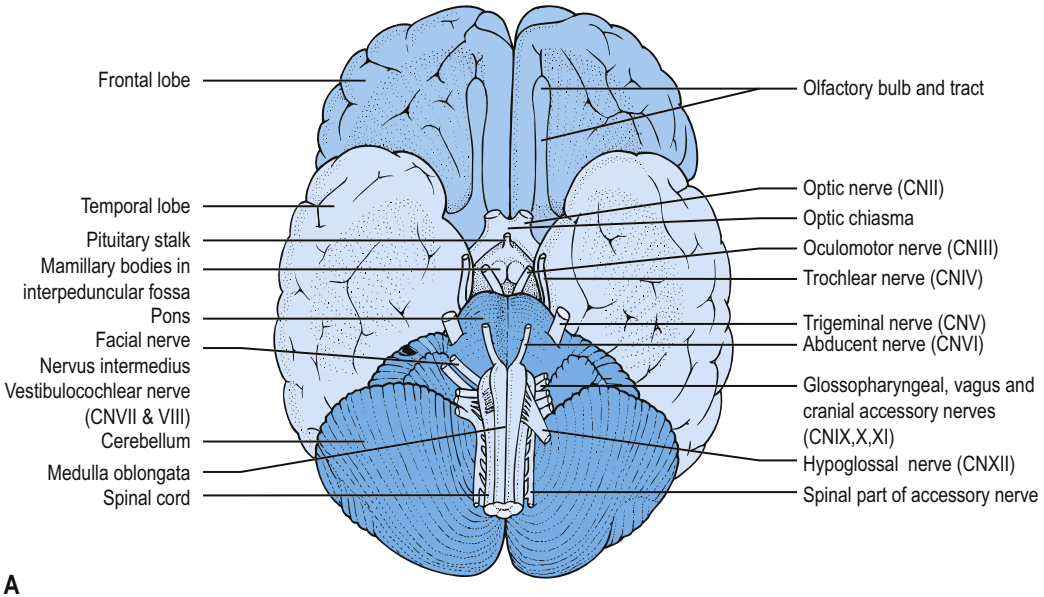
Intracranial and intracavernous course

The abducent nerve has the longest intracranial course of any cranial nerve. It courses upwards in the pontine cistern between the brainstem and the clivus, either side of the basilar artery. It is crossed or 'bound down' to the brainstem close to its origin by the anterior inferior cerebellar artery. It may pierce the dura early in its upward course upon the clivus close to the inferior petrosal sinus (2 cm below the posterior clinoid process) (Fig 1.52). On reaching the upper border of the apex of the petrous temporal bone, it crosses the inferior petrosal sinus from medial to lateral and changes direction sharply from a vertical to a horizontal course and runs forwards beneath the *petrosphenoidal ligament (of Gruber)* and *superior petrosal sinus* (Figs 1.56 and 1.57). The abducent nerve passes forward *within the cavernous sinus*, surrounded by venous spaces and suspended by fine connective tissue trabeculae. It lies *lateral* to the ascending portion of the internal carotid artery and then inferolateral to its horizontal portion (Fig. 1.9D,E). The abducent nerve enters the intraconal space of the orbit by passing within the tendinous ring (Figs 1.5 and 1.6).

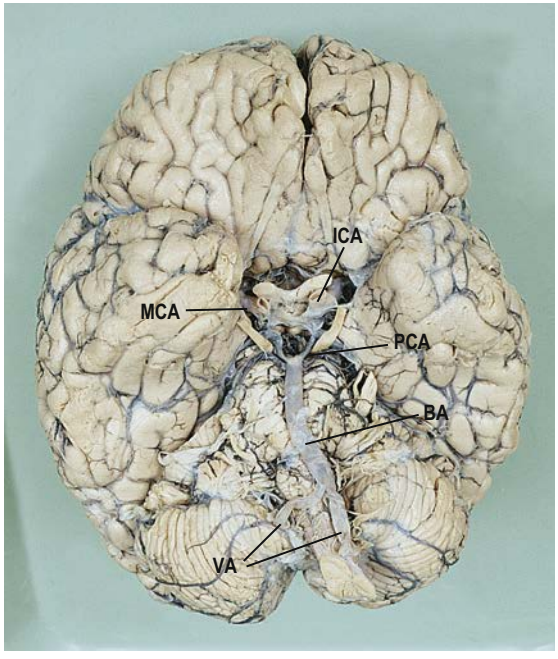
Intraorbital course

The abducent nerve has a short intraorbital course. It enters the bulbar surface of the lateral rectus one-third of the way from its origin (Fig. 1.57). This is the only muscle supplied by the abducent nerve.

The intracranial lesions that can affect the abducent nerve and their functional consequences are outlined in Box 1.31 Clinical correlates.



A



B

Fig. 1.51 Ventral views of the brain to demonstrate the origin of the cranial nerves: (**A**) diagram without vessels; (**B**) photograph of a whole brain with vessels. *BA*, Basilar artery; *ICA*, internal carotid artery; *MCA*, middle cerebral artery; *PCA*, posterior cerebral artery; *VA*, vertebral artery.

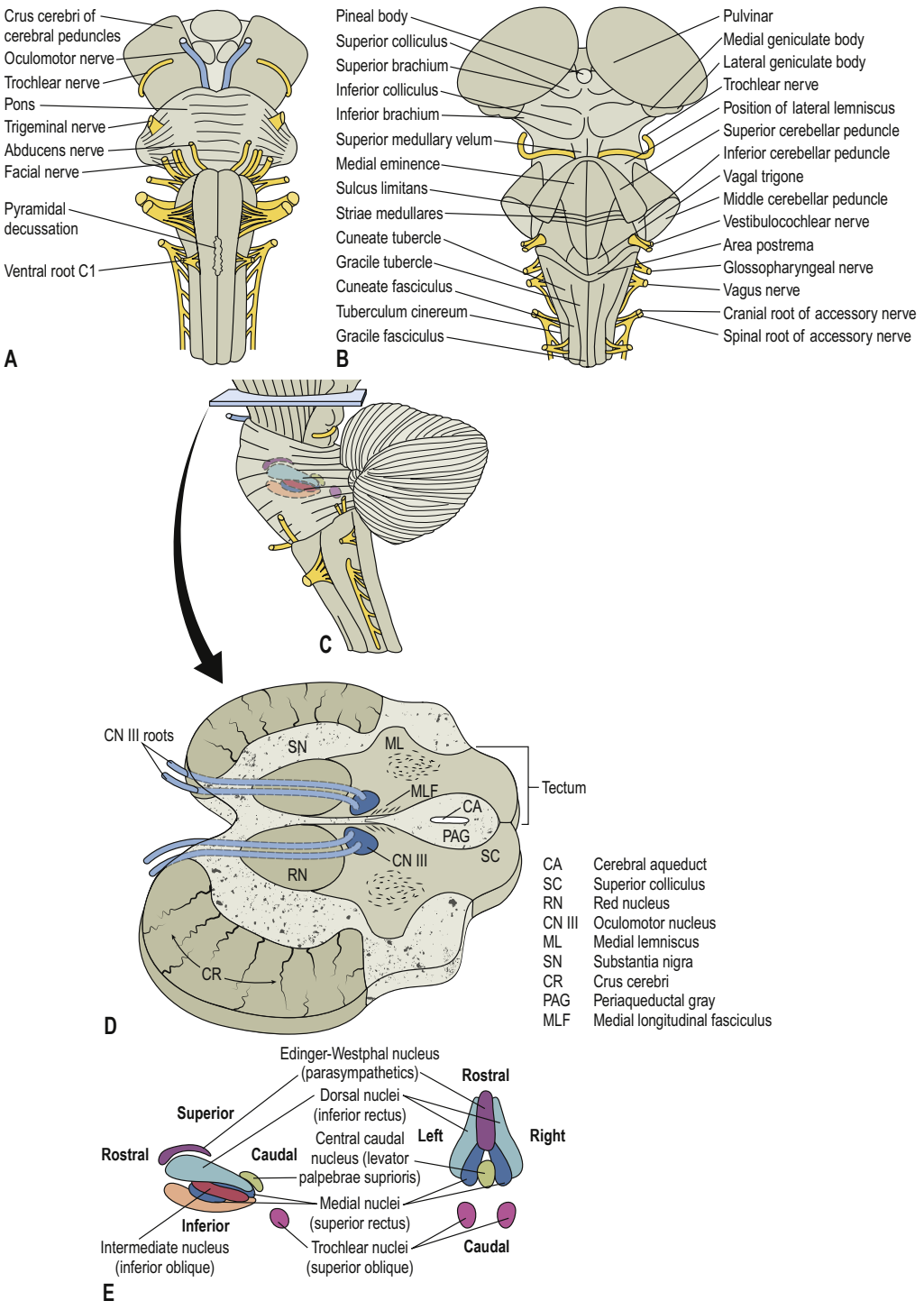


Fig. 1.52 (**A**) Ventral aspect of the entire brainstem showing the origin of the oculomotor nerve *(blue)*. (**B**) Dorsal view of the brainstem. (**C**) shows the level of sectioning of the brainstem (superior colliculus), which affords the view seen in the diagrammatic representation of a stained transverse section represented in (**D**). (**E**) Diagram of the arrangement of the oculomotor nuclei subserving the various extraocular muscles.

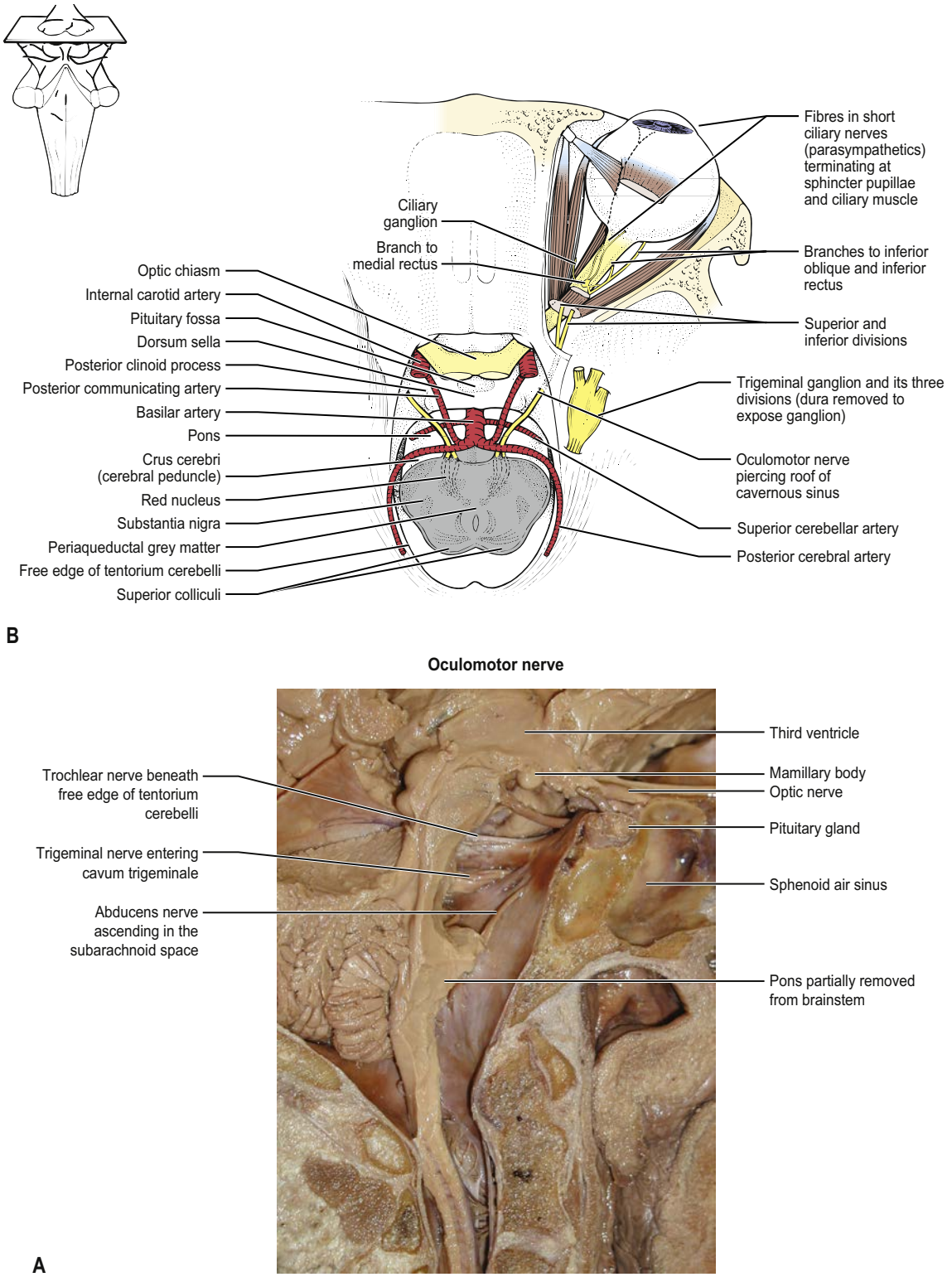


Fig. 1.53 (**A**) Diagram of the course of the oculomotor nerve from brainstem to the orbit. (**B**) Photograph of a hemisected head to reveal the brainstem in which the pons has been removed to reveal the origin and course of the oculomotor, trochlear and abducent nerves as well as the root of the trigeminal and the intracranial part of the optic nerve.

Box 1.28 **Clinical correlates**

Intracavernous lesions of cranial III nerve

The oculomotor, like other cranial nerves coursing through the cavernous sinus, can become involved in pathological processes such as venous thrombosis or aneurysms of the internal artery. Pituitary enlargements more commonly affect the oculomotor and trochlear nerves than the abducent, which is protected by the internal carotid artery. Meningioma or expanding lesions in the region of the superior orbital fissure can also compress the nerve.

Sensory endings in oculomotor nerves

The cell bodies of proprioreceptive fibres in the extraocular muscles are located in the mesencephalic nucleus of the trigeminal nerve. The mesencephalic nucleus also receives proprioreceptive terminals from neck and face musculature. The coordination of simultaneous movements of the head and eyes is dependent on conjugation of sensory (proprioceptive) information from the musculature of the neck and eyes, and input from cerebellar oculomotor centres (see Chapter 5).

Trigeminal nerve (cranial nerve V) (Figs 1.58 and 1.59)

The trigeminal is the largest of the cranial nerves. It is the main sensory nerve of the head, supplying the mucous membranes of the oronasal cavities, middle ear, paranasal sinuses, skin of the face, the teeth, the cornea, the temporomandibular joint, and the dura of the anterior and middle cranial fossae (Fig. 1.59A,B). The dermatomes corresponding to its three major subdivisions and the cutaneous branches of the ophthalmic and maxillary nerves are shown in Fig. 1.59A.

Trigeminal nerve and nucleus (Fig. 1.58A–D)

This nerve has GSA (general somatic afferent) and SVE (special visceral efferent) functions that manifest as a large nerve emerging from the mid pons of the brainstem in the posterior cranial fossa at the level of the middle cerebellar peduncle as a large *sensory root* and a small *motor root*. The GSA are concerned with sensation to the face in the three dermatomes (Fig. 1.59A and see later). The SVE fibres supply motor fibres to the muscles of mastication. The GSA component has cell bodies in the trigeminal ganglion.

There are three sensory nuclei in the brainstem associated with the trigeminal nerve. They are located in the brainstem in a vertical continuous series (Fig. 1.58B)

- The *mesencephalic nucleus* is the most rostral component and is the site for the first-order sensory neurones serving proprioreception from muscles of mastication and muscles of facial expression. It may also serve proprioreception for the extraocular muscles although this is still controversial. The neurone cell bodies lie at the level of the upper pons (not in trigeminal ganglion outside the CNS like other sensory neurone cell bodies). It is connected to the motor nucleus of V and is the basis for jaw-jerk reflex.
- The *main sensory or pontine nucleus* lies at the level of the mid pons (Fig. 1.58D), it processes pain, touch, pressure and temperature. Central processes from trigeminal sensory neurones entering the pons pass both in cranial and caudal directions. Conscious proprioreception, pressure and tactile sensations travel in the trigeminothalamic tract (akin to the medial lemniscus) to the VPN (ventral posteromedial nucleus) of the thalamus (ipsilateral fibres travel in the dorsal trigeminothalamic tract; the contralateral ones travel in the ventral trigeminothalamic tract).

The *spinal nucleus* and tract of trigeminal which may extend from the mid pons to the C2–4 level (Fig. 1.58B). It is continuous with the substantia gelatinosa of the posterior grey horn of the spinal cord. It receives pain and temperature sensation to the caudal aspect of the nucleus. It is connected via the ventral trigeminothalamic tract (akin to lateral spinothalamic tract) to the thalamus. It is concerned with tactile, nociceptive and thermal information from the territories of the three divisions.

Box 1.29 **Clinical correlates**

Lesions of the oculomotor nerve

Complete lesions of the oculomotor nerve (e.g. trauma) result in:

- inability to look upwards, downwards or medially
- lateral or *external strabismus* because of unopposed action of lateral rectus
- diplopia
- complete ptosis (paralysis of levator palpebrae superioris and unopposed orbicularis oculi)
- dilated non-reactive pupils (unopposed dilator pupillae)
- lack of accommodation.

Incomplete lesions: some of the symptoms above may be present.

Internal ophthalmoplegia: loss of parasympathetic components only. This may be the first sign of nerve palsy as the parasympathetic fibres are located superficially in the nerve and they may be damaged first in intracranial lesions; thus, pupil dilation is a crucial sign of compression within the cranial cavity following head injury.

External ophthalmoplegia: the loss of extraocular muscle supply.

Intracranial lesions affecting the oculomotor nerve: aneurysms of adjacent arteries around the brainstem may cause compression of the nearby nerve. Meningitis can involve the nerve along its course in the subarachnoid space.

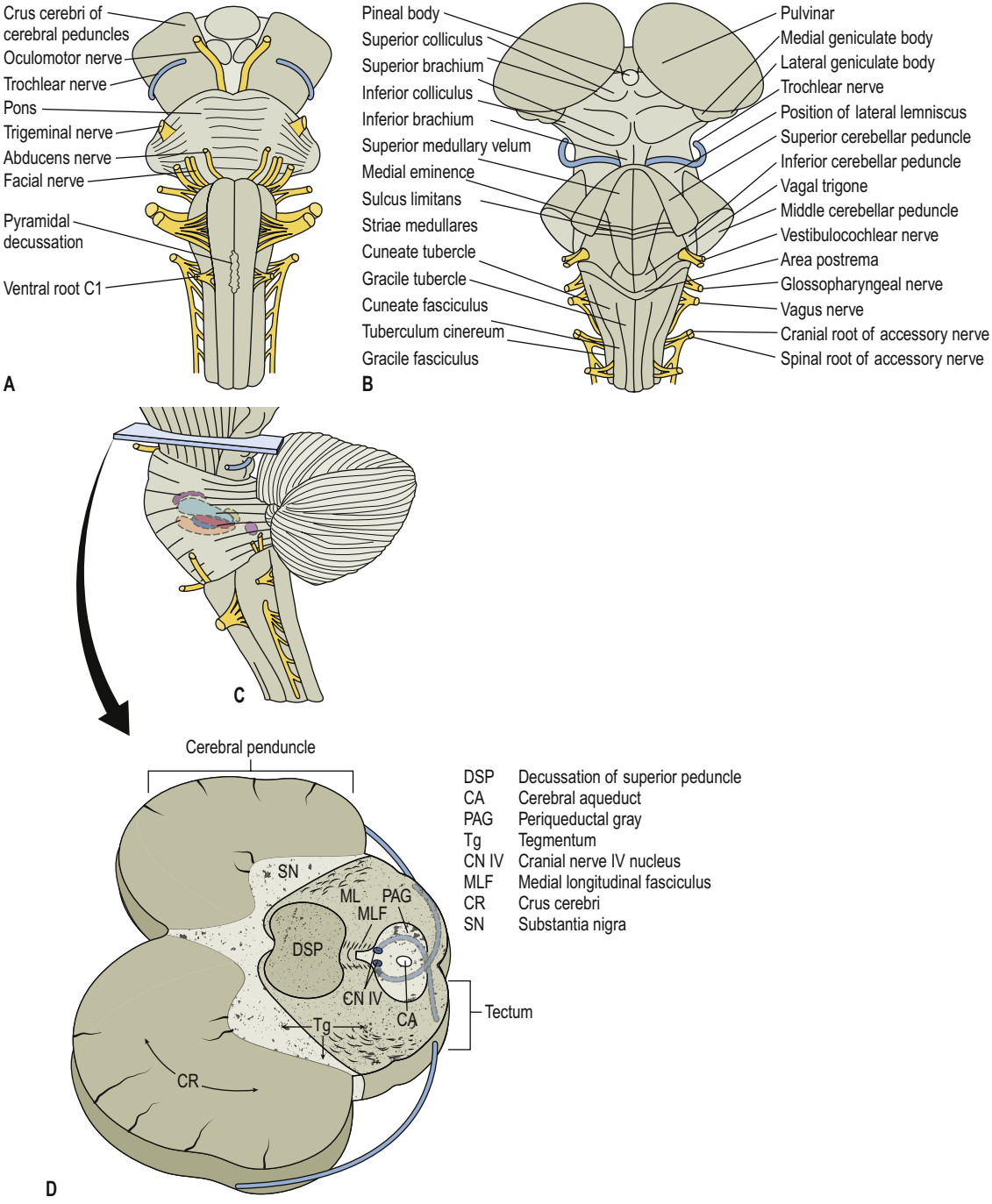


Fig. 1.54 (**A**) Ventral and (**B**) dorsal views of the brainstem showing the origin and course of the trochlear nerve *(blue)*. (**C**) The level of the brainstem section (inferior colliculus) that affords the view seen in the diagrammatic representation of a stained section in (**D**).

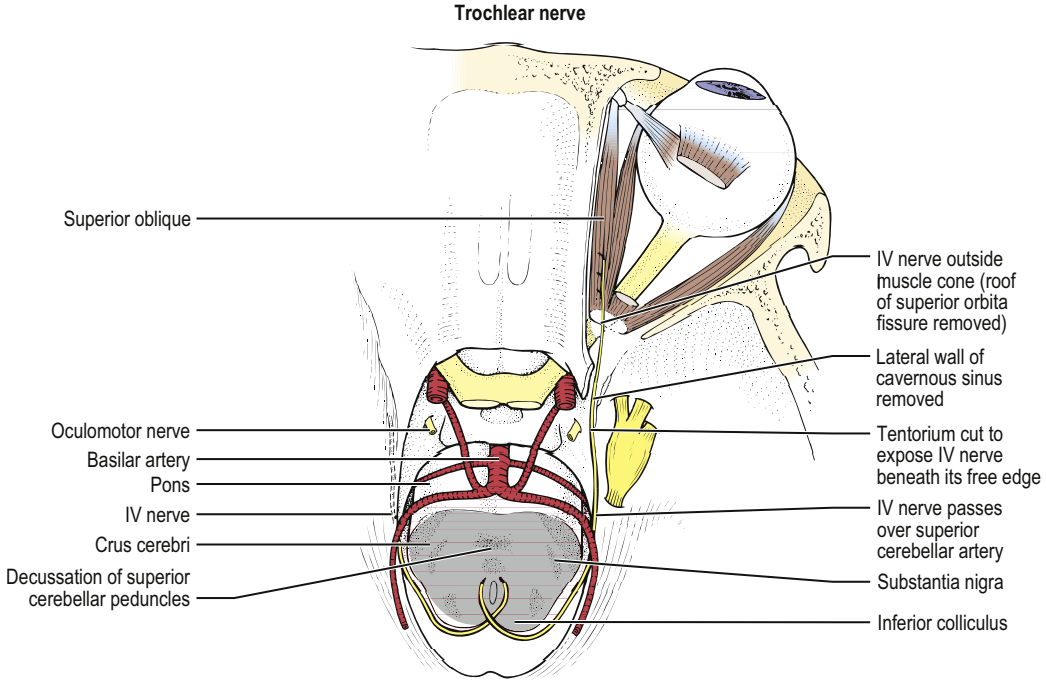


Fig. 1.55 Diagram of the course of the trochlear nerve from brainstem to the orbit.

Box 1.30 **Clinical correlates**

Intracranial lesions affecting the trochlear nerve

The trochlear nerve is rarely paralysed alone, although it is particularly vulnerable at its posterior exit from the brainstem and as it winds round the midbrain. Lesions causing compression on the undersurface of the tentorium may affect the trochlear nerve. It may also be involved in pathological processes in the cavernous sinus. Patients suffering paralysis of the superior oblique because of trochlear nerve lesions suffer diplopia when looking down and have difficulty in looking down when the eye is adducted because the superior oblique is the only depressor in the adducted state. Patients characteristically carry the head tilted to the non-affected side with the chin lowered to compensate for the overaction of the inferior oblique producing unopposed torsion on the eye.

The afferent fibres terminating in the last two nuclei are the central processes of sensory neurones whose cell bodies are located in the trigeminal (Gasserian) ganglion. The peripheral processes terminate in appropriate sensory receptors in the territories of the three divisions (Fig. 1.59A,B).

The *motor nucleus* is located medial to the main sensory nucleus (Fig. 1.58D). It receives input from both cerebral hemispheres, the reticular formation, the red nucleus, the tectum, the mesencephalic nucleus and the medial longitudinal fasciculus (MLF). Its motor fibres exit as the motor root of the trigeminal and supply the muscles of mastication.

The divisions of the trigeminal nerve

The trigeminal nerve arises from the brainstem in the posterior cranial fossa as a large *sensory root* and a small *motor root.* The sensory root consists of the central processes of pseudounipolar sensory neurones whose cell bodies lie in the large *trigeminal ganglion* located in the cavum trigeminale, a bony depression near the apex of the petrous temporal bone lined by evaginating dura mater from the edge of the tentorium cerebelli and roofed over by dura of the middle cranial fossa. The ganglion is partly surrounded by cerebrospinal fluid, which is continuous with the subarachnoid space of the posterior cranial fossa. The ganglion is homologous to a dorsal root or sensory ganglion of a spinal nerve. It is from the anterolateral convex surface of this flattened ganglion that the three named branches emerge: the ophthalmic (V_1), maxillary (V_2) and mandibular (V_3) (Fig. 1.59B).

The *ophthalmic division* or nerve (Fig. 1.60A) splits in the anterolateral portion of the cavernous sinus into three main branches: the *lacrimal, frontal* and *nasociliary.* The pathway and termination of these nerves are summarized in Figs 1.45–1.47, 1.59B and 1.59B.

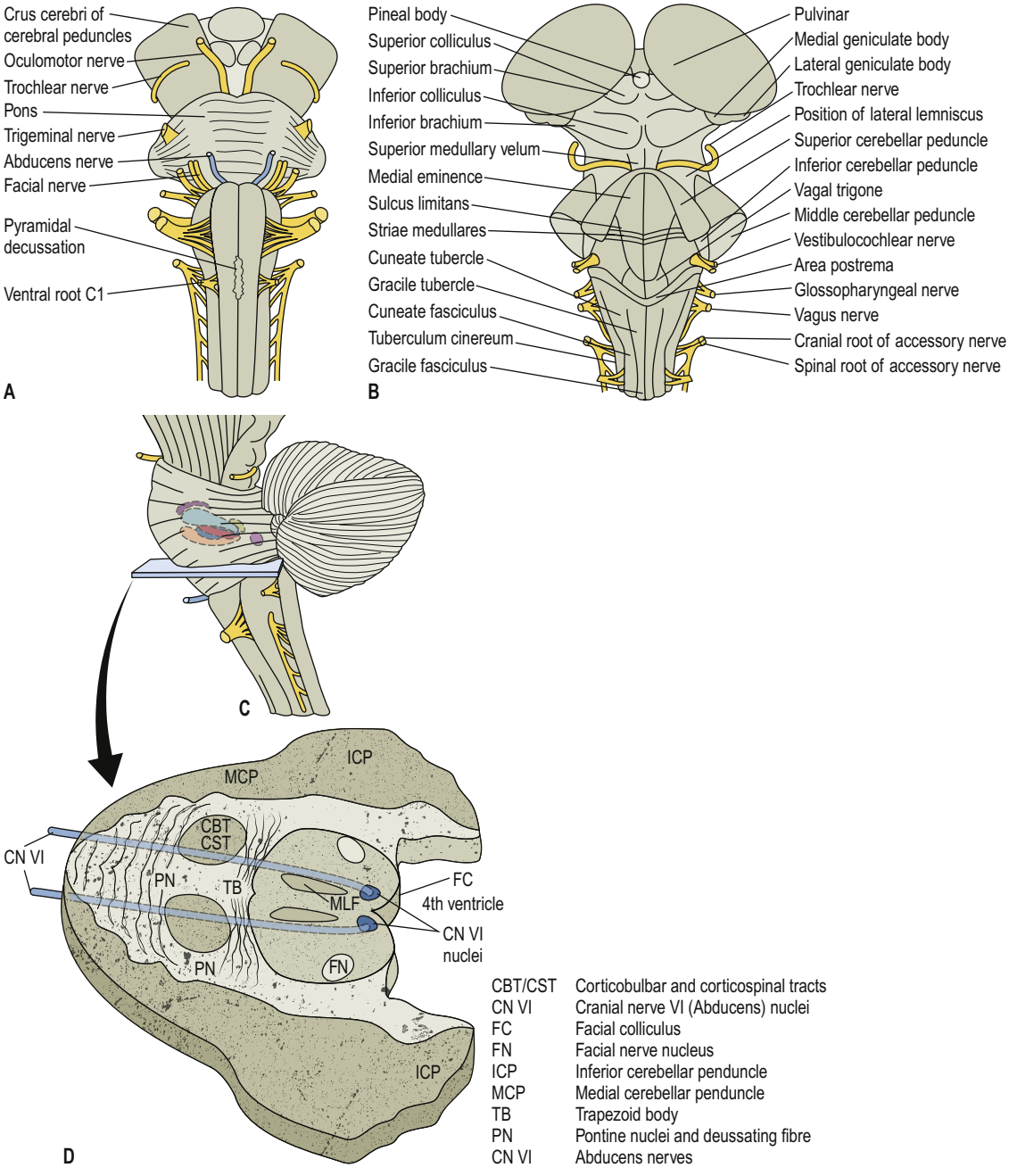


Fig. 1.56 (**A**) Ventral aspect and (**B**) dorsal view of the brainstem showing the origin and course of the abducent nerve *(blue)*. (**C**) The level of the brainstem section (mid-pons) which affords the view seen in the diagrammatic representation of a stained section in (**D**).

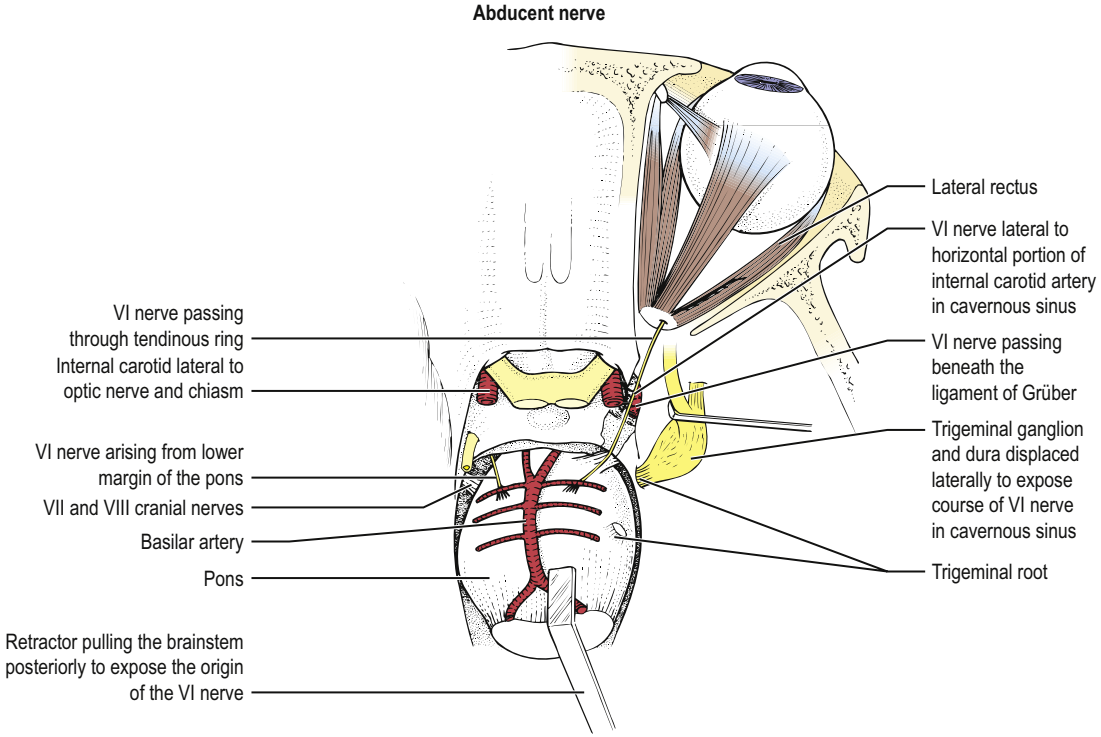


Fig. 1.57 Diagram of the course of the abducent nerve from brainstem to the orbit.

Box 1.31 **Clinical correlates**

Intracranial lesions affecting the abducent nerve

The abducent nerve is considered the weakling of the cranial contents. It is very susceptible to damage in head injuries, such as fractures to the base of the skull or any type of expanding cerebral lesion, owing to its long intracranial course. If the brainstem is displaced downwards (as a result of raised intracranial pressure), the nerve may be compressed against the inferior cerebellar artery or severed where it bends sharply over the apex or crest of the petrous temporal bone. Within the *cavernous sinus* the nerve, not being protected by the dura of the lateral wall, is more susceptible than the other intracavernous nerves. For example, atheromatous changes in the internal carotid artery may compress the abducent nerve. Lesions of the abducent nerve result in paralysis of lateral rectus; thus, the patient is unable to abduct the eye and suffers esotropia (*internal strabismus*) as a result of the unopposed action of the medial rectus.

The *maxillary nerve* (Figs 1.59B and 1.61) passes through the *foramen rotundum* and spans the *pterygopalatine fossa* before entering the orbit through the *inferior orbital fissure* as the *infraorbital nerve*. It lies beneath the periorbita and is thus not truly an orbital content. The nerve passes forward from the inferior orbital fissure to the *infraorbital groove*, which becomes the *infraorbital canal*, the nerve eventually emerging through the *infraorbital foramen* (Fig. 1.61). Here it radiates out as a number of cutaneous branches supplying the lower eyelid, the nose, upper lip and cheek (Fig. 1.59A). The *zygomatic nerve*, a branch of the infraorbital, runs along in the inferior orbital fissure, beneath the periorbita, to the lateral wall of the orbit where it pierces the zygomatic bone as two branches, the *zygomaticotemporal* and *zygomaticofacial* nerves (both cutaneous). Traditionally, a communicating branch is described as passing up the lateral wall of the orbit to join the lacrimal nerve; however, the presence and importance of this nerve have been disputed. Other branches of the maxillary nerve, which pass through the *pterygopalatine ganglion* (without synapsing), supply the nasal cavity, upper alveolar arch and hard and soft palate (Fig. 1.59B). More details of these branches and those of the mandibular division are provided in standard anatomical texts.

The *mandibular division*, besides its sensory component, supplies motor fibres to the muscles of mastication and

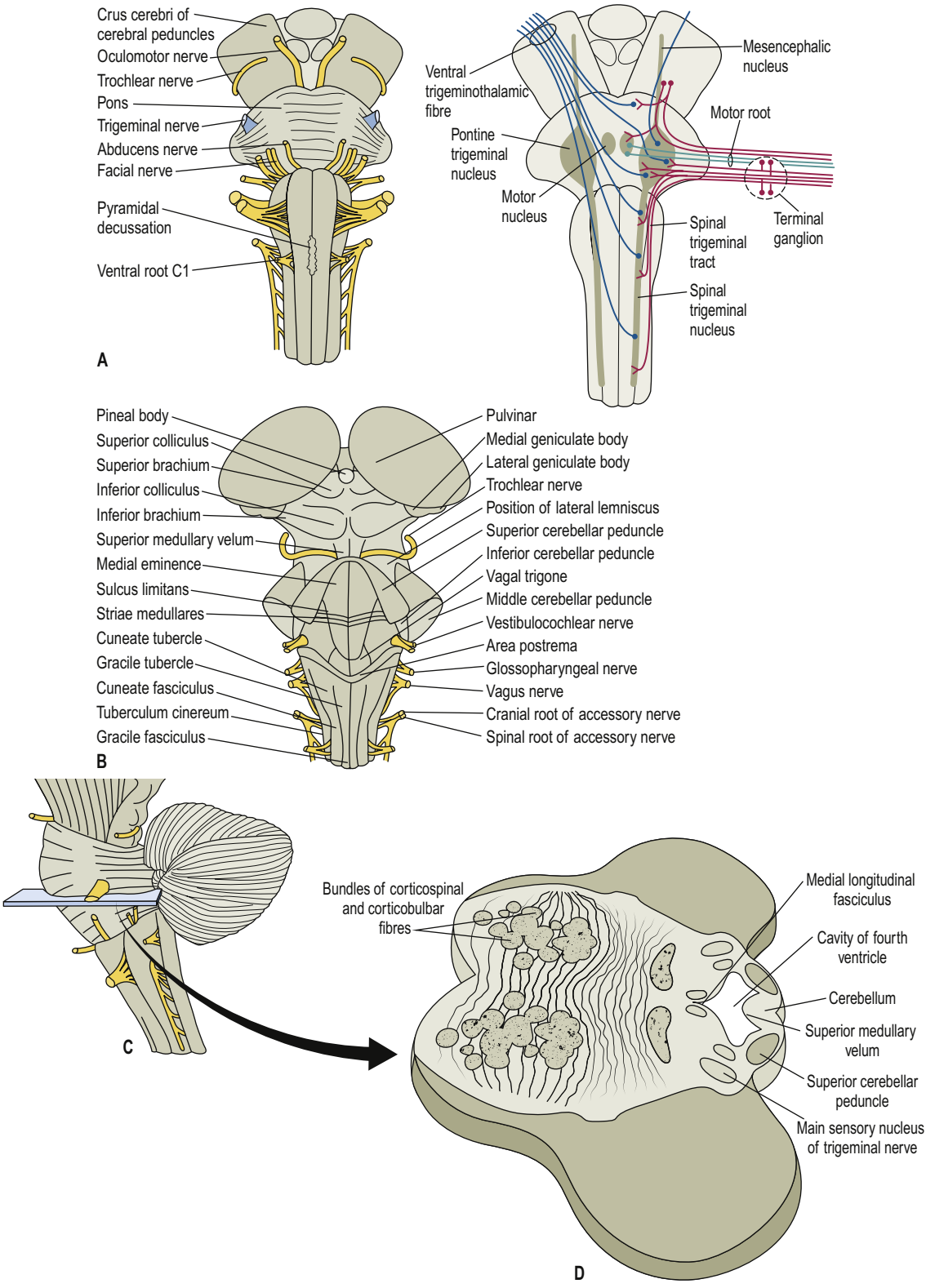


Fig. 1.58 (**A**) Ventral and (**B**) dorsal views of the brainstem showing the origin and course of the trigeminal nerve *(blue)*. In (**B**) the central pathways and nuclei of motor and sensory components of the nerve are shown. (**C**) shows the level of section in the mid-pons to illustrate where the cross-sectional diagram in (**D**) is taken from, which includes the superior and middle cerebral peduncles and the upper part of the fourth ventricle.

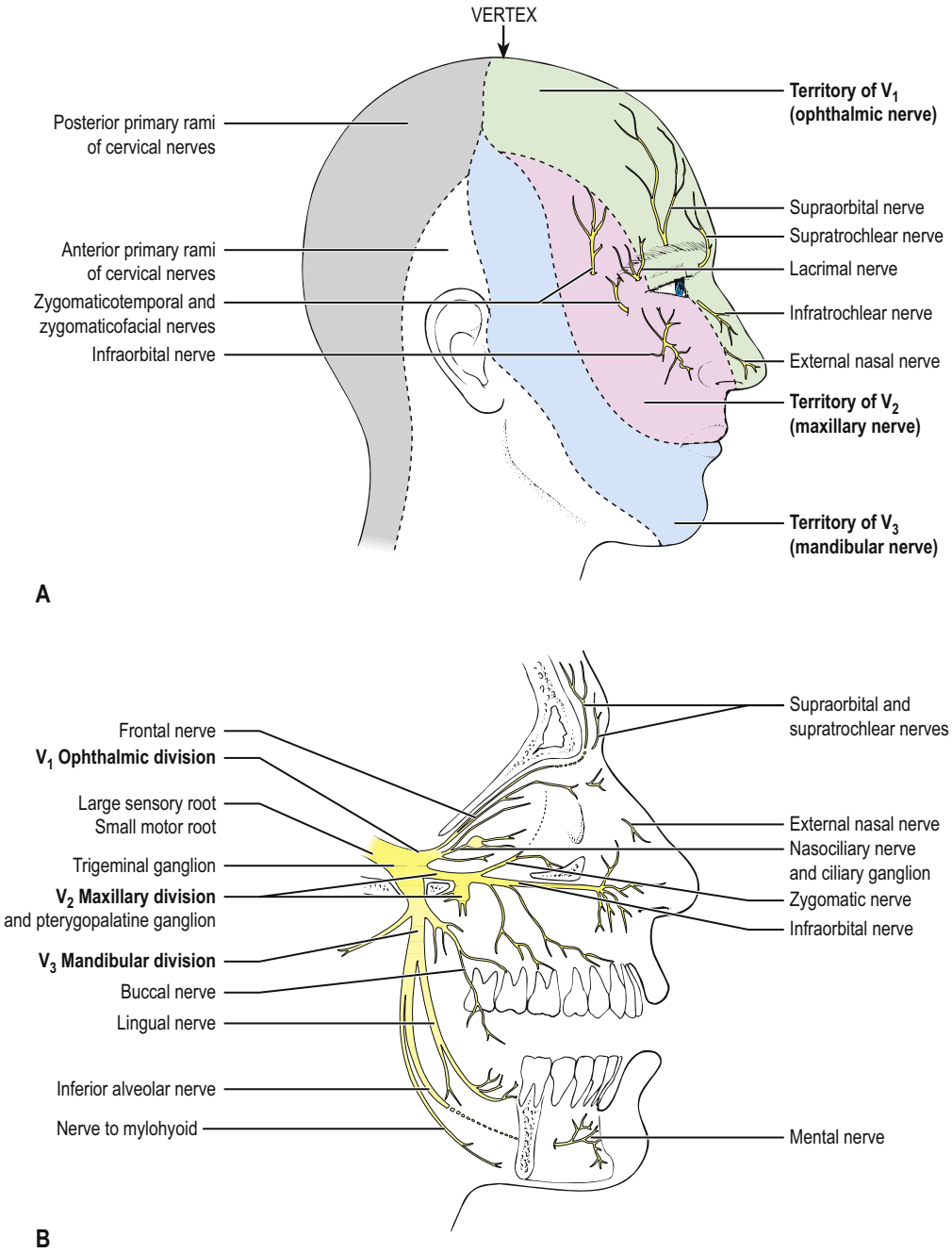


Fig. 1.59 (**A**) Sensory dermatome map of the head and neck. Note the limits of the territories of the skin of the face and scalp supplied by the three divisions of the trigeminal nerve (ophthalmic, maxillary and mandibular). Only the cutaneous branches of the former two are illustrated. (**B**) Stylized diagram showing the origin of the motor and sensory roots of the trigeminal, the position of the pterygopalatine and ciliary ganglia and the territories of some of the major branches of the three divisions.

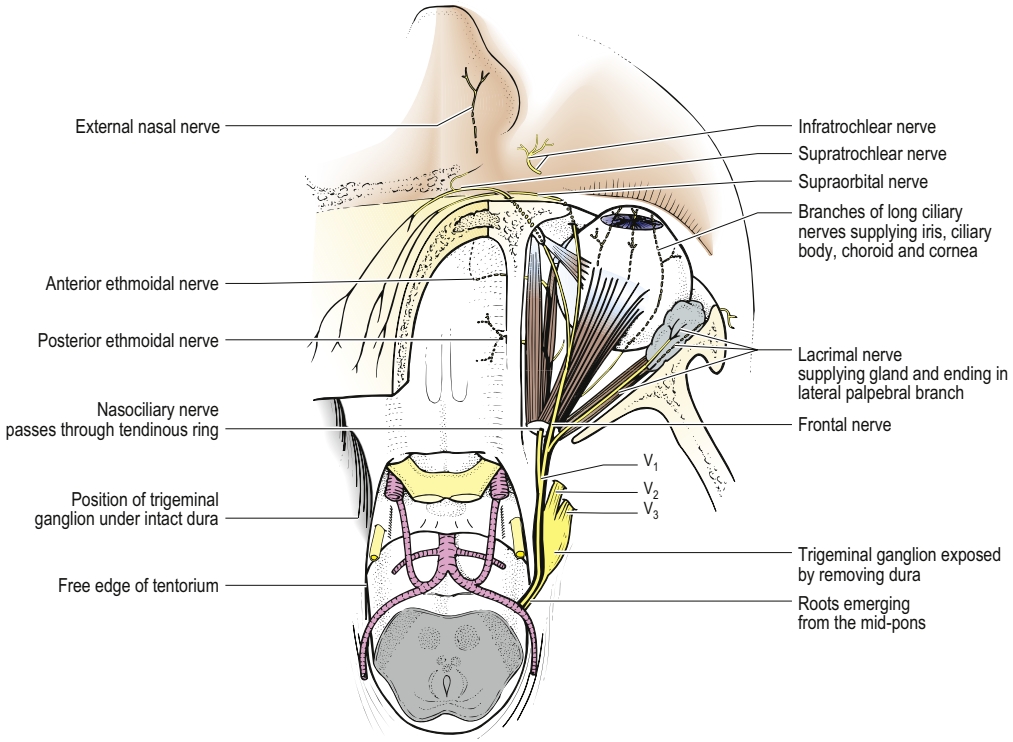


Fig. 1.60 Diagram summarizing the origin, intracranial, intracavernous and intraorbital course of the ophthalmic division of the trigeminal and its three main branches.

- *The frontal nerve* passes obliquely above the levator palpebrae superioris and divides into the supraorbital nerve and the supratrochlear nerve.
  - The supraorbital nerve passes through the supraorbital notch or foramen and ascends in the subcutaneous tissue of the forehead to supply the skin.
  - The supratrochlear nerve passes along the medial wall of the orbit to pass above the trochlea and supply the skin at the root of the nose, upper eyelid and the medial part of the forehead.
- The *lacrimal nerve* travels forward along the upper border of the lateral rectus and sends fibres to the gland before ending as *lateral palpebral branches* to the conjunctiva and skin of this region.
- The *nasociliary nerve* branches include the *sensory root* to the ciliary ganglion (from which the short ciliary nerves emerge), *long ciliary nerves, posterior ethmoidal nerve* (supplies ethmoidal and sphenoidal sinuses), the *anterior ethmoidal nerve* and *infratrochlear nerves*.
  - The anterior ethmoidal briefly re-enters the cranial cavity at the cribriform plate (beneath the dura) before piercing the bone to exit the cavity to terminate as the *medial* and *lateral internal nasal branches* (supply the nasal cavity), the latter of which ends as the *external nasal branch* that supplies the skin on the lower half of the nose.
  - The infratrochlear nerve runs close to the medial orbital wall, passes beneath the trochlea and supplies the skin at the angle of the eye and upper part of the skin of the nose.

also receives proprioceptive fibres from these muscles and the muscles of facial expression in the lower part of the face.

The importance of understanding the distribution of the territories of the trigeminal nerve is illustrated by the condition of trigeminal neuralgia (Box 1.32 Clinical correlates).

The four parasympathetic ganglia of the head and neck (ciliary, pterygopalatine, otic and submandibular) are associated with the branches of the trigeminal (Fig. 1.59B). They are generally connected by short stalks containing pre- and postganglionic fibres and the terminal fibres are distributed with the branches of the trigeminal; they might share the same perineural sheath for part of their course ('hitch-hikers'). Sympathetics may also hitch-hike with branches of the trigeminal for part of their course and be distributed to the territories of this large nerve.

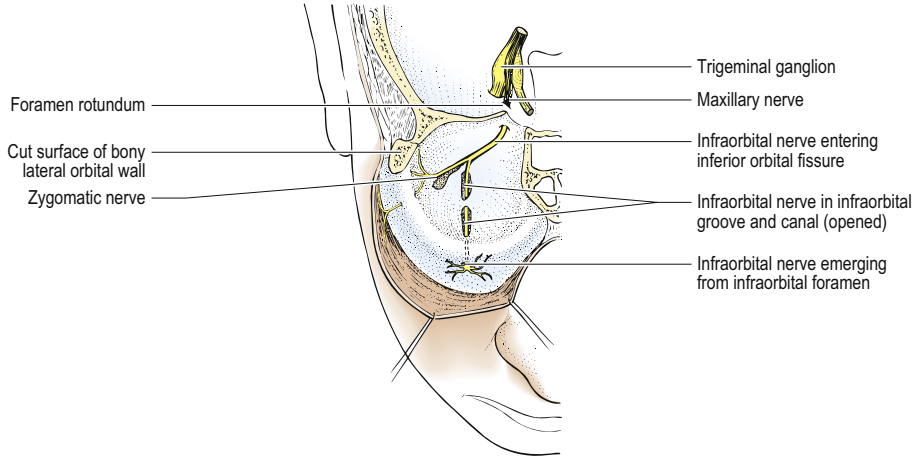


Fig. 1.61 Diagram of the orbital floor and the course of the maxillary and infraorbital nerves.

Box 1.32 **Clinical correlates**

Trigeminal neuralgia ('tic douloureux')

This is a condition characterized by excruciating pain in the territory of one or more of the divisions of the trigeminal nerve. Thus, the clinician should be familiar with structures supplied by each division. Many causes have been suggested, including osteitis of the petrous temporal bone or compression of the root or ganglion in the cavum trigeminale by enlarged or engorged vessels; however, in many cases the aetiology is unknown. The territory of the maxillary nerve is most frequently involved, then the mandibular, and less commonly the ophthalmic nerve. Three commonly performed surgeries are glycerol rhizotomy, stereotactic radio-surgery (Gamma Knife) and endoscopic vascular decompression.

Facial nerve (cranial nerve VII) (Figs 1.62 and 1.63)

Each facial nerve contains a number of functional components (see eTable 1.1). The *facial nerve proper* (containing *branchiomotor* or *special visceral efferents*) supplies the muscles of facial expression, stapedius (small muscle in middle ear), stylohyoid and the posterior belly of the digastric. The second component, the *nervus intermedius*, contains secretomotor or *parasympathetic* fibres (*general visceral efferent*), which synapse in the pterygopalatine ganglion and supply the lacrimal gland and choroid in addition to other glands in and around the nose and mouth. The nerve also transmits taste fibres (*special visceral afferents*) from the anterior two-thirds of the tongue.

Origin (Fig. 1.62)

The main motor nucleus (*branchiomotor* or *SVE)* lies in the ventrolateral aspect of the tegmentum in the lower pons and is the most medial of the three nuclei. Its efferent fibres take an unusual course and pass dorsomedially around the abducent nucleus creating a small elevation (the facial colliculus), which can be seen macroscopically in the floor of the IV ventricle. Fibres then pass laterally to the abducens nucleus before coursing ventrally and caudally through the body of the pons to emerge at the pontomedullary junction lateral to the abducent nerve. The fibres then pass through the internal acoustic meatus and through the petrous temporal bone in the facial canal. These fibres of cranial nerve VII exit the skull through the *stylomastoid foramen* and pierce the parotid gland to emerge at the anterior border of that gland and radiate on the face to supply the muscles of facial expression (Fig. 1.63). The clinical importance of corticonuclear input to the upper and lower part of the nucleus in understanding lesions of the facial nerve and their clinical consequences is outlined in Box 1.33 Clinical correlates.

The parasympathetic (*GVE*) nucleus lies lateral to the main motor nucleus. It has two components, the superior salivatory nucleus (SSN) and the lacrimal nucleus (LN). The SSN is located in the reticular formation of the lower pons. The fibres pass ventrolaterally in the pons and emerge as the nervus intermedius. The preganglionic fibres terminate in submandibular ganglion. From here, postganglionic parasympathetic fibres supply the submandibular and sublingual glands. Fibres from the more rostral LN travel in the petrosal nerve to reach the pterygopalatine ganglion. Postganglionic parasympathetic fibres from this ganglion supply the lacrimal, nasal and palatine glands. The LN has input from the hypothalamus (emotional responses) and sensory nucleus of the trigeminal (reflex lacrimation in

eTable 1.1 **Functional analysis of cranial nerve components**

Functional classification	Modality/target (ontogeny/ phytogeny)	Present in cranial nerves
Somatic efferent (general motor)	Supplies skeletal muscle of somatic origin (preotic somites: extraocular muscles; occipital somites: tongue musculature)	III, IV and VI XII
Somatic afferent (general sensory)	Pain, temperature and touch. Supplies skin and mucous membranes of the head and neck	Predominantly in V but several minor elements in VII, IX and X
General visceral efferents[a] (parasympathetic)	Supplies smooth muscle (viscera), cardiac muscle, glands, blood vessels and intrinsic eye muscles (ciliary muscle and sphincter pupillae)	III, VII, IX, X and XI X (vagus) is largest parasympathetic nerve in the body
General visceral afferents	Pain and sensibility of viscera	VII, IX and X
Special visceral efferents (branchiomotor)	Skeletal muscles of mastication and facial expression (i.e. pharyngeal arch or visceral evolutionary origin)	V, VII, IX, X and XI
Special somatic afferent (special senses concerned with body position, excluding vision)	Maintenance of balance and reception of sound (vestibulocochlear organ)	VII
Special visceral afferent (special visceral senses, taste and smell)	Olfactory epithelium in nasal cavity and taste receptors in tongue and palate	Olfaction (smell) in I; taste in VII, IX and X

[a]Sympathetic nerve fibres, originating from upper thoracic segments of the spinal cord and synapsing in cervical ganglia may hitch-hike with various cranial nerves and/or blood vessels to reach ocular and orbital structures, e.g. dilator pupillae muscle and tarsal muscle (see Fig. 1.65).

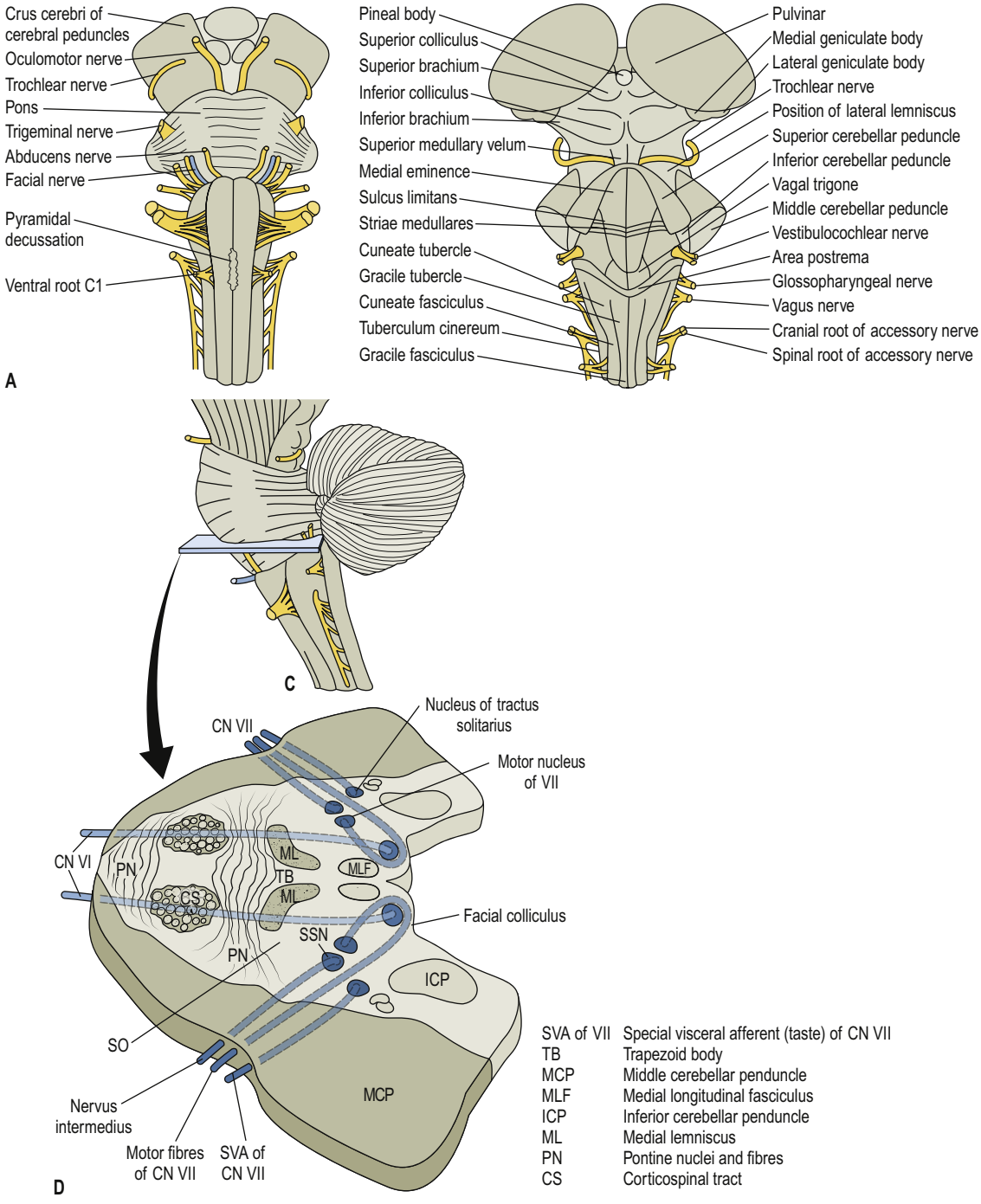


Fig. 1.62 (**A**) Ventral and (**B**) dorsal views of the brainstem showing the origin and course of the facial nerve *(blue)*. (**C**) shows the level of section in the lower pons to illustrate where the cross-sectional diagram in (**D**) is taken from which is at the level of the facial colliculus in the floor of the fourth ventricle.

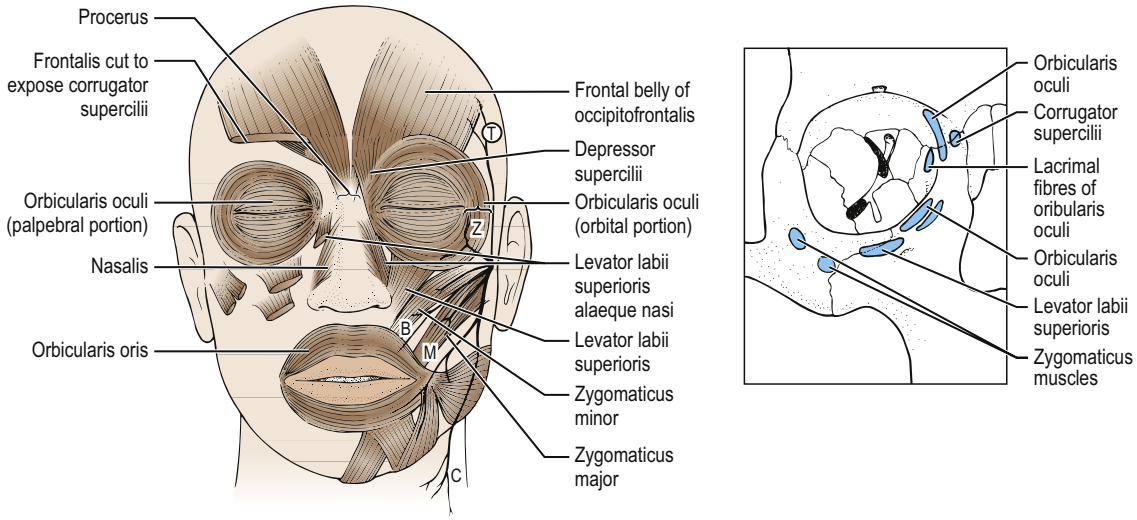


Fig. 1.63 The muscles of facial expression, particularly those of relevance to the eye and orbit. *Inset* shows the bony origins of some of these muscles around the orbit. Some of orbicularis oculi and frontalis muscle fibres have been removed on the right side of the face to display underlying musculature. The extracranial motor branches of the facial nerve that supply the muscles are shown on the right: *C*, Cervical branches *B*, buccal branches; *M*, marginal mandibular branch; *T*, temporal or frontal branches; *Z*, zygomatic branches.

response to irritation of the cornea and conjunctiva) (see Chapter 4, p. 207).

The sensory nucleus (*special visceral afferents*) lies close to the motor nucleus. It receives input from the sensory neurons located in the geniculate ganglion of the facial nerve, which subserve sensation from the anterior two-thirds of the tongue. Efferents from the nucleus decussate and ascend to the thalamus. Signals are relayed from there to the postcentral gyrus of the cerebral cortex.

The facial nerve is important to the eye and orbit primarily because of its parasympathetic supply to the lacrimal gland (and some intraocular branches) and its motor supply to the periorbital facial muscles (especially orbicularis oculi). It is the most frequently paralysed of all peripheral nerves. (Lesions of the facial nerve are discussed in Box 1.33 Clinical correlates.)

There are three reflex arcs in the brainstem involving the facial nerve, of which the corneal reflex is an important clinical test (Table 1.5).

Ocular appendages (adnexa)

Muscles of the eyelids and adjacent face

The muscles that are primarily responsible for movement of each eyelid include the *orbicularis oculi* (a muscle of facial expression), which is responsible for lid closure, and the *levator palpebrae superioris* (an extraocular muscle), which raises the lid.

Other *muscles of facial expression* of primary interest around the eye and eyelids are the *corrugator supercilii* and *occipitofrontalis.*

Orbicularis oculi (Fig. 1.63)

Shape. The orbicularis oculi muscle is a broad, flat sheet of skeletal muscle with orbital, palpebral and lacrimal portions. The circular orientation of the fibres is a reflection of the sphincter-like function of this muscle. The *orbital portion* arises from the medial palpebral ligament and the adjacent orbital margin (Fig. 1.63, inset). Its fibres run circumferentially in an elliptical fashion around and beyond the orbital margin. Most pass round the lateral orbital margin without interruption, although some fibres (known as *depressor supercilii*) are inserted into the skin and connective tissue of the eyebrow. The *palpebral part* is an extremely thin muscle that originates from the *medial palpebral ligament.* Its fibres pass laterally within the eyelid anterior to the orbital septum and tarsal plate (see later), and interlace to form the *lateral palpebral raphe.* The small *lacrimal component* of the muscle passes deep to the medial palpebral ligament and is attached to the posterior lacrimal crest (behind the lacrimal sac) as two muscle slips (upper and lower). These fibres are inserted laterally into the tarsi close to the lacrimal canaliculi; they help draw the eyelids and lacrimal papillae medially and in addition dilate the lacrimal sac during blinking. This helps to suck tears into the lacrimal punctum from the lacus lacrimalis.

Box 1.33 **Clinical correlates**

Lesions of the facial nerve

Supranuclear lesions caused by vascular stroke in which descending corticonuclear and corticospinal fibres are damaged in the internal capsule. The upper facial motor nucleus (supplying upper half of facial muscles) receives input from 'face' areas of both the ipsilateral and contralateral motor cortices. The lower part of the facial nucleus has only contralateral input. The effect of a stroke, therefore, is to cause contralateral paralysis or weakness of the limbs and lower face. The upper face survives because of the bilateral supranuclear supply to the upper part of the facial nucleus.

Nuclear lesions: direct damage to the facial nucleus, such as thrombosis of the pontine branches of the basilar artery, results in complete paralysis of structures supplied by the facial nerve (and abducent nerve; see Fig. 1.54) together with motor weakness of the limbs on the opposite side (owing to pyramidal decussation occurring below this level).

Infranuclear lesions: Bell's palsy involves direct neuritis of the facial nerve in the bony canal within the temporal bone and results usually in complete facial paralysis.The patient is unable to move their lips (saliva and food drools from the corner of the mouth) or eyebrows or to close their eyelids (lids may be lax, causing epiphora) and suffers hyperacusis (resulting from paralysis of the stapedius). Some patients may also have reduced lacrimal and salivatory secretions and loss of taste to the anterior two-thirds of the tongue. Other causes of infranuclear lesions include multiple sclerosis, tumours of the cerebellopontine angle (acoustic neuromas), middle ear disease and tumours of the parotid gland.

Table 1.5 **Reflexes involving the facial nerve**

	Corneal reflex	Blinking to light or fast-approaching object	Blinking to noise
Receptor	Sensory ending in corneal epithelium	Retina	Cochlea
Afferent pathway	Long ciliary nerves, nasociliary nerve, ophthalmic nerve	Optic nerve	Vestibulocochlear nerve
First synapse	Spinal nucleus of trigeminal	Superior colliculus	Inferior colliculus
Second synapse	Facial nucleus	Facial nucleus	Facial nucleus
Efferent pathway	Temporal and zygomatic branches of facial nerve	Temporal and zygomatic branches of facial nerve	Temporal and zygomatic branches of facial nerve
Effector muscle	Orbicularis oculi	Orbicularis oculi	Orbicularis oculi

Nerve supply. Temporal and zygomatic branches of the facial nerve.

Action. The *orbital* portion, owing to its elliptical form and medial attachments, acts like a purse string, drawing the skin of the forehead, temple, cheek and orbital margin towards the medial angle of the orbit, firmly closing the lids (for example, when in very bright light). The *palpebral* portion of orbicularis oculi can act under both voluntary and involuntary control to close the eyelids during normal blinking (and sleeping). The blink reflex (Table 1.5) is essential to the integrity of the ocular tear film and function of the cornea.

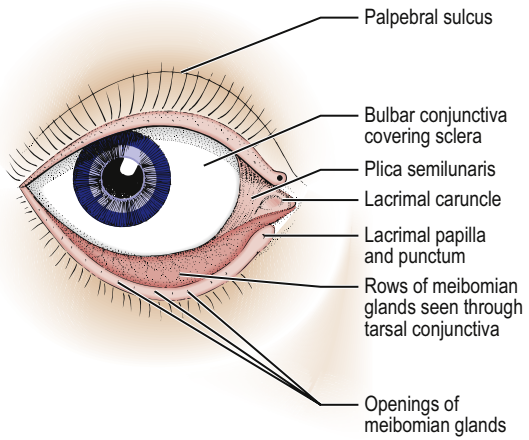


Fig. 1.64 Surface anatomy of the eye and eyelids. The bottom lid has been slightly everted to reveal the inner surface of the lid, lacrimal papillae and puncta and the openings of the meibomian glands along the lid margin.

Corrugator supercilii

This is a small pyramidal muscle at the medial aspect of the eyebrow (Fig. 1.58), beneath the occipitofrontalis and

orbicularis oculi. It draws the eyebrow downwards and medially (frowning), producing vertical skin furrows on the forehead. It assists in protecting the eyes in bright light. It is supplied by small subdivisions of the temporal branch of the facial nerve.

Occipitofrontalis

This fibromuscular layer covers the dome of the skull from the eyebrows to the nuchal lines. It consists of two occipital bellies posteriorly attached via a thick fibrous layer, the *galea aponeurotica*, to the two frontal bellies. Only the frontal part is of relevance to the eye and orbit. Its fibres form a thin quadrangular sheet that is attached to the superficial fascia above the eyebrows. The medial fibres are continuous with procerus (Fig. 1.58), the intermediate fibres with corrugator supercilii and orbicularis oculi, and the lateral fibres with orbicularis oculi. The frontal belly is supplied by the temporal branches of the facial nerve. Upon contraction it draws the scalp backwards and elevates the eyebrows, causing transverse wrinkles on the scalp as in expressions of surprise, horror, fright, or when glancing upwards.

Levator palpebrae superioris

This muscle lies within the orbit and is responsible for opening the eyelids and, upon relaxation, allows lid closure due to gravity. The muscle has its origin from the lesser wing of sphenoid, above and in front of the optic foramen, blending with the origin of superior rectus. The muscle belly passes horizontally forward above superior rectus, close to the orbital roof (Fig. 1.45). Behind the orbital margin it curves downwards into the lid where it becomes aponeurotic. The aponeurosis fans out, on either side, to form medial and lateral horns, which extend the whole width of the eyelid. The levator palpebrae superioris inserts into the skin of the upper lid (causing the horizontal palpebral sulcus or furrow) and the anterior surface of the tarsal plate (see Fig. 1.60C). The lateral horn of the aponeurosis forms the *lateral palpebral ligament*, which inserts into the lateral orbital tubercle, and the medial horn forms the *medial palpebral ligament*, which inserts into the frontolacrimal suture. The levator palpebrae superioris is supplied by the superior division of the oculomotor nerve (see p. 67 and Fig. 1.52). Upon contraction it elevates the upper lid, thereby opening the palpebral fissure. On the inferior aspect of the levator palpebrae superioris is a small band of smooth muscle, the *superior tarsal* or *Müller's muscle*. It is attached anteriorly to the upper surface of the tarsal plate and conjunctival fornix (see Fig. 1.60C). It has a sympathetic innervation and upon contraction assists the levator in elevating the eyelid. Although there is no equivalent muscle to the levator in the lower lid, there is a small group of smooth muscle fibres (inferior tarsal muscle) that originate from the fascial sheath of the inferior rectus and insert into the lower tarsus.

The eyelids

The eyelids are thin curtains of skin, muscle, fibrous tissue and mucous membrane that serve to protect the eyes from injury and excessive light and also to distribute tears over the ocular surface during blinking. The upper lid, when open, normally just overlaps the corneoscleral junction and it is this lid that undergoes most displacement during eyelid closure, the lower lid moving only minimally during normal blinking.

On external examination (Fig. 1.59) each lid is seen to be divided into *orbital* and *tarsal portions* by a horizontal *palpebral sulcus*, which is most evident on the upper lid. The upper lid is limited superiorly by the eyebrow, whereas the lower lid blends with the skin of the cheek. The upper and lower lids meet at the *medial* and *lateral canthi* or angles and are separated from one another by an elliptical opening: the *palpebral fissure*. The lateral canthus is an acute angle (60 degrees) and lies close to the eyeball; the medial canthus is rounded, elongated medially, and lies 6 mm from the eyeball. It is separated from the eye by a triangular zone, the *lacus lacrimalis* (lake of tears), in which a small raised red swelling, the *caruncula lacrimalis*, is situated. There are obvious racial differences in the shape and form of the eyelids and canthi, the most conspicuous being the vertical *epicanthal fold* in Oriental and Asian races.

The *eyelid margins* (Figs 1.64 and 1.65) are approximately 30 mm long, 2 mm thick and relatively square in profile along most of their length, except the medial one-sixth, which is rounded and lacks eyelashes. Eyelashes are modified, thick, stiff hairs that occur as double or triple rows close to the anterior lid margin. They curl away from the lashes of the opposite lid. Notable features on the lid margins include:

- *Lacrimal puncta* located at the medial ends of the upper and lower lids. These drain tears from the lacus lacrimalis. The puncta are more easily identified if tension is placed on the lids, causing the papillae to blanch.
- Openings of *tarsal (meibomian) glands* are visible to the naked eye as a row of minute openings on the lid margin posterior to the eyelash follicles. There are around 30 in the upper lid and slightly fewer in the lower lid.
- The *skin/conjunctival transition zone* mucocutaneous junction occurs at the level of the opening of the tarsal glands.
- The *grey line* marks the anterior boundary of the tarsal plate and is a useful landmark for surgical incisions.

The histological structure of the eyelid is summarized in Fig. 1.65A,C. Note that the fibrous framework of the lids is formed by the *orbital septum* arising from the orbital margin (Fig. 1.43) and the *tarsal plates*. The *tarsal plates* are modified regional thickenings of the orbital septum that provide rigidity to the upper and lower lids and separate the orbit

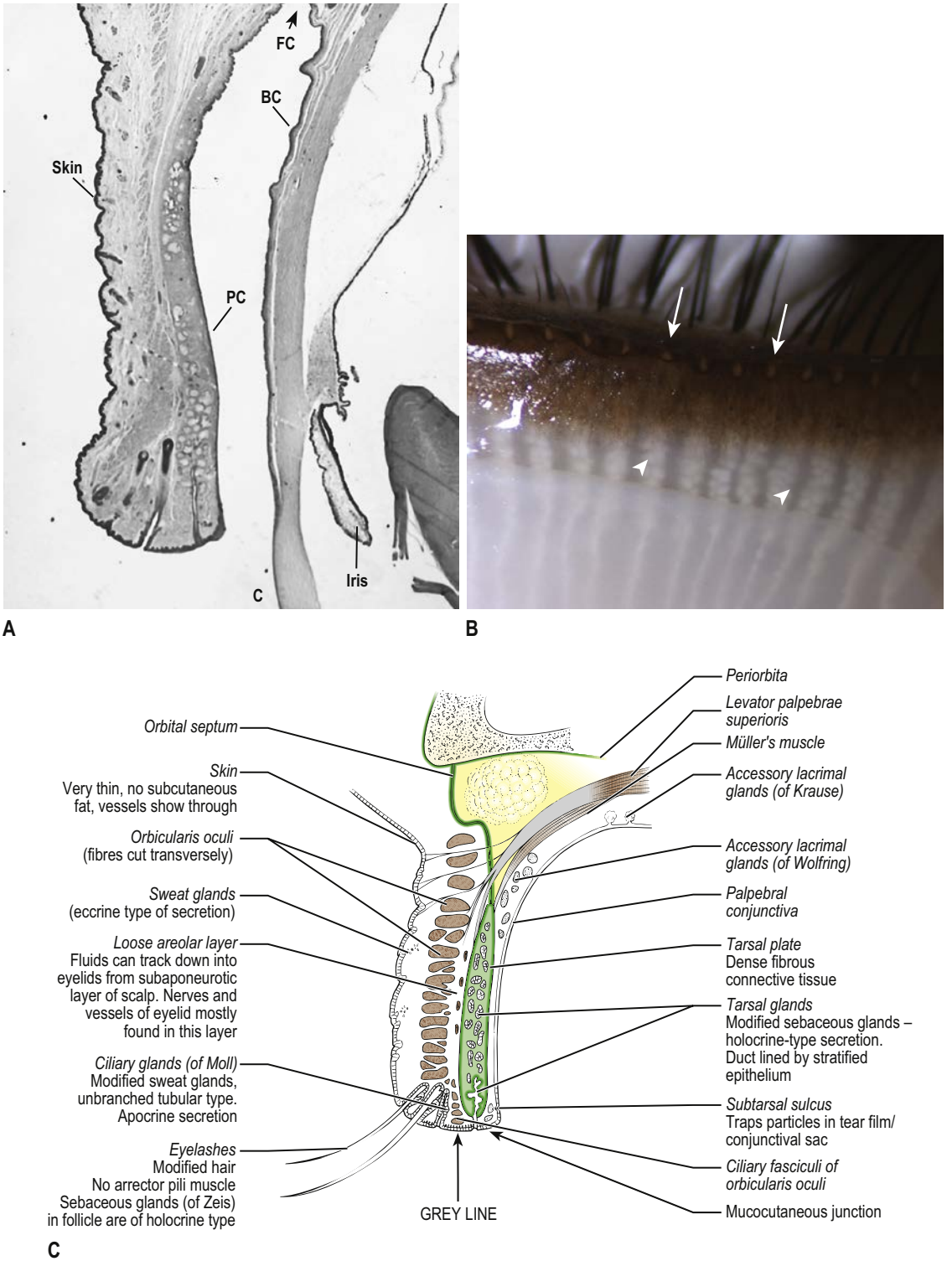


Fig. 1.65 (**A**) Histological preparation of the upper eyelid, conjunctiva and anterior segment. (**B**) Primate upper eyelid as viewed from the inner aspect to reveal the rows of tarsal glands *(arrowheads)* and their openings *(arrows)* on the lid margin. (**C**) Schematic diagram of the upper eyelid in longitudinal section (sagittal plane). *BC*, Bulbar conjunctiva; *C*, cornea; *FC*, forniceal conjunctiva; *PC*, palpebral conjunctiva. Original magnification: **A** × 15. (Fig. 1.65A courtesy W.R. Lee and Springer-Verlag.)

and its contents from the lids. Ageing changes in the septum are discussed in Box 1.34 Clinical correlates.

The tarsal plates consist of dense fibrous connective tissue and are approximately 25–30 mm from the medial to lateral borders. They are 1 mm thick and the upper plate is greater in height (10–12 mm) than the lower plate (5 mm). They are attached at either end via their continuations, the *medial* and *lateral palpebral ligaments*. Skin moves freely over their anterior surface, although the conjunctiva is tightly bound to the posterior surface. External examination of an everted eyelid reveals vertical rows of yellowish tarsal glands (Fig. 1.65B). They are embedded in the matrix of the tarsal plate and consist of modified sebaceous glands. Histologically the acinar cells are replete with lipid droplets that are secreted in a holocrine manner on to the eyelid margin, which retains tears in the conjunctival sac and contributes to the lipid layer of the precorneal tear film.

The blood supply and nerve supply of the lids and surrounding areas are summarized in Fig. 1.66. Lymphatics drain to the superficial parotid or submandibular lymph nodes. The pretarsal portion derives its arterial supply from the superficial temporal and facial arteries (branches of the external carotid), while the post-tarsal portion is supplied by branches of the ophthalmic artery (branch of internal carotid artery) (see Fig. 1.50). Venous drainage follows a similar pattern to the arterial supply (see p. 69, Table 1.4 for consideration of anastomoses between the internal and external carotid arteries). The post-tarsal venous drainage is via the ophthalmic veins to the cavernous sinus.

Clinical conditions relevant to the lids are discussed in Box 1.35 Clinical correlates.

Box 1.34 Ageing changes

Herniation of orbital fat through weakened regions of the septum occurs in the elderly, producing bulging, sagging lids (blepharochalasis).

Movements of the eyelids

The eyelids close as a result of the action of the palpebral fibres of the orbicularis oculi and relaxation of the levator palpebrae superioris. Opening of the lids occurs via the pull of levator palpebrae superioris on the skin, tarsal plate and forniceal conjunctiva. The nerve supply to these muscles is from three sources: orbicularis oculi—the facial nerve (VII); levator palpebrae superioris—the oculomotor nerve (III); while its smooth muscle component, superior tarsal (Müller's) muscle, is supplied by sympathetic nerves. The latter is important in times of fear or excitement when the width of the palpebral fissure is further increased. A common clinical condition causing drooping of the eyelid (ptosis) is a feature of Horner syndrome resulting from a variety of pathological conditions (Box 1.36 Clinical correlates).

The conjunctiva

The conjunctiva (Figs 1.65A,C and 1.67) is a thin translucent mucous membrane that derives its name from the fact that it attaches the eyeball to the lids. It consists of a superficial

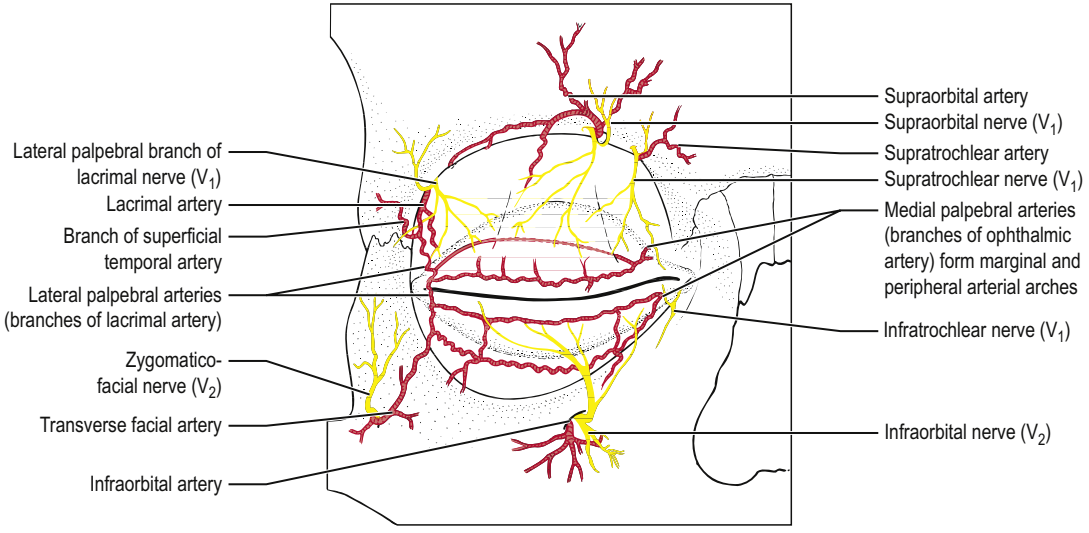


Fig. 1.66 Summary of the blood supply and sensory nerve supply of the eyelids and adjacent areas from cutaneous branches of the ophthalmic (V_1) and maxillary (V_2) divisions of the trigeminal nerve.

AL Grawany

Box 1.35 **Clinical correlate**

Chalazion, hordeolum and ectropion

A *chalazion* is a localized, painless swelling in the lid caused by obstruction and chronic inflammation of a tarsal gland (see Fig. 1.60B).

A hordeolum (sty) may be an acute infection of an eyelash follicle or its sebaceous gland, infection of a ciliary sweat gland (external hordeolum) or acute infection of a tarsal gland (internal hordeolum).

Ectropion is drooping of the lower lid owing to paralysis of orbicularis oculi. Because of paralysis of the fibres of orbicularis that enclose the lacrimal sac, the puncta no longer suck up tears, which may thus pass over the lid margin (*epiphora*).

Box 1.36 **Clinical correlates**

Horner syndrome

Because of the complex neuroanatomy of the sympathetic nervous system, Horner syndrome, which is characterized by classic symptoms of unilateral ptosis, miosis and dry facial skin (anhidrosis) and blushing on the affected side, may result from a wide variety of lesions in the central and peripheral nervous system. These include iatrogenic interruption of the sympathetic chain in the neck, dissection of the internal carotid artery, cervical disk dislocation and the lysis of the first rib affecting the stellate ganglion associated with Pancoast tumour. Other symptoms may include heterochromia and enophthalmos, although the latter is debatable in humans.

conjunctival epithelium overlying a loose connective tissue stroma. The epithelium is continuous with the corneal epithelium at the limbus and with the skin at the mucocutaneous junction on the lid margin. The conjunctiva is reflected from the anterior portion of the sclera at the superior and inferior fornices on to the tarsal surface of the eyelids. Thus, when the lids are closed a potential sac, the *conjunctival sac*, is formed. The volume of this sac is approximately 7 μL, which explains the tendency for eye drops from commercial dispensers (volume 50–70 mL) to overflow unless the lower lid is held away from the globe. The conjunctiva is responsible for the production of the mucous component of the tear film and, in common with other mucous membranes, has a variety of immunological defence mechanisms that protect the ocular surface from infection (see Chapter 7). For descriptive purposes the conjunctiva can be divided into three main regions.

Palpebral conjunctiva (Fig. 1.65A,C)

The palpebral conjunctiva lines the inner surfaces of the eyelids. It is tightly bound to the tarsal plate, the subepithelial connective tissue stroma being thin in this region. The lacrimal puncta open on to the palpebral conjunctiva; thus, the conjunctival epithelium is continuous with the lining of the inferior meatus of the nasal cavity, which explains the manner in which infection spreads between these two sites. A small *subtarsal sulcus*, close to the lid margin, is important in trapping and removing foreign particles and debris on the ocular surface. Accumulations of immune cells in the palpebral conjunctiva might become enlarged in various allergic conditions (Box 1.37 Clinical correlates).

Forniceal conjunctiva (Fig. 1.65A)

The superior and inferior fornices are continuous at the medial and lateral canthi, thus forming a circular *cul de sac*. The ducts of the main lacrimal gland and the bulk of accessory lacrimal glands empty into the superolateral fornix. The forniceal conjunctiva is loosely attached to the fascial sheaths of levator palpebrae superioris and the rectus muscles and thus moves slightly with the eye during contraction of these muscles.

Bulbar conjunctiva

The white sclera is visible through the normal translucent bulbar conjunctiva (see Figs 1.20A and 1.66). It clothes the anterior part of the eyeball including the extraocular muscle insertions and Tenon's capsule. Near the limbus, the conjunctiva is tightly bound to the globe, but further from the limbus there is a loose episcleral tissue layer (Fig. 1.67A) within which lies the pericorneal vascular plexus (Fig. 1.67D). These vessels can become dilated and conspicuous as a result of physical and inflammatory stimuli.

There are two specializations of the conjunctiva in the medial fornix. First, the semilunar fold (*plica semilunaris*), which is probably homologous to the nictitating membrane of lower mammals and many non-mammalian vertebrates. It is highly vascular and rich in goblet cells and interstitial immunocompetent cells. The function of this loose fold may be to facilitate lateral movement of the eye. Second, the *caruncle* (*caruncula lacrimalis*) is a highly vascular nodule of modified skin in the medial corner of the eye containing large nests of accessory lacrimal and sebaceous glandular tissue.

The conjunctiva is a highly visible mucous membrane and as such can manifest a variety of systemic clinical conditions (Box 1.38 Clinical correlates).

Structure of the conjunctiva

Histologically, the conjunctival epithelium varies in structure, depending on location, from a stratified squamous

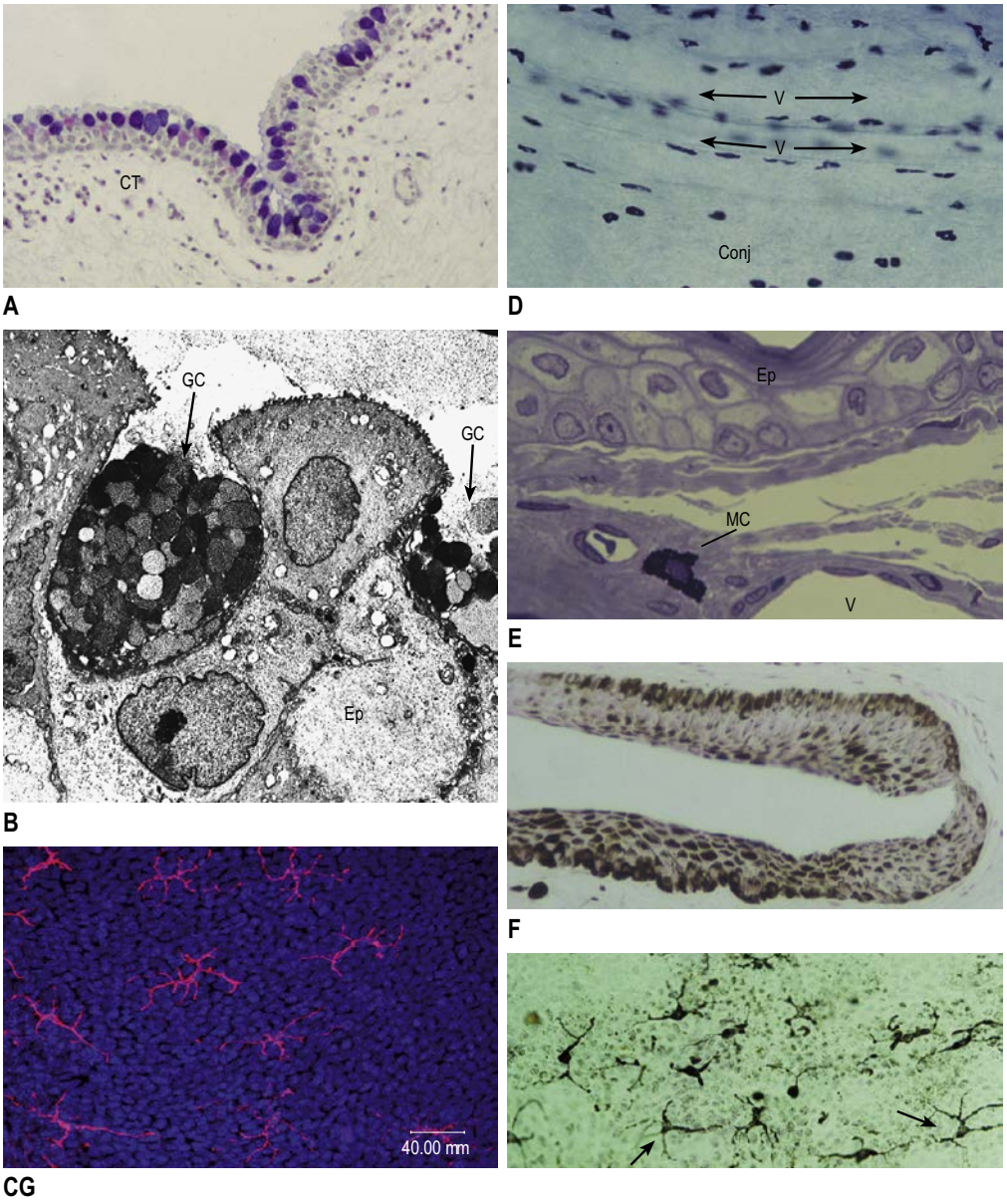


Fig. 1.67 Histology of the conjunctiva. (**A**) Low-power light micrograph of human conjunctiva (periodic acid–Schiff [PAS] stain) showing the irregular nature and goblet cell content *(purple and red PAS+ profiles)* of the epithelium (compare with corneal epithelium; see Figs 1.12, 1.14 and 1.16). Note the accumulations of lymphoid cells in the highly vascular connective tissue stroma, a common feature in eyes of elderly patients. (**B**) Electron micrograph of goblet cells *(GC)* in the conjunctival epithelium

.

(**C**) Plan view of + intraepithelial dendritic cells (sometimes called Langerhans cells) in the limbal/conjunctival epithelium in wholemount preparation from Cx3cr1-GFP transgenic mouse (*green*: Cx_3cr1_+ myeloid-derived cells; *red*: MHC class II; *blue*: DAPI-stained nuclei). Note the Langerhans cells are Cx_3cr1_+ MHC class II_+. The Z-profile shows how some of the cell processes act as periscopes and project towards the superficial aspect of the conjunctival/limbal epithelium. (**D**) Toluidine blue-stained conjunctival wholemount illustrating the orientation and distribution of mast cells around limbal vessels *(V)* and in the bulbar conjunctiva *(Conj)* where they are more rounded. (**E**) Toluidine blue-stained semi-thin resin section of the limbal region showing a mast cell

(MC) adjacent to a large venule *(V)*. (**F**) Primate conjunctiva (histological preparation, H&E) showing melanocytes in the basal layer of the conjunctival/limbal epithelium. Note the intraepithelial melanin granules throughout the conjunctival epithelial layers. (**G**) Melanocytes as seen in a limbal wholemount preparation. Note their highly dendriform shape *(arrows)* and how they form a halo of melanin granules within the adjacent epithelial cells (epithelial-melanin unit). *CT*, Subepithelial connective tissue or connective tissue stroma; *Ep*, epithelium. Original magnifications: **A** × 150; **B** × 3000; **C** × 200; **D** × 100; **E** × 650; **F** × 150; **G** × 160. (Fig. 1.67

Box 1.37 **Clinical correlates**

Allergic responses in the conjunctiva

Clinical examination of the everted lid for signs of ocular allergy and infection is a common procedure. Two major types of abnormal accumulations of immune cells may occur: *follicles*, which are similar to mucosal-associated lymphoid follicles elsewhere and consist primarily of lymphocytes; and *papillae*, which are focal aggregates of chronic inflammatory cellular infiltrates and accompanying vascular changes. These are usually associated with allergic conditions and irritation of the ocular surface, such as in contact lens wear.

Box 1.38 **Clinical correlates**

Systemic disease evident in the conjunctiva

The conjunctiva may manifest signs of several important systemic diseases: pathognomonic signs are present in sickle cell anaemia (comma sign), jaundice (scleral icterus [yellowing]) and vitamin A deficiency (Bitot's spots).

non-keratinizing epithelium (close to the lid margin) to a stratified columnar epithelium (bulbar). In general, it consists of between two and seven layers of epithelial cells that are organized into three main types: basal, intermediate and superficial. There is no 'prickle' layer as found in corneal epithelium, indicating that there are fewer desmosomes between the conjunctival epithelial cells. There are numerous other cell types resident within the epithelium, reflecting its protective function, including the following:

- *Goblet cells* (Fig. 1.67A,B): unicellular mucus-secreting cells that vary in density in different regions of the conjunctiva, being most numerous in the fornices and plica semilunaris. They are responsible for the secretion of the majority of conjunctival mucins.
- *Melanocytes* (Fig. 1.67F,G): degree of melanization varies dependent on race, although melanocytes are present in all eyes. Melanosomes are synthesized within the melanocyte before exocytosis and subsequent uptake by surrounding epithelial cells as occurs in the epidermis.
- *Intraepithelial MHC class II-positive dendritic cells* (sometimes referred to as Langerhans cells because of their similar morphology to analogous dendritic cell populations in the epidermis of the skin) (Fig. 1.67C) function as 'sentinels' on the ocular surface and are responsible for trapping and internalizing antigens and transporting these signals to either local lymph nodes (such as the preauricular nodes) or conjunctival associated lymphoid tissue (CALT) or follicles, where they are capable of presenting antigens to naive T cells and inducing primary immune responses or driving antigen-specific B-cell maturation and immunoglobulin production (see Chapter 7, p. 425-429).
- *Intraepithelial lymphocytes*: a feature of normal conjunctiva, but increased numbers occur in inflammatory conditions and close to subepithelial lymphoid accumulations. These are predominantly $CD3^+$ T cells, of which half are $CD3^+CD4^+$ T cells, half are CD8+ T cells with occasional $CD19^+$ B cells being present. There is significant variation in ratios in different regions of the conjunctiva and recent data suggest that IL-22 producing γδ T cells and IL-22 producing γδ - innate lymphoid cells are also present in the human conjunctiva (see also Chapter 7 p. 434)

The epithelium has an irregular basal aspect adjacent to the underlying connective tissue that is sometimes described as having a looser lymphoid layer and a deeper fibrous layer. Distinct papillae, finger-like protrusions of connective tissue stroma that project into the epithelium, are found only near the limbus. The subepithelial connective tissue contains numerous immunocompetent cells such as mast cells (Fig. 1.67D,E), eosinophils, plasma cells and lymphocytes scattered among the matrix. In some eyes, particularly of older individuals, these may form diffuse or discrete aggregates. Some of these follicles contain pale germinal centres and represent the local mucosal lymphoid tissue (MALT) or CALT. There is no clear evidence of the specialized antigen-transporting intraepithelial M cells that are typically found in other similar MALT such as Peyer's patches or tonsils. The diffuse subepithelial aggregates along with intraepithelial lymphocytes form the efferent arm of the immune system and aid in immunological protection of the ocular surface. The topographical distribution of the small (~0.3 mm) lenticular lymphoid follicles and epithelial crypts in the tarso-orbital conjunctiva suggests that they may approximate to the cornea during eye closure and thus act as an 'immunological cushion'.

The loose connective tissue stroma contains a rich vascular network, similar to the eyelids. In addition, it also receives blood from the anterior ciliary arteries (see Fig. 1.50).

The sensory nerve supply of the palpebral conjunctiva is almost entirely from branches of the ophthalmic division of the trigeminal (supraorbital, supratrochlear and lacrimal nerves). These contain the neurotransmitters substance P, calcitonin gene-related peptide (CGRP) and gallanin. Only the medial portion of the inferior forniceal and palpebral conjunctiva derives its nerve supply from the maxillary division (infraorbital nerve) (see Fig. 1.66). The long ciliary nerves supply the bulbar conjunctiva. Parasympathetic nerves from the pterygopalatine ganglion (containing the neurotransmitters acetylcholine and VIP) and sympathetics (containing norepinephrine and neuropetide Y) travelling

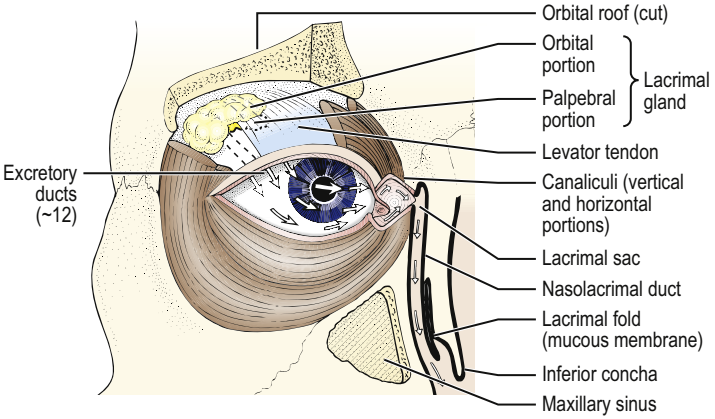


Fig. 1.68 Schematic diagram summarizing the entire lacrimal apparatus. *Arrows* indicate the direction of tears from the site of production to the site of drainage.

with branches of the ophthalmic artery are also present in the conjunctiva. Both parasympathetics and sympathetics have been identified around goblet cells, whereas the sensory nerve endings occur only among the stratified squamous epithelial cells.

Glands in the conjunctiva

Besides the unicellular mucous glands (goblet cells) distributed throughout the conjunctiva, there are several small collections of named glands, some of which are accessory lacrimal glands (glands of Krause in the upper fornices, glands of Wolfring in the upper border of the tarsus); others secrete mucus (glands of Henle). The accessory lacrimal glands are under sympathetic stimulation and are responsible for baseline tear production.

Lacrimal apparatus (Fig. 1.68)

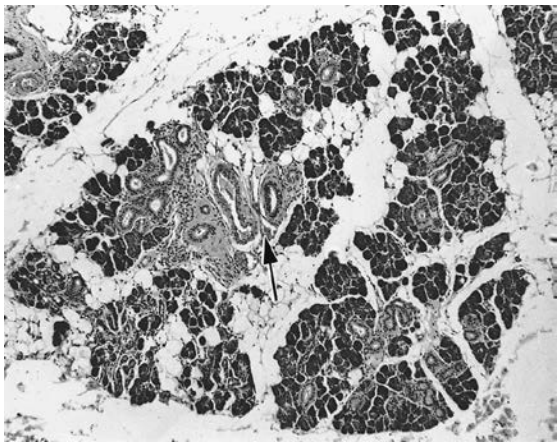
The lacrimal apparatus consists of the *lacrimal gland, lacrimal puncta, lacrimal canaliculi, lacrimal sac* and *nasolacrimal duct.* The lacrimal apparatus produces tears that moisten the ocular surface, thus preventing desiccation of delicate ocular cells and tissues and facilitating non-friction-bearing movements of the lids on the globe. Tears are thus essential in maintaining the functional integrity of the eye.

Tear film. The tear film (7–9 μm) is composed of three layers: an outer oily or lipid layer (from meibomian and Zeis glands), a middle aqueous layer containing protein, electrolytes and water (mainly from lacrimal glands but also small contributions from conjunctival epithelia and cornea) and a deep hydrophilic mucin layer (from goblet cells and conjunctival epithelial cells and some from the corneal epithelium) associated with the microplicae-rich surface of the conjunctival epithelium (Fig. 1.67B).

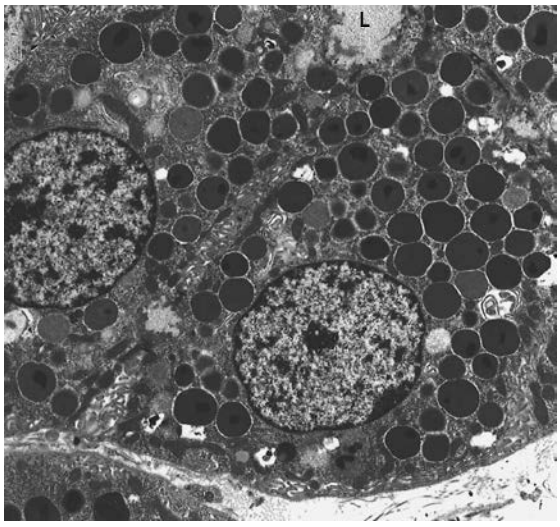
Lacrimal gland

The lacrimal gland measures approximately 20 × 12 × 5 mm and weighs approximately 78 mg. It is divided by the lateral horn of the aponeurosis of the levator palpebrae superioris into a large *orbital* and small *palpebral portion,* which are continuous via a small isthmus around the lateral border of the aponeurosis. The orbital portion is shaped like an almond with a convex outer surface that is lodged in the *lacrimal fossa* (see Fig. 1.47). The concave inferior surface is moulded around the tendons of levator palpebrae superioris and lateral rectus (Fig. 1.70). The palpebral portion of the gland is approximately one-quarter of the total gland and its inferior surface lies close to the eye; indeed, the gland can usually be seen when the upper lid is everted. In the orbital portion, fine interlobular ducts unite to form three to five main excretory ducts that then traverse the palpebral portion, joining a further five to seven from this part of the gland before entering the superotemporal conjunctival fornix. As a result of this arrangement, removal of the palpebral portion renders the entire gland non-functional.

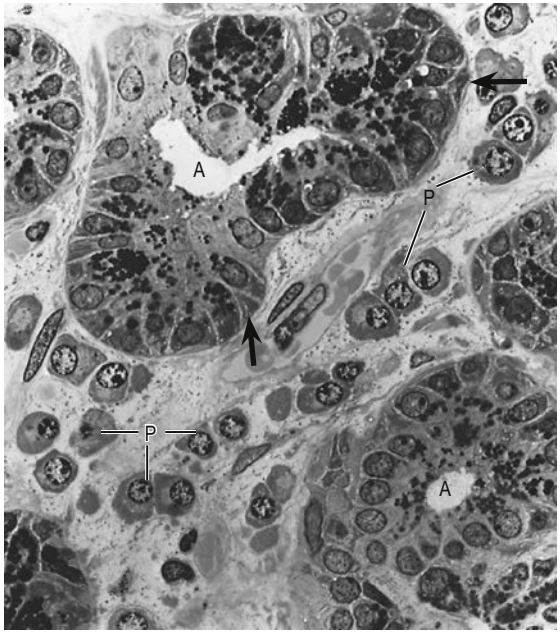
Histological structure (*Fig. 1.69A–C*). The lacrimal gland is a branched tubuloacinar gland of the serous type. It is composed of many lobules separated by interstitial fibrovascular septae that are continuous with the poorly developed capsule. On section, each lobule contains numerous acini separated by abundant loose intralobular connective and adipose tissue. Histologically, the acini resemble those of the parotid gland and appear as a series of rounded profiles in cross-section (Fig. 1.69A,B). Each acinus or tubuloacinar unit consists of a single layer of cuboidal or columnar cells whose apices are directed towards a central lumen (Fig. 1.69B–D). A layer of stellate-shaped myoepithelial cells surrounds each acinus. The central lumen of several acini unite to form intralobular ducts (Fig. 1.69A), which eventually form larger interlobular ducts that unite to form the main



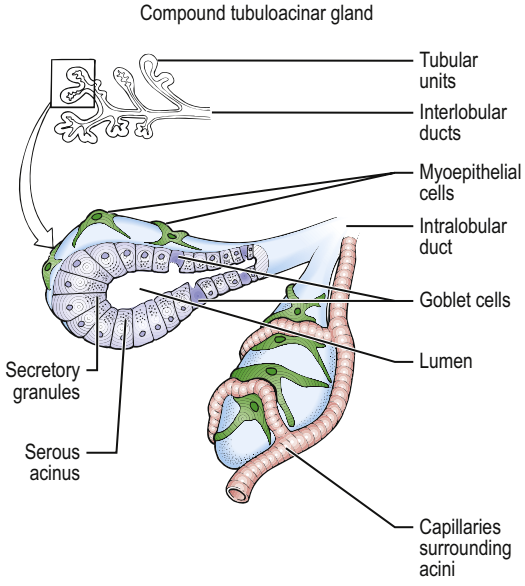
A



C



B



D

Fig. 1.69 Histology of the lacrimal gland. (**A**) Low-power micrograph of an entire lobule containing a series of large intralobular ducts *(arrow)*. (**B**) Semi-thin section illustrating the arrangement of the glandular epithelial cells, containing numerous secretory granules, in acinar units *(A)*. The vascular intralobular connective tissue is extremely rich in mature plasma cells *(P)*. *Arrows*, myoepithelial cells. (**C**) Electron micrograph revealing the ultrastructure of a few pyramidal-shaped acinar cells whose apices are directed toward the central lumen *(L)*. Note the numerous electron-dense zymogenic granules in the apical portion of the cells. An intracellular canaliculus is indicated *(arrow)*: *N*, nucleus. (**D**) Three-dimensional diagram summarizing the arrangement of the epithelial cells, myoepithelial cells and capillaries in the lacrimal gland. Original magnifications: **A** × 40; **B** × 630; **C** × 4400. (Fig. 1.69B courtesy W.R. Lee.)

excretory duct system. The glandular epithelial cells have the characteristic histological and ultrastructural appearance of serous cells, namely basophilic cytoplasm, owing to large numbers of round or oval secretory (zymogen) granules. The epithelial cells have been subdivided into various subtypes depending on the size and electron density of these granules; however, there is still debate as to whether these are functional subtypes or different stages in the life cycle of one cell type. The presence of true intracellular canaliculi (Fig. 1.69C), as observed in salivary glands, is also

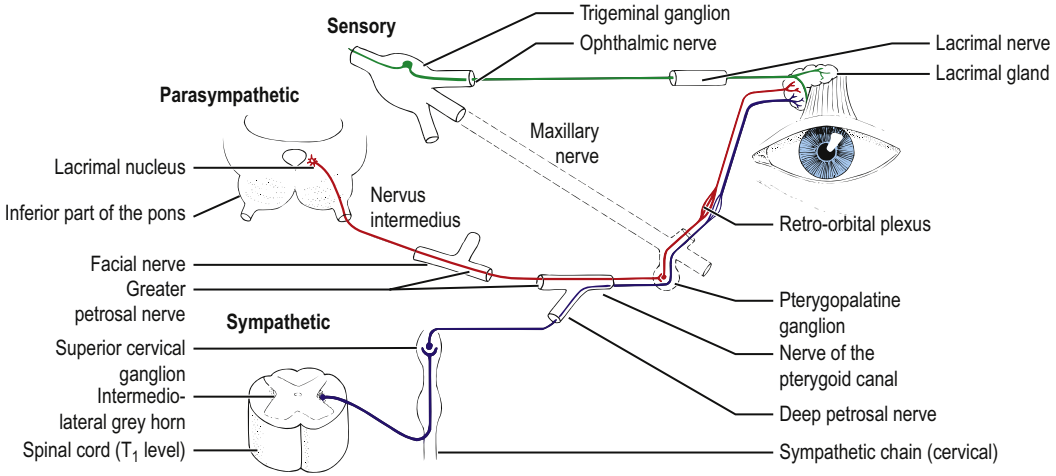


Fig. 1.70 Diagram summarizing the sensory (trigeminal nerve: *green*), secretomotor (facial nerve: *red*) and sympathetic (*blue*) innervation of the lacrimal gland.

still controversial. The secretion of the gland is primarily proteinaceous, although some granules contain glycosaminoglycans and the lumen on histological examination contains strongly eosinophilic and mucoid-like secretory material. The secretion also contains lysozymes, lactoferrin, B-lysin and immunoglobulin A (IgA) (see Chapter 7 p. 455), which are important in defence of the ocular surface against microbial infection. The IgA is derived from numerous plasma cells, present with other immunocompetent cells such as lymphocytes and mast cells, in the intralobular connective tissue (Fig. 1.69B). The number of these cells increases with age, concomitant with increased fibrosis and fatty infiltration and a decrease in the acinar elements especially in the orbital lobe.

Nerve supply. The nerve supply of the lacrimal gland and the complex course of the secretomotor fibres are summarized in Fig. 1.70. The lacrimatory nucleus of the facial nerve lies at the rostral end of the general visceral efferent column of cell bodies in the brainstem, which include the superior and inferior salivatory nuclei (see p. 81). The cells are under the influence of the hypothalamus via descending autonomic pathways, thus explaining the neuronal pathways involved in excess lacrimation, which accompanies various emotional states. Reflex excess lacrimation occurs following irritation of the cornea, conjunctiva and nasal epithelia (afferent pathways in ophthalmic and maxillary divisions of the trigeminal). Interneurones connect the trigeminal sensory nuclei with the lacrimatory nucleus.

Blood supply. The lacrimal gland derives its blood supply principally from the lacrimal artery, an early branch of the ophthalmic artery, although a variable branch from the infraorbital artery (originating indirectly from the external carotid) might also aid in its supply. Venous blood drains posteriorly, usually to the superior ophthalmic vein in the orbit, and lymph drains to the preauricular node.

Collecting portion of the lacrimal apparatus (Fig. 1.69)

The collecting system serves to drain normal tears that have not evaporated (normally only a very small quantity) and those produced in times of increased lacrimation. Excess tears are drained from the medial aspect of the conjunctival sac via the canaliculi into the lacrimal sac and nasolacrimal duct to empty into the inferior meatus of the nasal cavity (Fig. 1.68).

The *puncta* are small openings (visible to the naked eye on the medial margin of each lid) at the summit of small swellings, the *papillae lacrimalis* (see Fig. 1.64). Tears that enter the puncta from the *lacus lacrimalis* during blinking pass into the *lacrimal canaliculi* situated in the upper and lower lid behind the medial palpebral ligament. Each canaliculus is about 10 mm long (0.5 mm in diameter) and has vertical and horizontal components (Fig. 1.68) that might unite to form a common canaliculus before entering the *lacrimal sac*. The canaliculi are lined by stratified squamous non-keratinizing cells. They enter the lacrimal sac by piercing the fascial covering. This sac is 12 mm long and in its upper portion the walls are usually in apposition. It lies in the lacrimal fossa (Fig. 1.5B), protected by the medial palpebral ligament anteriorly and the lacrimal fibres of orbicularis oculi posteriorly. It is related medially to the ethmoidal air cells and the middle meatus of the nasal cavity. The walls of the lacrimal sac consist of

fibroelastic tissue and are lined by a mucous membrane consisting of stratified cuboidal/columnar epithelium containing goblet cells. This epithelium is continuous with that lining the canaliculi and the nasolacrimal duct inferiorly.

The nasolacrimal duct empties into the anterior part of the inferior meatus of the nasal cavity, the opening being protected by a flap of mucous membrane, which prevents air and debris passing up the duct during nose blowing. The duct lies in a bony nasolacrimal canal formed by the maxilla, the lacrimal bone and the inferior nasal concha. The duct is lined by a stratified columnar ciliated epithelium, which rests upon a vascular substantia propria. Knowledge of the position of a variety of constrictions and mucous membrane folds or valves along the course of the canal is important during reconstruction of congenitally malformed lacrimal drainage systems.

Anatomy of the visual pathway

The visual pathway is made up of the *retina, optic nerves, optic chiasma, optic tracts, lateral geniculate bodies, optic radiations* and *visual cortex* (summarized in Fig. 1.71). There are other areas of the cortex also associated with vision such as the frontal eye fields (see Chapter 5 for a full description of visual physiology). The visual pathway is effectively a tract within the central nervous system because the retinae develop as evaginations of the diencephalon (see Chapter 5) and, as discussed previously, the optic nerves are covered by layers of meninges; the corneoscleral envelope and uveal tract of the eye itself are homologous to the dura mater and pia–arachnoid, respectively.

The retina has been described on p. 37. The intraocular, orbital and intracanalicular portions of the optic nerve were described on p. 56. Description of the visual pathway is in the later section 'Intracranial portion of the optic nerve'.

Intracranial portion of the optic nerve (Figs 1.51 and 1.71)

The optic nerves leave the cranial end of the optic canal and pass medially, backwards and slightly upwards within the subarachnoid space of the middle cranial fossa. They end by forming the optic chiasma in the floor of the third ventricle. Important relations include the olfactory tracts, frontal lobe (gyrus rectus) and the anterior cerebral arteries above. Each internal carotid artery as it emerges from the roof of the cavernous sinus lies lateral to the junction of the optic nerve and chiasma (Fig. 1.45). Below the optic nerves lies the jugum of the sphenoid and the sulcus chiasmaticus or optic groove.

Optic chiasma

The optic chiasma (Figs 1.71 and 1.72) is situated at the junction of the anterior wall and the floor of the third ventricle, approximately 5–10 mm above the diaphragma sella and the hypophysis cerebri. It is a flattened quadrangular bundle of nerves measuring 12 × 8 mm whose anterolateral angles are continuous with the optic nerves and its posterolateral angles form the optic tracts. It usually lies just behind the optic groove or sulcus chiasmaticus but rarely might lie partly within the sulcus. The tuber cinereum (a sheet of grey matter that forms a median eminence around the base of the pituitary stalk or infundibulum) lies behind and below the chiasma between the mamillary bodies, important as a memory relay to the hippocampus. The anterior perforated substance is an important lateral relation. The anterior communicating artery passes between the two anterior cerebral arteries and lies above the chiasma. The partial crossing of optic nerve fibres in the optic chiasma is an essential requirement for binocular vision. The fibres from the nasal hemiretina of each eye cross the midline to enter the contralateral optic tract after taking a short loop in the ipsilateral tract or into the contralateral optic nerve. Nerve fibres from the temporal hemiretina do not cross at the chiasma (Fig. 1.71).

Optic tracts

The optic tracts (Figs 1.71 and 1.72) wind round the cerebral peduncles of the rostral midbrain and each divides into a large *lateral root*, which terminates posteriorly in the lateral geniculate body and is concerned with conscious visual sensation, and a smaller *medial root*. The medial root is connected to the *pretectal area* and the *superior colliculus* by the superior brachium and carries around 10% of tract fibres, which functionally are not concerned with conscious vision. They contain six groups of fibres, three of which target the superior colliculus (involved in the visual grasp reflex, automatic scanning of images and visual association pathways); the remaining three enter either the pretectal nucleus (serve the pupillary light reflex), the parvocellular reticular formation (arousal function), or the retinohypothalamic tract, which terminates in the suprachiasmatic nucleus of the hypothalamus (possibly involved in photoperiod regulation and has been invoked to account for the beneficial effect of bright artificial light or sunshine on mood).

The *superior colliculi* are two small rounded elevations located on the dorsal surface of the midbrain above the inferior colliculi, visible on external examination of the brainstem. The pineal body lies between and above the superior colliculi. The two pairs of colliculi are referred to collectively as the *tectum*. The mesencephalic or tectal termination of optic tract fibres is phylogenetically older than the forebrain termination (visual cortex).

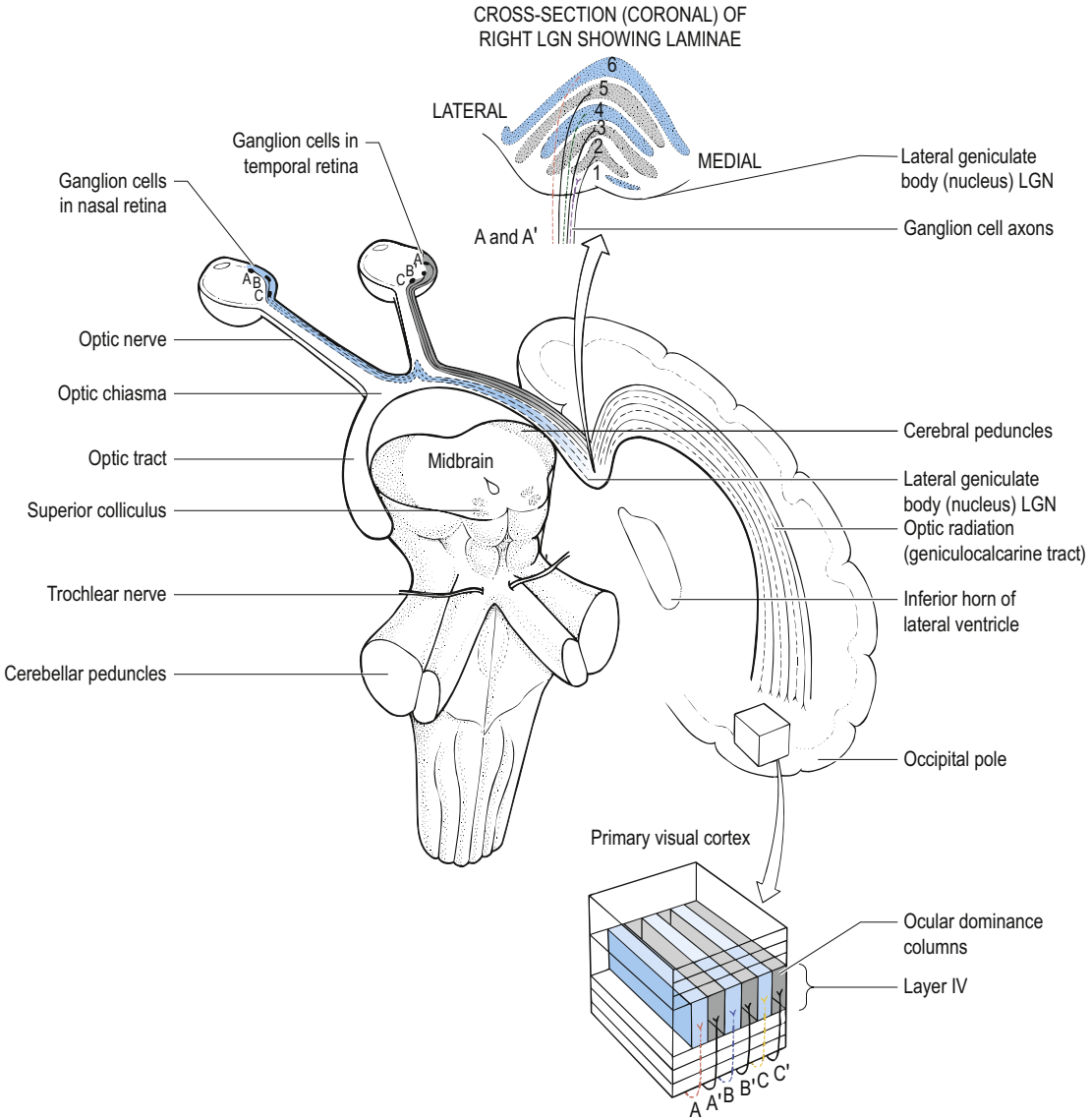


Fig. 1.71 Diagram summarizing the visual pathways. The manner of retinotopic projection to the lateral geniculate nucleus *(LGN)* and the ocular dominance columns in the primary visual cortex from left and right eyes are illustrated by three imaginary points or images (A, B, C) from the left visual field (not shown) falling on the right half of each retina (A, B, C in left eye and A′, B′, C′ in right eye).

The *lateral root of the optic tract* passes backwards, a little upwards, and terminates in the lateral geniculate nucleus (LGN) part of the thalamus (a relay station for ascending sensory information). The lateral root does not lie completely free because its medial aspect is attached to the outer wall of the third ventricle by a narrow band of tissue. It rotates slightly on its own axis (90 degrees inward twist) as it passes round the cerebral peduncles. It runs above the dorsum sella and crosses the third nerve from medial to lateral. The optic tract runs below and parallel to the posterior cerebral artery (Figs 1.52 and 1.71). The middle portion of the tract is overlapped by the uncus and parahippocampal gyrus.

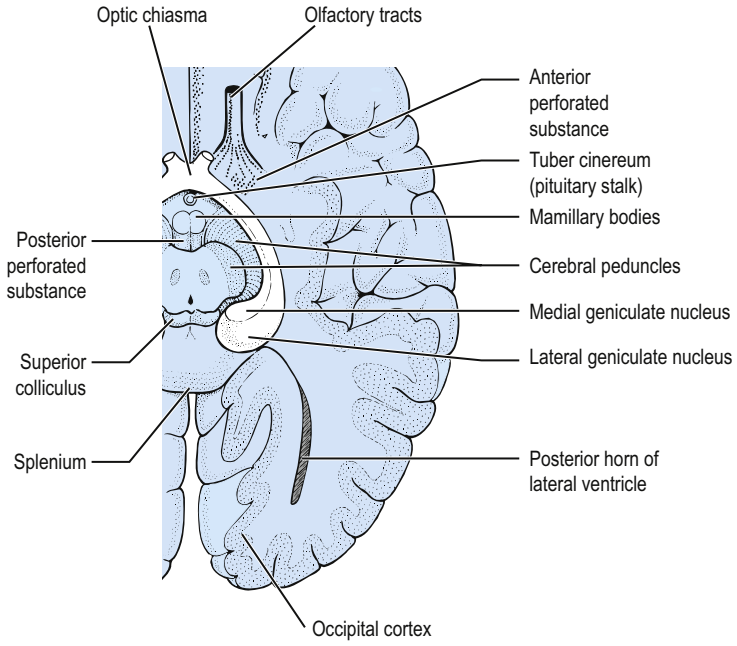


Fig. 1.72 The base of the brain with the brainstem and cerebellum removed to expose the optic nerves, optic chiasma, optic tracts and lateral geniculate nucleus and their relations.

Lateral geniculate bodies

Each lateral geniculate body (Figs 1.71 and 1.72) is distinguishable on the surface of the brain as an ovoid projection on the posteroinferior aspect of the thalamus, partly obscured by the overhanging temporal lobe (Fig. 1.73). It consists of a body, head, spur and hilum. The hilum is continuous with the groove between the medial and lateral root of the optic tract, which enters its anterior aspect. It lies at the anterior aspect of the pulvinar, which also partly surrounds it, particularly from above. The LGN in which the great majority of the optic tract fibres terminate consists of six laminae or cell layers (numbered 1 to 6 beginning at the hilum), oriented in a dome-shaped mound similar to a stack of hats (Fig. 1.71). On coronal section, the layers of cell nuclei (approximately 1 million) are separated by white matter (optic tract fibres). Nerve fibres derived from the contralateral eye (crossed fibres from the nasal half of the retina) terminate on cell bodies in layers 1, 4 and 6. Those of the ipsilateral eye (uncrossed) terminate in layers 2, 3 and 5. Thus each LGN receives information from both retinae. Each retinal ganglion cell axon may terminate on up to six geniculate cells; however, these are located in one lamina. Fibres from the upper quadrants of peripheral retinae synapse on the medial aspect of the LGN and those of the lower quadrant on the lateral aspect. The *macula* projects to a disproportionately large central wedge of the LGN. The posterior aspect of the LGN is dome-shaped and it is from here that the geniculate cell axons that form the optic radiation emerge. The bulk of the LGN sends its fibres via the optic radiation to the visual cortex (area 17). The LGN has input from areas 17, 18, 19, oculomotor centres and the reticular formation.

Optic radiations (geniculocalcarine tracts)

These tracts (Fig. 1.71) consist of nerve fibre bundles whose cell bodies lie in the LGN. Their axons terminate in the visual (striate) cortex. The fibres form a wide forward and inferiorly directed fan-shaped loop (of Meyer), first into the retrolenticular portion of the internal capsule (posterior to sensory fibres and medial to auditory fibres). The fibres then pass into the temporal lobe around the inferior horn of the lateral ventricle (Fig. 1.73). Each tract then passes posteriorly along the lateral aspect of the posterior horn of the lateral ventricle before turning medially to enter the visual cortex. The optic radiations are of major clinical importance as they are frequently involved in cerebrovascular disturbance or tumours (Fig. 1.80). Not all fibres loop to

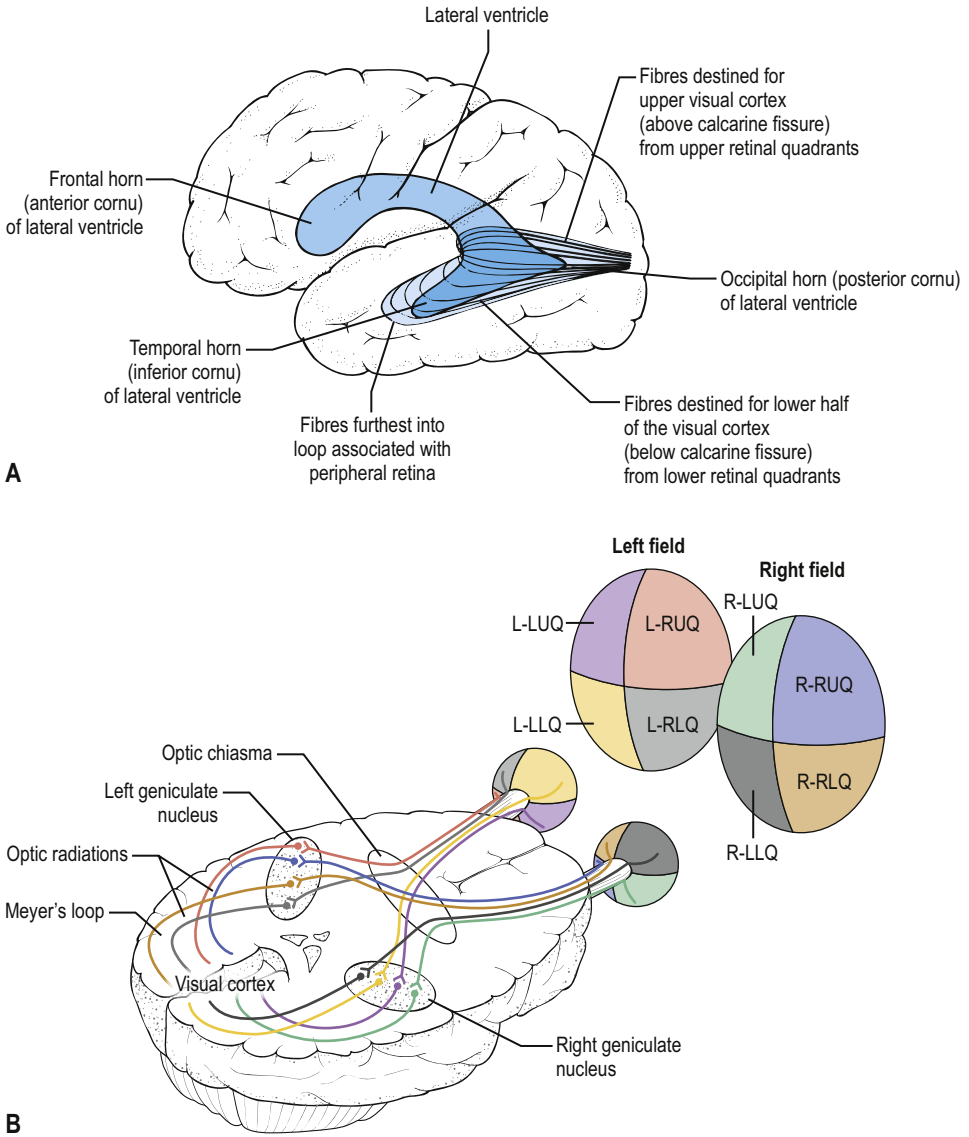


Fig. 1.73 (**A**) The optic radiation and its relation to the lateral ventricle, viewed from the left side. (**B**) Diagram showing the retinotopic organization of fibres within the optic radiations. R-RUQ, right eye–right upper quadrant; R-RLQ, right eye–right lower quadrant; L-LUQ, left eye–left upper quadrant; L-LLQ, left eye–left lower quadrant. (Fig. 1.73B courtesy Wikimedia Commons.)

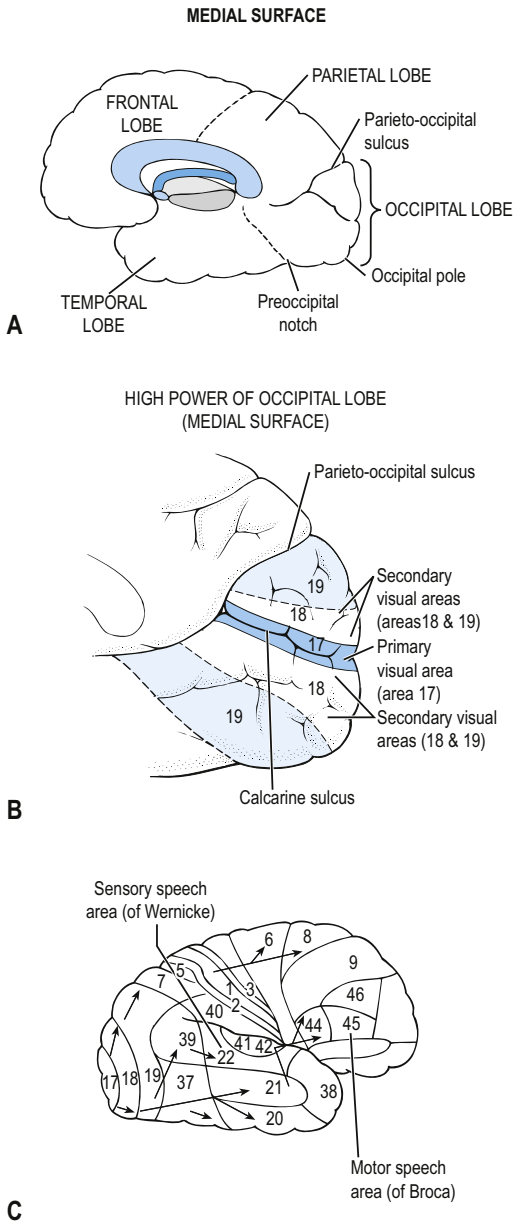


Fig. 1.74 (**A**) Simplified diagram showing the boundaries of the lobes of the brain as viewed from the medial aspect. (**B**) The medial surface of the right occipital lobe indicating the sites of the primary and secondary visual areas. (**C**) The cytoarchitectural areas of the cortex as described by Brodmann. Higher visual projections are from area 17 of the left hemisphere. Those to and from area 20/21 are concerned with detail and colour. Projections to and from area 7 are associated with stereopsis and movement, while those to and from area 39 are concerned with recognition of letters and numbers.

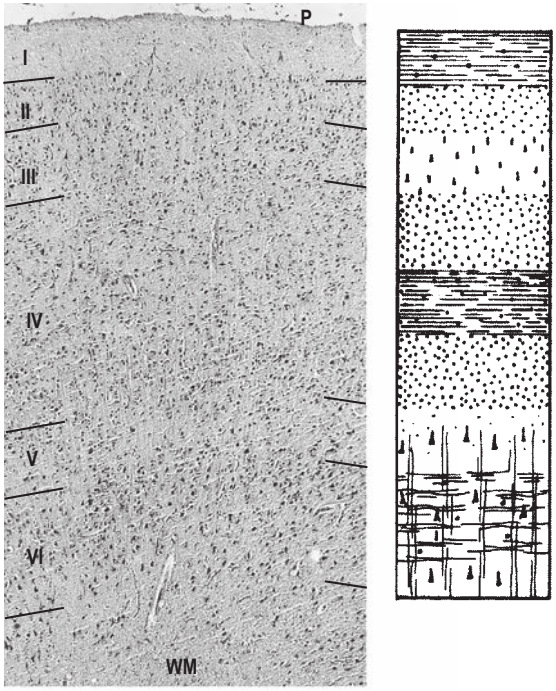


Fig. 1.75 Histology and cytoarchitecture of the primary visual cortex. Layers I–VI are indicated on the micrograph and adjacent diagram. *P*, Pia; *WM*, white matter. Original magnification: × 35.

the same degree (Fig. 1.73A,B). Those destined for the lower half of the visual cortex take a wider sweep into the loop around the tip of the inferior horn of the lateral ventricle than those designed for the upper half of the visual cortex (Fig. 1.73B). The fibres that swing furthest into the loop are associated with peripheral retina; those that pass more directly posteriorly originate closer to or within the macula region of the retina.

Primary visual cortex (area 17)

The myelinated fibres of the geniculocalcarine tract (containing fibres from both eyes) enter the primary visual cortex, which lies within the depths of the calcarine sulcus and extends both above and below its margins on the medial surface of the occipital cortex, extending posteriorly as far as the occipital pole (Fig. 1.74B) and anteriorly as far as the parieto-occipital sulcus. The area above the fissure is known as the *cuneus gyrus* and below is the *lingual gyrus*.

Fibres from the superior retinal quadrants (representing the inferior visual field) pass to the upper lip of

the calcarine sulcus. The fibres representing the macula account for one-third of the visual cortex (posterior portion of area 17). The myelinated fibres of the geniculocalcarine tract entering this area of cortex create the conspicuous white line or stria (of Gennari). This represents layer IV in the cortex (Fig. 1.75). The six basic layers of the primary visual cortex are shown in Fig. 1.75. This region of cortex, although thinner (1.5 mm), is more cellular than other areas of cortex, the predominant cell type not being pyramidal but small stellate cells. Alternating *ocular dominance columns* of these cells receive input from right and left eyes (Fig. 1.71). The geniculocalcarine projection is ordered in a manner whereby matching points from the retinae of both eyes are registered side by side in contiguous columns (see Chapter 5 for a discussion of binocular vision, colour vision, etc.). The cells in laminae II and III project to the secondary visual cortex (areas 18 and 19; Fig. 1.74B,C). Those in lamina V project to the superior colliculus and those of VI are a major source of 'feed-forward' to the LGN.

Secondary visual association areas (areas 18 and 19)

These association areas (Fig. 1.74B,C), which lack the characteristic extra stria found in the primary visual cortex, lie above and below area 17 and extend on to the lateral surfaces of the cerebral hemispheres. They possess the usual six layers, although layer IV is less extensive. Areas 18 and 19 receive afferent input fibres from area 17, the thalamus and pulvinar, together with other regions of the cerebral cortex. The connections of areas 18 and 19 mainly follow dorsal and ventral pathways (Fig. 1.75C). Outputs to area 7 in the parietal cortex are mainly involved in stereopsis and movement. Ventral outputs to the inferotemporal cortex are concerned with analysis of colour and form and connections to area 37 are associated with recognition of faces. Area 18 is also likely to be involved in sensory–motor eye coordination, because this is known to be linked to the frontal eye fields and oculomotor nuclei via descending pathways. This area also integrates information from two halves of the visual field via commissural fibres crossing the midline in the splenium of the corpus collosum (Fig. 1.72).

Frontal eye field

This frontal area (Fig. 1.74C) corresponds to Brodmann's areas 6, 8 and 9 and is concerned with voluntary control of eye movements (saccades). Fibres pass from here to the superior colliculus, and in turn are connected to the extraocular cranial nerve nuclei (III, IV and VI) and anterior horn cells (motor neurones) in cervical spinal cord segments, thus allowing coordination of head and neck movements with eye movements.

Retinotopic organization of the visual pathway and visual pathway disturbances (Figs 1.76–1.82)

A large amount of neurobiological research in primates and non-primates, together with observations of visual dysfunction or abnormalities in human subjects by neuro-ophthalmologists, has led to a considerable body of knowledge regarding the position along the visual pathway of fibres originating from various points on the retina. This information has been crucial to our understanding of the physiology of vision (see Chapter 5) and helps to explain the specific patterns of visual field disturbances following localized lesions in the pathway. Examples of these lesions and the resultant visual field loss are provided in Figs 1.77–1.82.

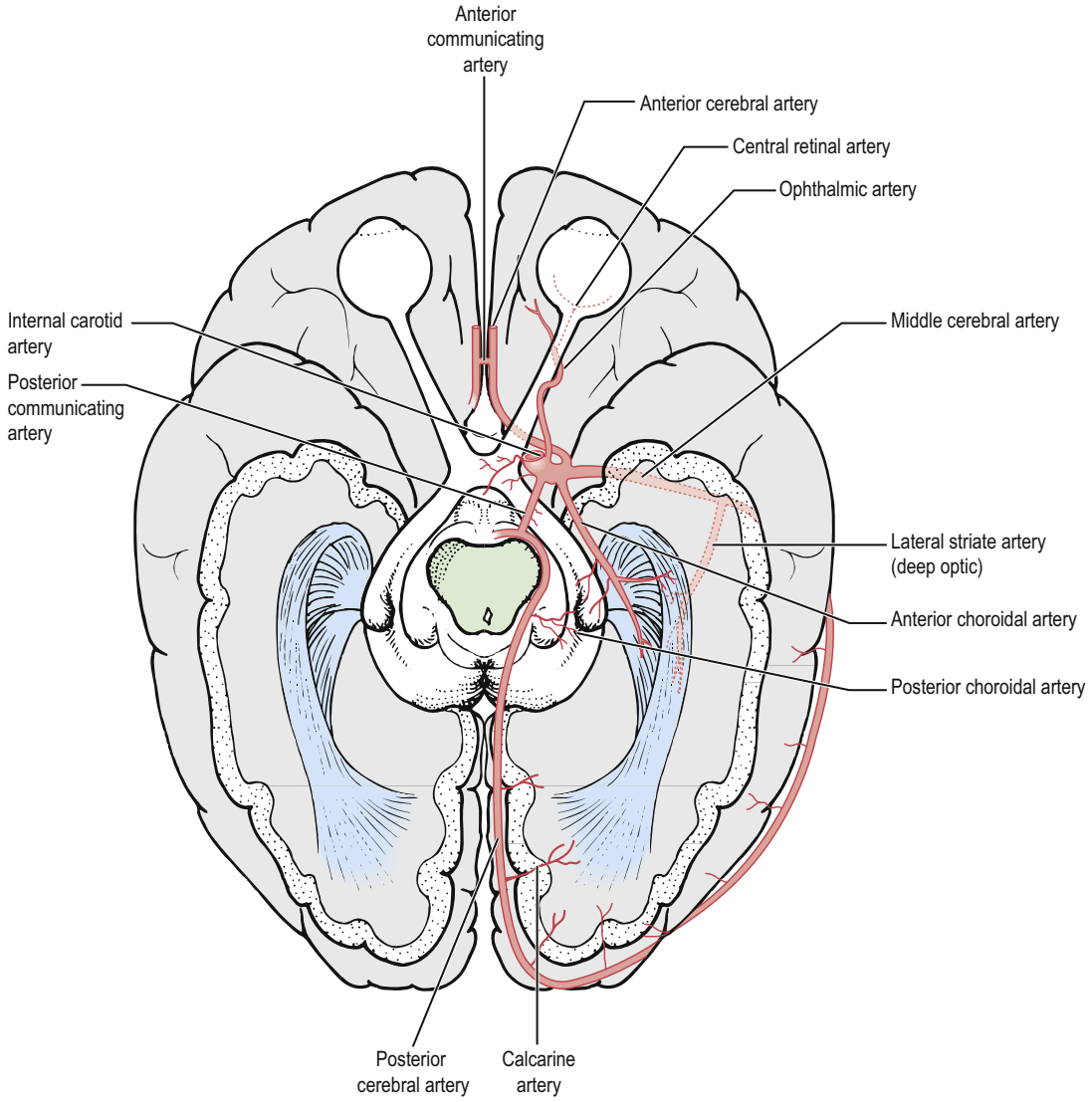


Fig. 1.76 Diagram of a brain dissected to display the visual pathways as seen from the ventral aspect. The blood supply to the various parts of the visual pathways is shown in *red* on the right-hand side of the diagram (corresponding to the *left side of the brain*). Note the blood supply to the following areas:

- Intracranial optic nerve: ophthalmic artery (an important inferior relation) and pial branches of the hypophyseal artery.
- Optic chiasma: adjacent related vessels including the superior hypophysial, internal carotid, posterior communicating , anterior cerebral and anterior communicating artery.
- Lateral root of the optic tract: anterior choroidal artery.
- Lateral geniculate body: anterior choroidal artery and branches of posterior cerebral artery.
- Commencement of the optic radiation (geniculocalcarine tract): anterior choroidal artery.
- Posteriorly directed fibres: lateral striate (deep optic) branch of the middle cerebral artery.
- Termination of geniculocalcarine tract and visual cortex: perforating branches of cortical arteries, principally the calcarine branch of the posterior cerebral although the middle cerebral may anastomose and aid in the supply of the cortex at the anterior end of the calcarine sulcus and at the posterior pole.

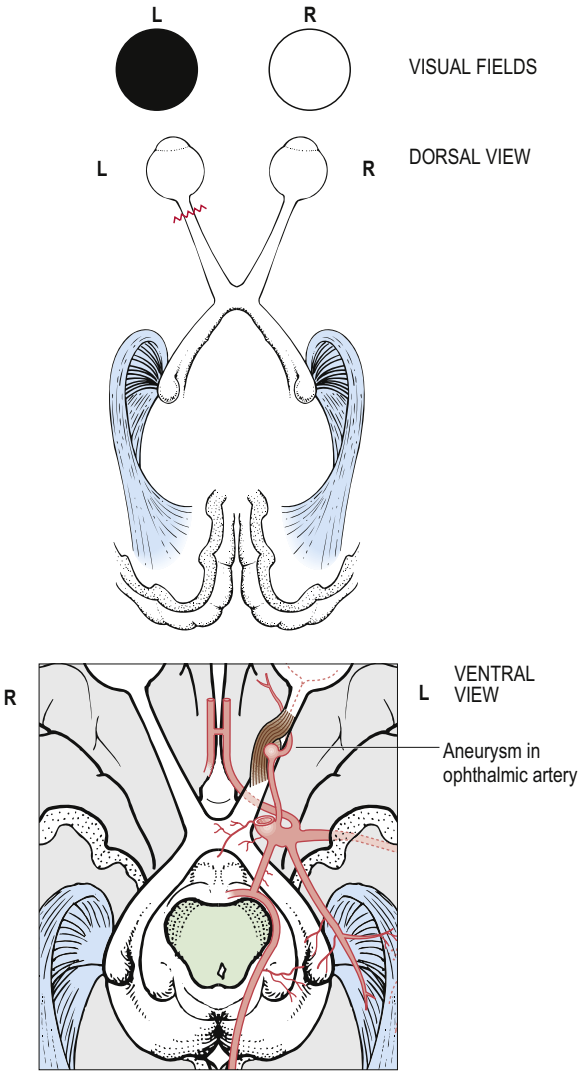


Fig. 1.77 Blindness in the left eye (*top panel* shows visual field deficit) caused by a lesion in the left optic nerve *(middle panel)*. An example of such a lesion, an aneurysm in the ophthalmic artery, is illustrated (*bottom panel:* ventral view of the brain).

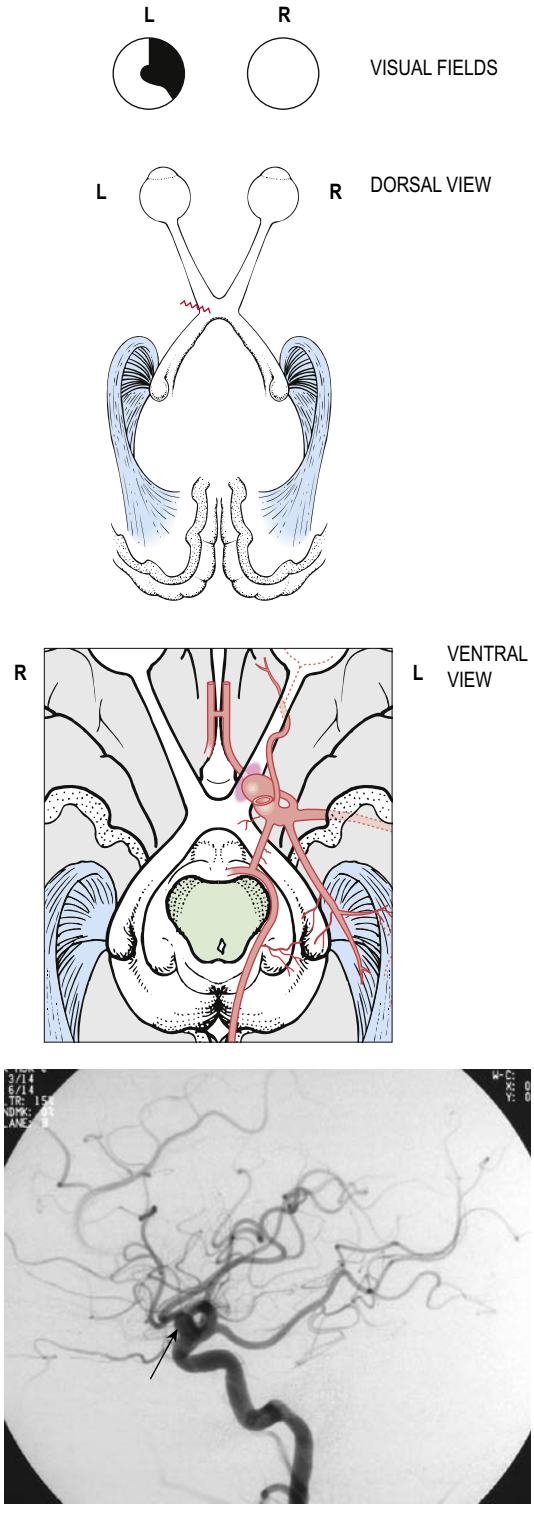
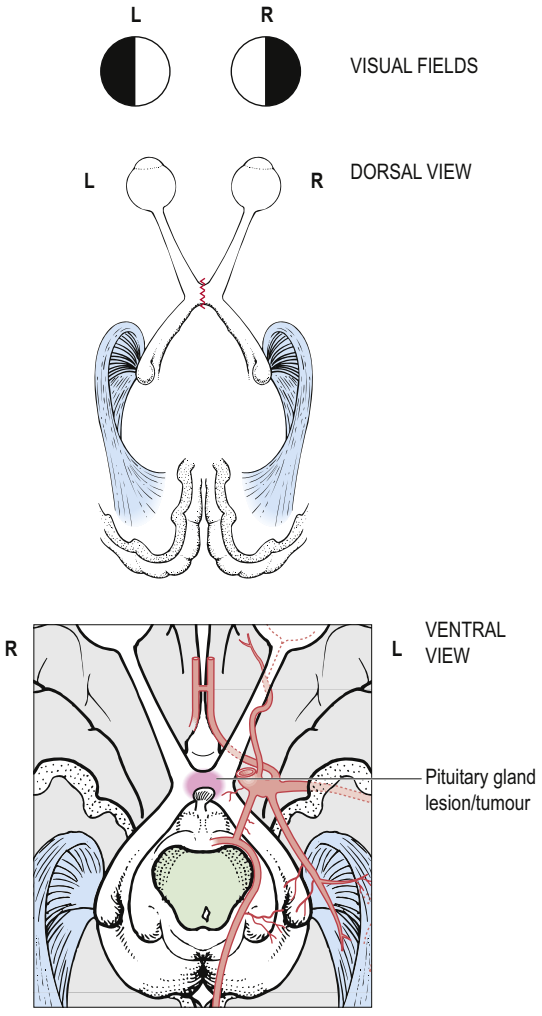
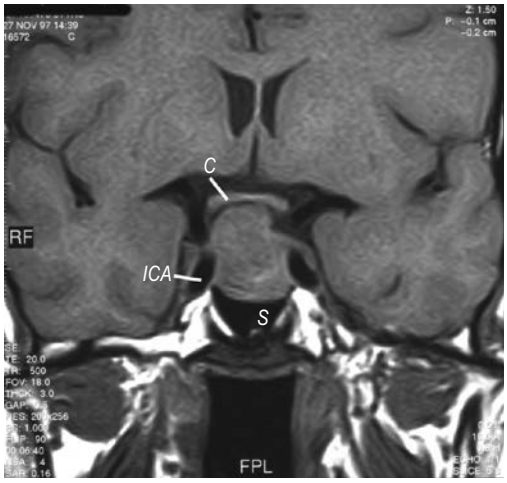


Fig. 1.78 (see right) Incongruous ipsilateral nasal hemianopia *(top panel)* caused by a lesion on the left side of the optic chiasma *(middle panel)*. An example of such a lesion is an aneurysm of the terminal portion of the internal carotid artery *(bottom panel)*. The radiographic image shows the digitally subtracted arterial phase of a carotid arteriogram of a 48-year-old patient suffering incongruous ipsilateral nasal hemianopia caused by such an aneurysm *(arrow)*. (Radiographic image courtesy Prof. T. Chakera, Royal Perth Hospital.)





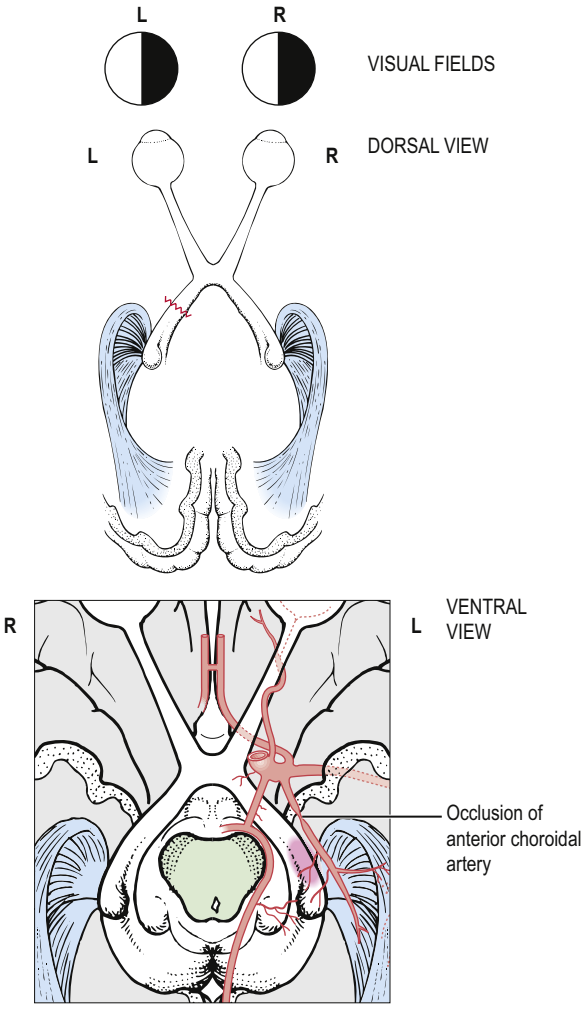


Fig. 1.80 Contralateral homonymous hemianopia *(top panel)* caused by a lesion in the left optic tract *(middle panel)*. Such a lesion would damage uncrossed fibres from the temporal retina of the left eye and crossed fibres from the nasal retina of the right eye and therefore cause disturbances in the right visual field. Causes of this type of deficit may include vascular disturbances such as occlusion of the anterior choroidal artery *(bottom panel)*.

Fig. 1.79 (see left) Contralateral bitemporal homonymous hemianopia *(top panel)* caused by interruption or damage to the nasal retinal fibres decussating in the optic chiasma *(middle panel)*. A common cause of such deficits is pituitary tumours *(bottom panel)*. Radiograph: coronal MR image of the sellar region in a 31-year-old patient who presented with bitemporal hemianopia. A large pituitary tumour can be seen compressing the optic chiasma *(C)*. *ICA*, Internal carotid artery; *S*, sphenoid sinus. (Radiographic image courtesy Prof. T. Chakera, Royal Perth Hospital.)

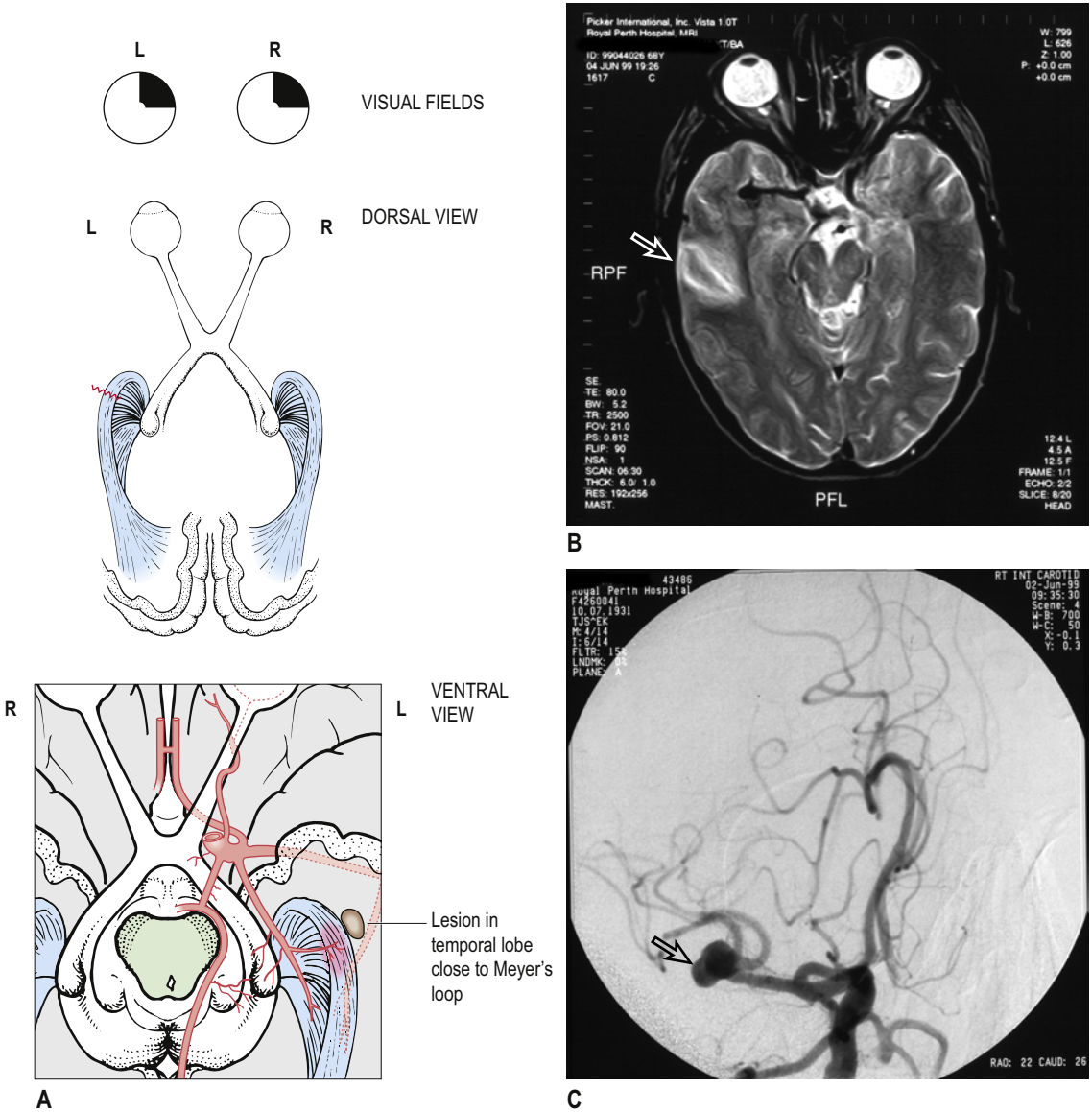
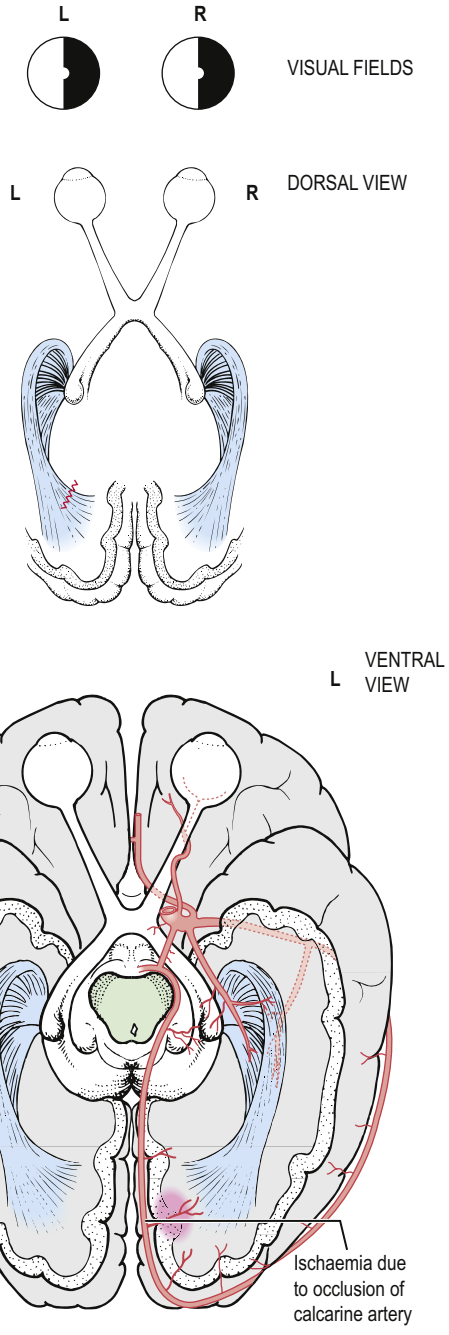


Fig. 1.81 (**A**) Contralateral homonymous superior quadrantanopia *(top panel)*, so-called 'pie in the sky' defects, may result from lesions in the temporal lobe affecting the fibres furthest into the optic radiation *(bottom panel)* (see Fig. 1.73B), which are derived from the inferior retinal quadrants and therefore cause deficits in the superior visual field. (**B**) Radiographic image of the brain of a 68-year-old patient with a stroke in the right middle cerebral artery (MCA) territory. Axial MR scan (diffusion weighted) shows the size and limits of the right temporoparietal infarct *(arrow)*. (**C**) The digital subtraction carotid arteriogram (anteroposterior projection) shows the aneurysm *(arrow)* at the bifurcation of the right MCA. The extent of the ischaemic changes in the temporal lobe (involving the optic radiation) are consistent with the patient's loss of the left visual field causing him to bump into objects. (Radiographic images courtesy Prof. T. Chakera, Royal Perth Hospital.)



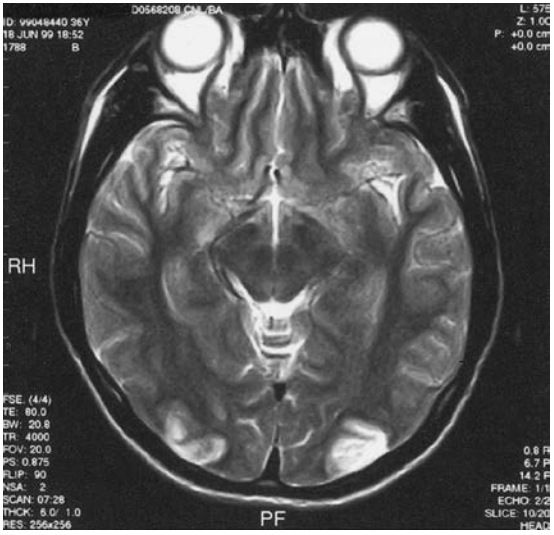


Fig. 1.82 Contralateral homonymous hemianopia with macular sparing *(top left panel)*. This may be the result of lesions affecting portions of the occipital cortex, such as tumours or infarcts *(bottom panel)*. Lesions affecting the entire occipital cortex can cause complete blindness. One such case is shown: MR scan of the brain in a 36-year-old intravenous drug user who presented with occipital headaches after injecting speed (amphetamine sulphate) with a dirty needle and waking 2 days later with total blindness and occipital headaches. Early CT scan (not shown) showed no sign of infarction but MR scan and diffusion-weighted imaging (not shown) clearly showed bilateral occipital lobe infarcts, which would explain the bilateral cortical blindness in this subject. (Radiographic image courtesy Prof. T. Chakera, Royal Perth Hospital.)

Blood supply of the visual pathway

The blood supply of the visual pathways is summarized in Fig. 1.76.

A further reading list is available online at https//expertconsult.inkling.com

References

Chinnery, H.R., Pearlman, E., McMenamin, P.G., 2008. Cutting edge: Membrane nanotubes in vivo: a feature of MHC class II+ cells in the mouse cornea. J. Immunol. 180 (9), 5779–5783. PMID:18424694.

Chinnery, H.R., McMenamin, P.G., Dando, S.J., 2017. Macrophage physiology in the eye. Pflugers. Arch. 469 (3–4), 501–515. https://doi.org/10.1007/s00424-017-1947-5. Epub 2017 Feb 23. Review. PMID: 28233124.

Forrester, J.V., McMenamin, P.G., Dando, S.J., 2018. CNS infection and immune privilege. Nat. Rev. Neurosci. 19 (11), 655–671. https://doi.org/10.1038/s41583-018-0070-8. Review. PMID: 30310148.

Freddo, T.F., 2013. A contemporary concept of the blood-aqueous barrier. Prog. Retin. Eye. Res. 32, 181–195. https://doi.org/10.1016/j.preteyeres.2012.10.004. Epub 2012 Nov 2. Review. PMID: 23128417.

Gonzalez, J.M. Jr, Ko, M.K., Hong, Y.K., Weigert, R., Tan, J.C.H., 2017. Deep tissue analysis of distal aqueous drainage structures and contractile features. Sci. Rep. 7(1), 1701. doi: 10.1038/s41598-017-16897-y. PMID: 29213129

Labetoulle, M., Baudouin, C., Calonge, M., Merayo-Lloves, J., Boboridis, K.G., Akova, Y.A., et al., 2019. Role of corneal nerves in ocular surface homeostasis and disease. Acta Ophthalmol 97 (2), 137–145. https://doi.org/10.1111/aos.13844. Epub 2018 Sep 17. Review. PMID: 30225941.

O'Koren, E.G., Yu, C., Klingeborn, M., Wong, A.Y.W., Prigge, C.L, Mathew R, et al., 2019. Microglial function is distinct in different anatomical locations during retinal homeostasis and degeneration. Immunity. pii: S1074-7613(19)30073-1. doi: 10.1016/j.immuni.2019.02.007. [Epub ahead of print]. PMID: 30850344

Reinoso, R., Martín-Sanz, R., Martino, M., Mateo, M.E., Blanco-Salado, R., Calonge, M., et al., 2012. Topographical distribution and characterization of epithelial cells and intraepithelial lymphocytes in the human ocular mucosa. Mucosal. Immunol 5 (4), 455–467. https://doi.org/10.1038/mi.2012.27. Epub 2012 May 2. PMID: 22549743.

Siegal A, Sapru HN. Essential Neuroscience. Essential Neuroscience, second ed. Lippincott Williams & Wilkins. ISBN-13: 978-1496382405, ISBN-10: 1496382404

Seyed-Razavi, Y., Hickey, M.J., Kuffová, L., McMenamin, P.G., Chinnery, H.R., 2013. Membrane nanotubes in myeloid cells in the adult mouse cornea represent a novel mode of immune cell interaction. Immunol Cell Biol 91 (1), 89–95. https://doi.org/10.1038/icb.2012.52. Epub 2012 Nov 13. PMID 23146944.

Truong, T.N., Li, H., Hong, Y.K., Chen, L., 2014. Novel characterization and live imaging of Schlemm's canal expressing Prox-1. PLoS One 9 (5), e98245. https://doi.org/10.1371/journal.pone.0098245. eCollection 2014, PMID 24827370.

Further reading

Ahnelt, P., Kolb, H., 1994. Horizontal cells and cone photoreceptors in human retina: a Golgi-electron microscopic study of spectral connectivity. J. Comp. Neurol. 343, 406–427.

Ambati, J., Anand, A., Fernandez, S., Sakurai, E., Lynn, B.C., Kuziel, W.A., et al, 2003. An animal model of age-related macular degeneration in senescent Ccl-2- or Ccr-2-deficient mice. Nat. Med. 9 (11), 1390–1397.

Balazs, E.A., 1982. Functional anatomy of the vitreous. In: Jakobiec, F.A. (Ed.), Ocular Anatomy, Embryology and Teratology. Harper and Row, Philadelphia, pp. 425–440.

Bishop, P.N., 2000. Structural macromolecules and supramolecular organisation of the vitreous gel. Prog. Retin. Eye Res. 19, 323–344.

Bishop, P.N., Takanosu, M., Le Goff, M., Mayne, R., 2002. The role of the posterior ciliary body in the biosynthesis of vitreous humour. Eye 16 (4), 454–460.

Butler, T.L., McMenamin, P.G., 1996. Resident and infiltrating immune cells in the uveal tract in the early and late stages of experimental autoimmune uveoretinitis. Invest. Ophthalmol. Vis. Sci. 37, 2195–2210.

Ceelen, P.W., Lockridge, A., Newman, E.A., 2001. Electrical coupling between glial cells in the rat retina. Glia 35, 1–13.

Chan-Ling, T., 1994. Glial, neuronal and vascular interactions in the mammalian retina. Prog. Ret. Eye Res. 13, 357–389.

Chinnery, H.R., Ruitenberg, M.J., Plant, G.W., Pearlman, E., Jung, S., McMenamin, P.G., 2007. The chemokine receptor CX3CR1 mediates homing of MHC class II-positive cells to the normal mouse corneal epithelium. Invest. Ophthalmol. Vis. Sci. 48, 1568–1574.

Christensen, A.M., 2004. Assessing the variation in individual frontal sinus outlines. Am. J. Phys. Anthropol. 127, 291–295.

Cruzat, A., Witkin, D., Baniasadi, N., Zheng, L., Ciolino, J.B., Jurkunas, U.V., et al, 2011. Inflammation and the nervous system: the connection in the cornea in patients with infectious keratitis. Invest. Ophthalmol. Vis. Sci. 52 (8), 5136–5143.

Dacey, D.M., Liao, H.W., Peterson, B.B., Robinson, F.R., Smith, V.C., Pokorny, J., et al, 2005. Melanopsin-expressing ganglion cells in primate retina signal colour and irradiance and project to the LGN. Nature 17 (433), 749–754.

Dartt, D.A., 2002. Regulation of mucin and fluid secretion by conjunctival epithelial cells. Prog. Ret. Eye Res. 21, 555–576.

Demer, J.L., 2004. Pivotal role of orbital connective tissue in binocular alignment and strabismus. Invest. Ophthalmol. Vis. Sci. 45, 729–738.

Demer, J.L., Poukens, V., Miller, J.M., Micevych, P., 1997. Innervation of extraocular pulley smooth muscle in monkeys and humans. Invest. Ophthalmol. Vis. Sci. 38, 1774–1785.

Diaz-Araya, C.M., Provis, J.M., Penfold, P.L., Billson, F.A., 1995. Development of microglial topography in human retina. J. Comp. Neurol. 363, 53–68.

Dick, A.D., Carter, D., Robertson, M., Broderick, C., Hughes, E., Forrester, J.V., et al, 2003. Control of myeloid activity during retinal inflammation. J. Leuk. Biol. 74 (2), 161–166.

Eggers, H.M., 1982. Functional anatomy of the extraocular muscles. In: Jakobiec, F.A. (Ed.), Ocular Anatomy, Embryology and Teratology. Harper and Row, Philadelphia, pp. 783–834.

Eisner, G., 1982. Clinical anatomy of the vitreous. In: Jakobiec, F.A. (Ed.), Ocular Anatomy, Embryology and Teratology. Harper and Row, Philadelphia, pp. 391–424.

Ertrurk, M., Kayalioglu, G., Govsa, F., Varol, T., Ozgur, T., 2005. The cranio-orbital foramen, the groove on the lateral wall of the human orbit, and the orbital branch of the middle meningeal artery. Clin. Anat. 18, 10–14.

Fitzgerald, M.J.F., 1992. Neuroanatomy. Basic and Clinical, second ed. Baillière Tindall, London.

Forrester, J.V., 2003. Macrophages eyed in macular degeneration. Nat. Med. 9, 1350–1351.

Gausas, R.E., 2004. Advances in applied anatomy of the eyelid and orbit. Curr. Opin. Ophthalmol. 15, 422–425.

Gebert, A., Pabst, R., 1999. M cells at locations outside the gut. Semin. Immunol. 11, 165–170.

Guymer, R.H., Bird, A., Hageman, G., 2005. Cytoarchitecture of choroid capillary endothelial cells. Invest. Ophthalmol. Vis. Sci. 45, 1660–1666.

Halfter, W., Dong, S., Schurer, B., Ring, C., Cole, G.J., Eller, A., 2005. Embryonic synthesis of the inner limiting membrane and vitreous body. Invest. Ophthalmol. Vis. Sci. 46 (6), 2202–2209.

Hamrah, P., Zhang, Q., Liu, Y., Dana, M.R., 2002. Novel characterization of MHC class II-negative population of resident corneal Langerhans cell-type dendritic cells. Invest. Ophthalmol. Vis. Sci. 43, 639–646.

Hingorani, M., Metz, D., Lightman, S.L., 1997. Characterisation of the normal conjunctival leukocyte population. Exp. Eye Res. 64, 905–912.

Ihanamaki, T., Pelliniemi, L.J., Vuorio, E., 2004. Collagens and collagen-related matrix components in the human and mouse eye. Prog. Ret. Eye Res. 23, 403–434.

Jakobiec, F.A., 1982. Ocular Anatomy, Embryology and Teratology. Harper and Row, Philadelphia.

Kabil, M.S., Eby, J.B., Shahinian, H.K., 2005. Endoscopic vascular decompression versus microvascular decompression of the trigeminal nerve. Minim. Invasive Neurosurg. 48 (4), 207–212.

Knop, E., Knop, N., 2005. The role of eye-associated lymphoid tissue in corneal immune protection. J. Anat. 206, 271–285.

Kolb, H., Linberg, K.A., Fisher, S.K., 1992. Neurons of the human retina: a Golgi study. J. Comp. Neurol. 318, 147–187.

Kolb, H., Fernandez, E., Schouten, J., Ahnelt, P., Linberg, K.A., Fisher, S.K., 1994. Are there three types of horizontal cell in the human retina? J. Comp. Neurol. 343, 370–386.

Koornneef, L., 1982. Orbital connective tissue. In: Jakobiec, F.A. (Ed.), Ocular Anatomy, Embryology and Teratology. Harper and Row, Philadelphia, pp. 835–857.

Last, R.J., 1968. Wolff's Anatomy of the Eye and Orbit. Lewis, London.

Lockwood, C., 1886. The anatomy of the muscles, ligaments, fascia of the orbit etc. J. Anat. Physiol. 20, 1.

Lütjen-Drecoll, E., 1998. Functional morphology of the trabecular meshwork in primate eyes. Prog. Retin. Eye. Res. 18, 91–119.

Mariani, A.P., 1990. Amacrine cells of the rhesus monkey retina. J. Comp. Neurol. 301, 382–400.

May, C.A., 2005. Non-vascular smooth muscle cells in the human choroid: distribution, development and further characterization. J. Anat. 207, 381–390.

May, C.A., Neuhuber, W., Lutjen-Drecoll, E., 2004. Immunohistochemical classification and functional morphology of human choroidal ganglion cells. Invest. Ophthalmol. Vis. Sci. 45, 361–367.
McClung, J.R., Allman, B.L., Dimitrova, D.M., Goldberg, S.J., 2006. Extraocular connective tissues: a role in human eye movements? Invest. Ophthalmol. Vis. Sci. 47, 202–205.
McMenamin, P.G., Lee, W.R., Aitken, D.A.N., 1986. Age-related changes in the human outflow apparatus. Ophthalmol 93, 194–208.
McMenamin, P.G., Forrester, J.V., 1999. Dendritic cells in the central nervous system and eye and their associated supporting tissues. In: Lotze, M.T., Thomson, A.W. (Eds.), Dendritic Cells: Biology and Applications. Academic Press, San Diego, pp. 205–248.
Muller, L.J., Pels, L., Vrensen, F.J.M., 1995. Novel aspects of the organisation of human corneal keratocytes. Invest. Ophthalmol. Vis. Sci. 36, 2557–2567.
Muller, L.J., Marfurt, C.F., Kruse, F., Tervo, T.M.T., 2003. Corneal nerves: structure, contents and function. Exp. Eye Res. 76, 521–542.
Nimmerjahn, A., Kirchhoff, F., Helmchen, F., 2005. Resting microglial cells are highly dynamic surveillants of brain parenchyma in vivo. Science 308 (5726), 1314–1318.
Obata, H., Yamamoto, S., Horiuchi, H., Machinami, R., 1995. Histopathologic study of human lacrimal gland. Ophthalmol 102, 678–686.
Penfold, P.L., Madigan, M.C., Provis, J.M., 1991. Antibodies to human leucocyte antigens indicate subpopulations of microglia in human retina. Vis. Neurosci. 7, 383–388.
Qiao, H., Hisatomi, T., Sonoda, K.H., Kura, S., Sassa, Y., Kinoshita, S., et al, 2005. The characterisation of hyalocytes: the origin, phenotype, and turnover. Br. J. Ophthalmol. 89 (4), 513–517.
Rada, J.A., Shelton, S., Norton, T.T., 2006. The sclera and myopia. Exp. Eye Res. 82, 185–200.
Raviola, G., 1977. The structural basis of the blood–ocular barriers. Exp. Eye Res. 25(suppl.), 27–63.
Robinson, S., Dreher, Z., 1990. Müller cells in the adult rabbit retinae: morphology, distribution and implications for function and development. J. Comp. Neurol. 292, 178–192.
Rohen, J., 1964. Ciliarkörper. In: von Möllendoff, W. (Ed.), Das Auge und seine Hilfsorgane. Ergänzung zu Band 111/2. Haut und Sinnesorgane. 4 Teil. Handbuch Der Mikroskopischen Anatomie des Menschen. Springer, Berlin, pp. 189–238.
Rowe, M.H., 1991. Functional organisation of the retina. In: Dreker, B., Robinson, S.R. (Eds.), Neuroanatomy of the Visual Pathways and Their Development. In: Cronly-Dillon, J., general. (Eds.), Vision and Visual Dysfunction, vol. 3. Macmillan UK, Basingstoke, pp. 1–68.
Rowland, L.P., 1985. Blood–brain barrier, cerebrospinal fluid, brain edema, and hydrocephalus. In: Kandel, E.R., Schwartz, J.H. (Eds.), Principles of Neuroscience. Elsevier, New York, pp. 837–844.
Ruskell, G.L., 1971a. Facial parasympathetic innervation of the choroidal blood vessels in the monkey. Exp. Eye Res. 12, 166–172.
Ruskell, G.L., 1971b. The distribution of autonomic postganglionic nerve fibres to the lacrimal gland in monkeys. J. Anat. 109, 229–242.
Ruskell, G.L., 2004. Distribution of pterygopalatine ganglion efferents to the lacrimal gland in man. Exp. Eye Res. 78, 329–335.
Ruskell, G.L., Kjellevold Haugen, I.B., Bruenech, J.R., van der Werf, F., 2005. Double insertions of extraocular rectus muscles in humans and the pulley theory. J. Anat. 206, 295–306.
Sebag, J., 2004. Seeing the invisible: the challenge of imaging vitreous. J. Biomed. Opt. 9 (1), 38–46.
Snell, R.S., Lemp, M.A., 1989. Clinical Anatomy of the Eye. Blackwell Scientific, Boston.
Stewart, P.A., Tuor, U.I., 1994. Blood–eye barriers in the rat: correlation of ultrastructure with function. J. Comp. Neurol. 340, 566–576.
Stiemke, M.M., Watsky, M.A., Kangas, T.A., Edelhauser, H.F., 1995. The establishment and maintenance of corneal transparency. Prog. Ret. Eye. Res. 14 (1), 109–140.
Thale, A., Tillman, B., 1993. The collagen architecture of the sclera: SEM and immunohistochemical studies. Ann. Anat. 175, 215–220.
Thermos, K., 2003. Functional mapping of somatostatin receptors in the retina: a review. Vision Res. 43 (17), 1805–1815.
Van Buskirk, E.M., 1989. The anatomy of the limbus. Eye 3, 101–108.
Waring, G.O., Bourne, W.M., Edelhaauser, H.F., Kenyon, K.R., 1982. The corneal endothelium. Normal and pathologic structure and function. Ophthalmol 89, 531–590.
Watson, P.G., Young, R.D., 2004. Scleral structure, organisation and disease. A review. Exp. Eye Res. 78, 609–623.
Williams, P.L., Warwick, R., 1980. Gray's Anatomy, Thirty, sixth ed. Churchill Livingstone, Edinburgh.
Wotherspoon, A.C., Hardman-Lea, S., Isaccson, P.G., 1994. Mucosa-associated lymphoid tissue (MALT) in the human conjunctiva. J. Pathol. 174, 33–37.
Yemelyanov, A.Y., Katz, N.B., Bernstein, P.S., 2001. Ligand-binding characterization of xanthophyll carotenoids to solubilized membrane proteins derived from human retina. Exp. Eye Res. 72 (4), 381–392.
Young, R.W., 1967. The renewal of photoreceptor cell outer segments. J. Cell Biol. 33, 61–72.
Zhang, H.R., 1994. Scanning electron-microscopic study of corrosion casts on retinal and choroidal angioarchitecture in man and animals. Prog. Ret. Eye Res. 13, 243–270.
Zhivov, A., Stave, J., Vollmar, B., Guthoff, R., 2005. In vivo confocal microscopic evaluation of Langerhans cell density and distribution in the normal human corneal epithelium. Graefes Arch. Clin. Exp. Ophthalmol. 243 (10), 1056–1061.

Chapter 2

Embryology and early development of the eye and adnexa

Contents

Introduction
General embryology
Ocular embryology: general introduction
Genetic regulation of eye development
Eye field transcription factors
Periocular mesenchyme is derived from a mixture of neural crest and mesoderm
The neural retina and retinal pigment epithelium are derived from neuroepithelium
Axes in the neural retina
Retinal morphogenesis
Macula and foveal development
Peripheral retina
Development of retinal vasculature
Development of the retinal pigment epithelium
Optic nerve and disc development
Development of the fibrous coat of the eye
Development of the sclera
Development of the cornea
Development of the intraocular contents
Lens development
Development of the vitreous and hyaloid system
Development of the uveal tract
The choroid
The ciliary body
The iris
Development of the anterior chamber angle and aqueous outflow pathways
Development of extraocular muscles

Introduction

This chapter aims to provide an embryological basis for understanding the anatomy of the eye and adnexa. It describes basic embryological and developmental events from the earliest formation of diverticula in the forebrain at the beginning of the fourth week to the maturation of the various components of the eye in fetal life that influence the final differentiation and functional specialization of the adult eye. The contribution of the neural ectoderm, surface ectoderm and periocular mesenchyme to the final configuration of the adult eye and surrounding tissues of the head is emphasized. The cellular interaction and influence of gene expression and transcription factors in determining cell fate will be outlined and examples are given where disruption of gene function underlies disturbances in these embryological events and the interactive processes between the various embryonic tissue and cell types lead to congenital abnormalities in humans.

General embryology

Recent innovations in genetics, genomics, genome engineering and live cell imaging have powered advances

in experimental embryology and developmental biology and helped provide an increasing understanding of the mechanisms underlying the developmental process. This is the process by which approximately 230 different cell types originate from a single fertilized egg and how they arrange themselves into tissues and organs within the dynamic 3-dimensional tissue spaces of the developing embryo. Before considering eye development, which commences in the fourth week, it might be useful to review the embryological events of the first 3 weeks after fertilization.

Further information on the first 3 weeks of embryogenesis is available online at https://expertconsult.inkling.com

Ocular embryology: general introduction

Although eye development can be considered from an embryological perspective to commence around day 22, when the *optic sulci* (optic primordium) appear as shallow grooves or pits in the inner aspect of the neural plate or neural folds (Fig. 2.1A) and the embryo is around 2 mm in length with eight somites, the group of cells that constitute the eye primordium or eye field have already begun to express a set of eye field transcription factors (EFTFs) that are highly conserved in our evolutionary ancestry (see later). The neural folds have commenced fusion to form the neural tube but the optic sulci form before they have completed their closure rostrally and caudally. When the folds in this area fuse shortly afterwards they give rise to the future *diencephalon* region of the *prosencephalic (forebrain) vesicle.* The optic sulci evaginate to form hollow diverticulae: the *optic vesicles* (Fig. 2.1B).

By about day 25 (20-somite stage) the hollow optic vesicles enlarge and become ensheathed by mesenchymal cells, except at the apex of the vesicle, which is closely apposed to the *surface ectoderm* on the lateral aspect of the developing head. This mesenchyme is derived from a mixture of cephalic neural crest and both para-axial and prechordal mesoderm.

Further information on mesenchyme is available online at https://expertconsult.inkling.com

A disc-shaped thickening of the neural ectoderm, the *retinal disc* (future neural retina), lies beneath a localized thickening of the surface ectoderm, which on day 27 is recognizable as the *lens placode* (Fig. 2.1B). Formation of the lens placode directly adjacent to the underlying neural ectoderm is regarded as one of the best examples of *induction* in developmental biology. Lens placode formation coincides with the formation of a constriction in the optic vesicle at its attachment to the wall of the forebrain to form the *optic stalk.* The *cavity of the optic vesicle* or *optic ventricle* (future subretinal space) is continuous, via the lumen in the optic stalk, with the future third ventricle (Figs 2.1B and 2.2A).

The single-layered spheroidal *optic vesicle* undergoes active invagination to fold in on itself creating two 'nested' layers of the goblet-shaped *optic cup.* The distal part of the optic vesicle becomes the inner layer or presumptive neural retina, whereas the proximal part of the optic vesicle becomes the outer layer of the optic cup, namely the presumptive retinal pigment epithelium. The thickened retinal disc at the tip or distal part of the vesicle (Fig. 2.1B,C) and the ectoderm-derived lens placode invaginate via a combination of differential growth (cell elongation and mitosis) and buckling to form the dorsal hemisphere of the optic cup and the lens vesicle. This combined invagination may also be aided by temporary fine cellular bridges between the lens placode and the retinal disc and is mediated by the extracellular matrix protein fibronectin-1 (*Fn1*), whose extent and circumferential limit beneath the lens placode is regulated by *Pax6* expression (Fig. 2.3). The active growth of the optic cup is not uniform around the circumference, which leads to the development of a groove at the distal and ventral aspect where the margins form the *choroidal* or *optic fissure.* By day 29 invagination of the retinal disc and lens placode is almost complete (Figs 2.1C and 2.2A). A small *lens pit* can be identified just before the surface ectoderm seals over the site of lens placode invagination. By the start of day 36 the *lens vesicle* separates from the surface ectoderm. The lens epithelial cells enclose the *lens cavity* and are surrounded externally by a basal lamina, which will form the future *lens capsule.* The longitudinal groove of the optic fissure, which extends into the optic stalk, acts as a temporary deficiency in the expanding and invaginating cup through which vascular mesenchyme and a branch of the ophthalmic artery, the *hyaloid artery*, become incorporated into the fissure and thus gain access to the lentoretinal space. By the end of the sixth week the growing edges of the choroidal or optic fissure meet and fuse; thus, the hyaloid vessels and associated mesenchyme become situated in the centre of the optic stalk and form the future central retinal artery and vein

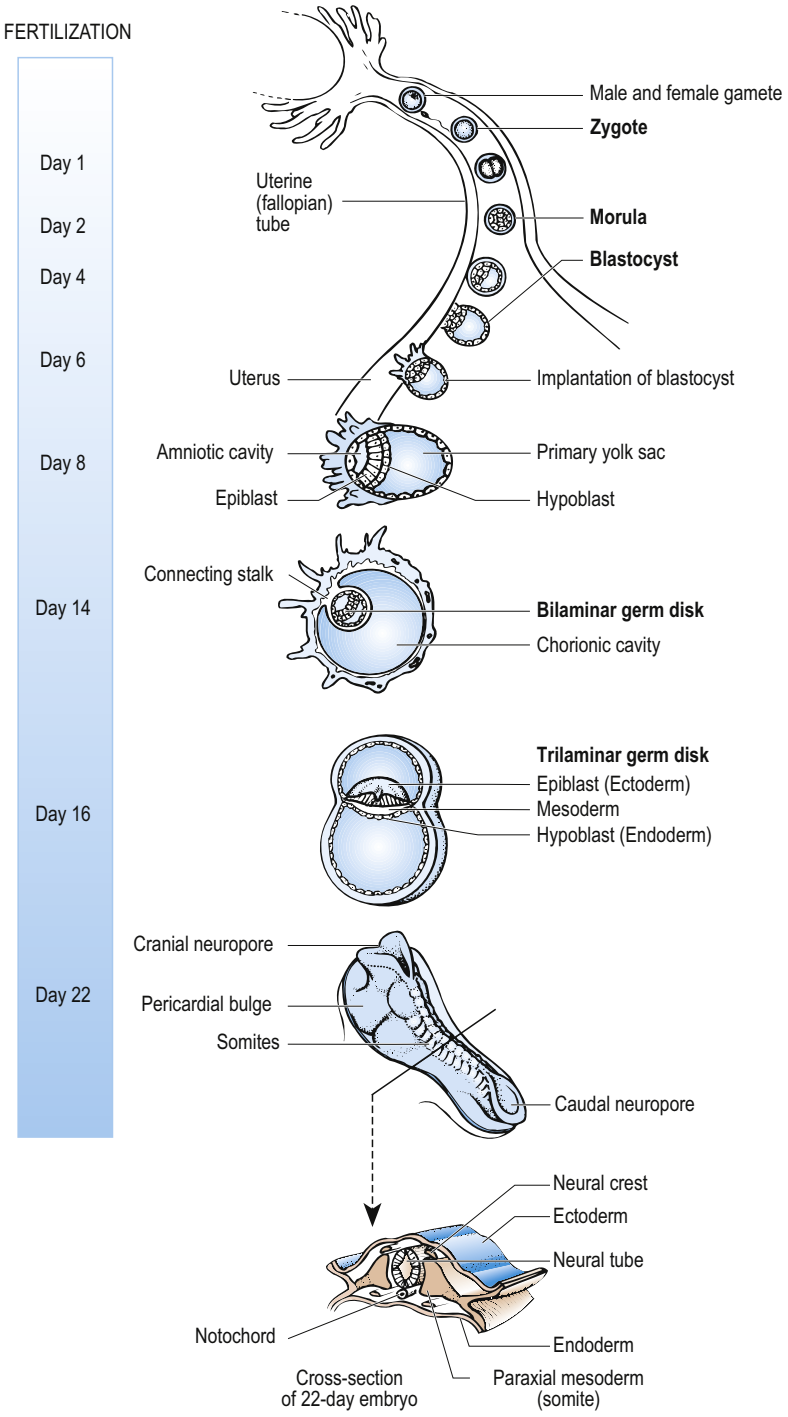
The first 3 weeks of life (see also Videos 2.1 -2.4)

The male and female *gametes* unite at fertilization, which generally occurs within the oviduct or uterine tube, to form the *zygote.* The newly formed diploid cell undergoes *cleavage* as it travels towards the uterus. Within the solid mass of cells (30-cell stage) or *morula*, the *blastocyst cavity* appears. The earliest *differentiation* of the embryo occurs around this stage when two groups of cells are formed: a peripheral *outer cell mass* or trophoblast and a central *inner cell mass* or embryoblast. The *embryoblast* will give rise to the embryo proper and some of its attached membranes. By day 5 or 6 the embryo is a hollow spheroid of around 100 cells, known as the *blastocyst.* The *embryoblast* is evident as a heterogeneous collection of cells at one pole of the blastocyst; the remainder forms the *trophoblast.* This differentiates into the syncytiotrophoblast and the cytotrophoblast, which contribute to the fetal component of the placenta. Around this time, the blastocyst enters the uterus and commences implantation into its rich endometrial lining.

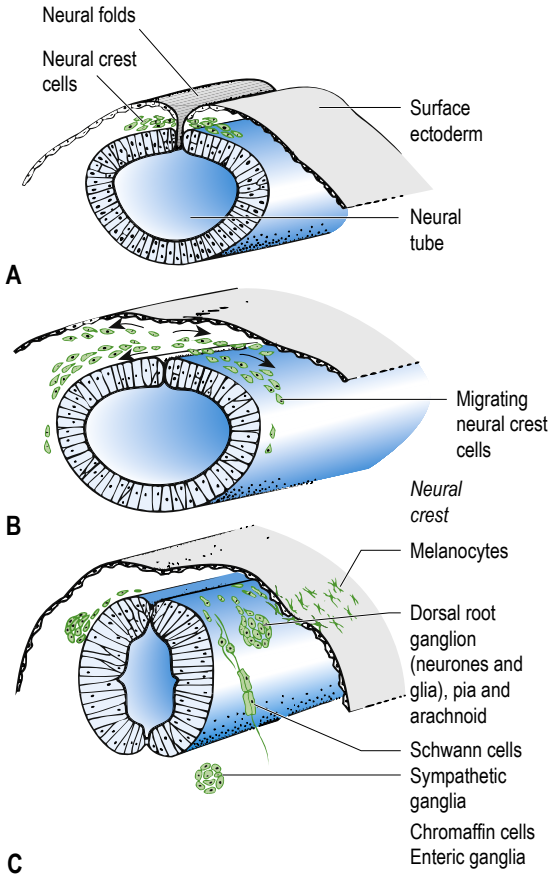
At the beginning of the second week, two cavities appear within the embryoblast. One forms the *amniotic cavity*, lined by epiblast cells, while the other forms the *yolk sac*, lined by cells derived from the hypoblast. Where the two cavities impinge there is a double discoid layer of cells, the *epiblast* or *primary ectoderm* and the *hypoblast* or *primary endoderm.* These two flat discs of cells constitute the *bilaminar germ disc*, which will develop into the embryo proper.

The third week of development commences with formation of the *primitive streak* and *primitive knot* or node at the caudal (tail) end of the epiblast. It is here that epiblast cells detach and migrate laterally and cranially into the potential space between the epiblast and the hypoblast to form the *intraembryonic mesoderm* or third germ layer. Some of the epiblast cells also replace the original hypoblast to form the *definitive* or *secondary hypoblast.* The epiblast is now known as the *ectoderm.* Formation of the primitive streak establishes the craniocaudal axis and bilateral symmetry of the future embryo. The formation of the *notochord* by budding from the primitive knot induces the formation of the *neural plate.* A series of *paraxial mesodermal* condensations or *somites* form along each side of the neural plate as it folds to form the *neural tube*, the precursor of the central nervous system (eFig. 2.1). A special group of cells, the *neural crest* cells, detach or delaminate from the margins of the neural folds and undergo extensive migration throughout the embryo where they differentiate into a remarkable variety of cells and tissues (eFig. 2.2). Derivatives of trunk neural crest cells are shown in eFig. 2.2. Derivatives in the head are more extensive.

Anatomically, the human embryo is described as being in the prone (face-down) position. The terms dorsal and ventral correspond to posterior and anterior in the adult. The terms rostral and caudal correspond to superior (head end) and inferior (tail end).



eFig. 2.1 Summary of embryological events occurring in the first 3 weeks following fertilization.



eFig. 2.2 Diagrammatic summary of (**A**) the origin, (**B**) migratory pathway, and (**C**) derivatives of neural crest cells in the trunk.

eBox 2.1

Mesenchyme

This is a term used to describe the tissue occupying the embryo between the surface ectoderm (and derivatives such as neuroectoderm) and the endoderm-derived epithelial layers. It is a loose tissue consisting of stellate amoeboid mesenchymal cells embedded in a matrix rich in glycosaminoglycans. Mesenchymal cells may be derived from several sources, namely mesoderm (dermatome or sclerotome component of the somites or lateral plate mesoderm) or neural crest. Thus, the descriptive term 'mesenchyme' does not imply an origin from any particular embryonic germ layer.

(Fig. 2.1D,E). The fusion or closure of the optic fissure commences at the mid-portion of the optic stalk and continues both proximally and distally. It is completed distally at the margins of the optic cup that will eventually form the pupil. Failure of closure of the optic fissure can occur anywhere between the cup margin (future pupil) to the optic stalk (future optic nerve) and result in a coloboma (Box 2.1 Clinical correlate). Abnormalities in more basic processes such as formation of the neural tube and optic vesicles can result in an absence of eyes, a single eye or fusion of the two eyes (Box 2.2 Clinical correlate).

Following separation of the lens vesicle from the surface ectoderm this layer regenerates and closes the lens pit and ectodermal layer in this region forms the future corneal epithelium (Fig. 2.2E). Around this time (day 39) a 'wave' of mesenchyme passes over the rim of the optic cup, directly beneath the surface ectoderm (Figs 2.1D,E and 2.2C,D). The cells in the first of the three waves of mesenchyme, which lie posteriorly closest to the lens, become flattened and form apicolateral contacts. These contacts become continuous bands of junctional complexes and thus form an endothelium. The other waves of mesenchyme will form the remainder of the cornea, iris stroma and iridocorneal angle mesenchyme.

By the end of the embryonic period (defined as the end of week 8 in humans) the retina can be clearly differentiated into a thin outer layer, which will form the *retinal pigment epithelium* (RPE) and a much thicker inner *neural retina* (Figs 2.1F and 2.2F). These two layers are separated by a narrow *intraretinal or subretinal space*, the remains of the almost obliterated ventricular cavity of the optic vesicle. Melanin first appears in the RPE around 5 weeks (Fig. 2.2C,D) and is visible on external examination of embryos of this gestational age. The neural retina commences its centrifugal differentiation into an *inner* and *outer neuroblastic layer* near its continuity with the optic stalk. The lens cavity disappears as the posterior cells elongate to form the *primary lens fibres* (Figs 2.1D and 2.2D). Mesenchyme, mostly derived from neural crest cells, condenses around the external surface of the optic cup. The innermost layer of this mesenchyme is loose and highly vascular and will form the *choroid* (uveocapillary lamina) (Fig. 2.2C,D). It lies adjacent to a distinct basement membrane formed by the RPE, which is continuous with a similar membrane around the forebrain. Indeed, the choroid is homologous in its embryonic origin with the *pia mater* and *arachnoid* investing the brain. The outer layer of the condensed mesenchyme will form the *sclera*, which is homologous to, and indeed continuous with, the dura mater around the optic nerve and brain posteriorly.

In the eighth and last week of the embryonic period (Figs 2.1F and 2.2F), ganglion cell axons grow from the inner retina towards the optic stalk. The axons travel within the stalk towards the brain, thus forming the optic nerve. Other major landmarks in development occurring in the eighth week include formation of secondary lens fibres, lens sutures and the secondary vitreous.

In summary, by the end of the embryonic period the eye comprises a double-layered neural ectoderm-derived optic cup containing a surface ectoderm-derived lens, both enveloped by condensed mesenchyme comprising a dense outer layer (the bulk of the future cornea and sclera) and an inner vascular layer, which will form the choroid and stroma of the iris and ciliary body. At this stage the human embryo is 30 mm in length (crown–rump length) and the developing eye is 1.5–2.0 mm in diameter (Figs 2.1F and 2.2F).

Genetic regulation of eye development

It is now clear that *epigenetic development*, the process by which invertebrates and vertebrates develop their definitive characteristics through the gradual alteration of simpler precursors, is regulated by *cascades of gene expression*. That is, early acting regulatory genes initiate developmental processes and induce the expression of *'downstream' genes*, which may subsequently lead to expression of further genes and so on until genes encoding actual structural and functional characteristics of specific cells and tissues are activated. There is overwhelming evidence that these cascades have been conserved throughout evolution from insects to fish to mammals. The *Drosophila*, or fruit fly, zebra fish and mouse are the most intensively studied experimental species in each respective group and recent research has shed light on some of the critical genes in ocular development, some of which have been remarkably conserved in evolution. The clues from animal studies have aided clinical geneticists to screen the DNA from patients with abnormalities and have helped reveal how mutations in single genes can cause congenital abnormalities, and the eye has provided science with some of the more elegant examples of the relevance of such genetic studies (Table 2.1) (see Chapter 3).

The basic body plan of all animal embryos is initially established by a class of regulatory genes called *selector*

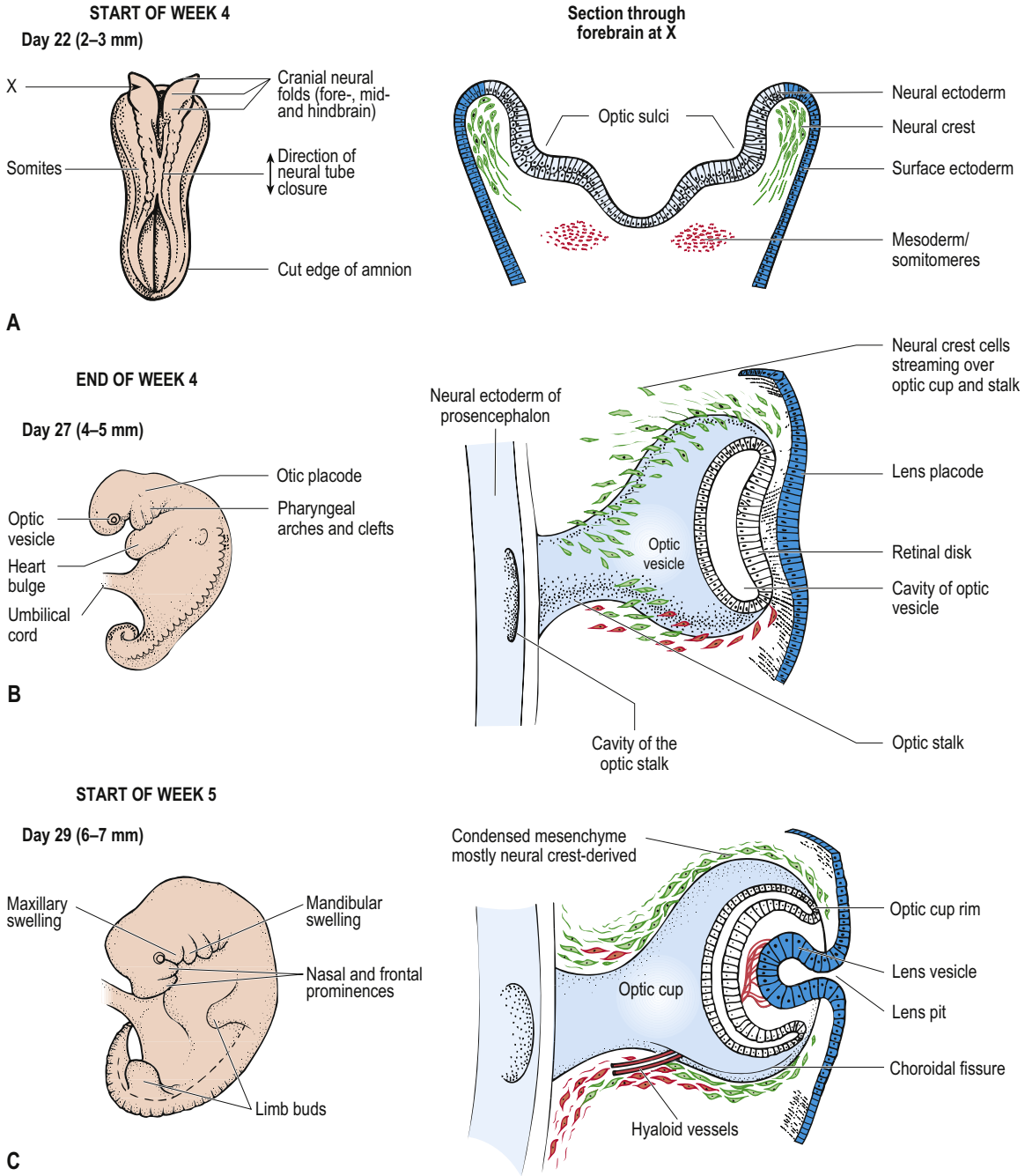
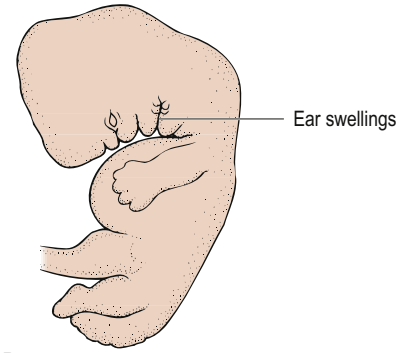


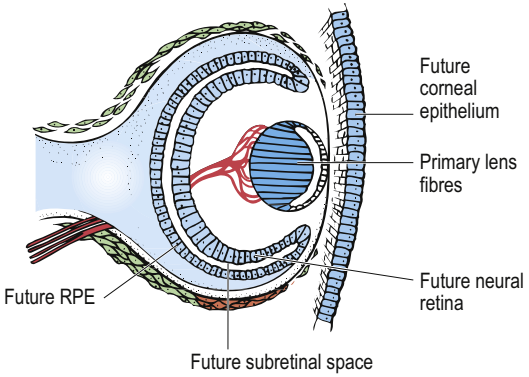
Fig. 2.1 Diagrammatic summary of ocular embryonic development from day 22 to week 8. The external appearance of the whole embryo at the equivalent period is shown on the left. The various germ layers are colour-coded to illustrate their origin and final contribution to the eye and periocular tissues.

START OF WEEK 6

Day 37 (8–11 mm)

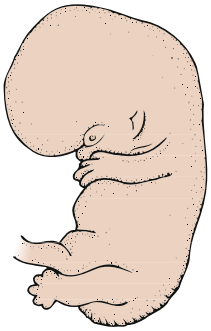


D

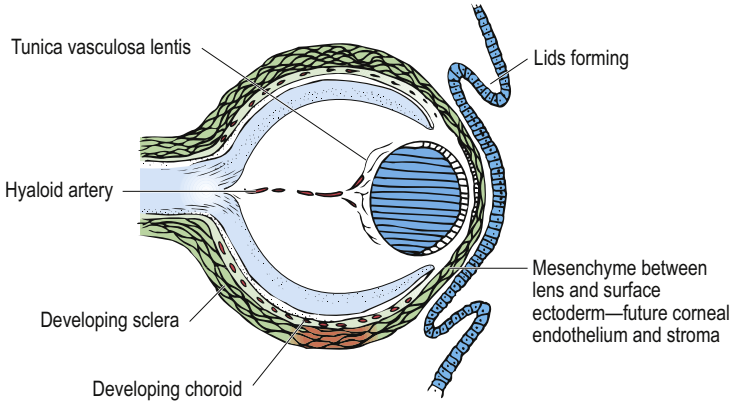


START OF WEEK 7

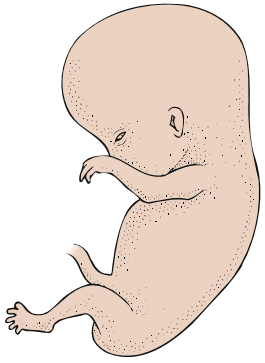
Day 44 (13–17 mm)



E



WEEK 8 (20–30 mm)



F

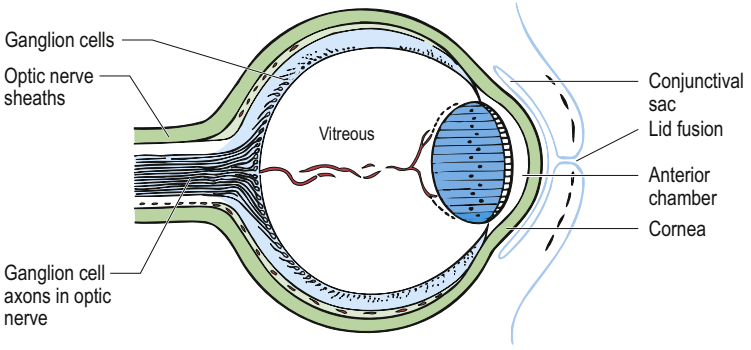


Fig. 2.1 cont'd

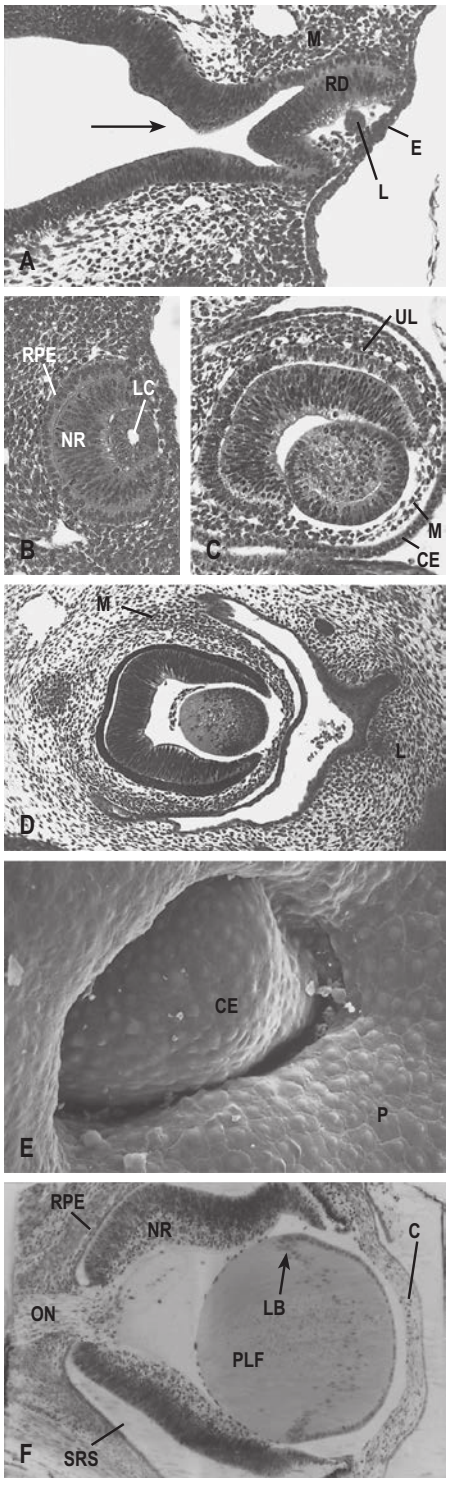


Fig. 2.2 Cont'd

Fig. 2.2 cont'd Histological and scanning electron micrographs of early mammalian eye development (chronological sequence **A–F**). (**A**) The cavity of the optic vesicle is clearly in continuity via the cavity of the optic stalk with the forebrain ventricle *(arrow)*. (**B**) The lens vesicle, containing a distinct lens cavity *(LC)*, fills the optic cup, which consists of two layers, an outer retinal pigment epithelium *(RPE)* and neural retina (*NR*). (**C**) Mesenchyme has condensed around the optic cup and migrated over the cup margin to form the future corneal endothelium/stroma *(M)* beneath the surface ectoderm-derived corneal epithelium *(CE)*. Beneath the RPE the vascular mesenchyme has already formed a distinct row of vessels, the uveocapillary lamina (*UL*) or presumptive choroid. (**D**) Pigment is identifiable in the RPE layer. The lentoretinal space contains vascular mesenchyme. The mesenchyme *(M)* around the developing eye has now formed two layers, an outer dense avascular layer (future sclera) and an inner vascular layer (future choroid). Beneath the newly formed lids *(L)* lies the conjunctival sac. (**E**) Scanning electron micrograph of the embryonic corneal epithelial surface *(CE)* and periderm *(P)* just before lid closure. (**F**) Late embryonic eye with well-developed cornea *(C)*, large lens consisting of primary lens fibres *(PLF)* whose nuclei form the lens bow *(LB)*. The neural retina *(NR)* is artefactually detached from the RPE, producing a large subretinal space (*SRS*). Axons have commenced migration along the optic nerve *(ON)*. *E*, surface ectoderm/periderm; *L*, lens vesicle; *M*, mesenchyme condensation; *RD* retinal disc. Original magnifications: **A** × 120; **B** × 180; **C** × 160; **D** × 55; **E** × 100; **F** × 65. (From McMenamin and Krause, 1993, with permission.)

or *switch genes* which, like the *maternal effect genes* of the fruit fly, establish longitudinal or antero-posterior (head–tail), dorsoventral and left–right axes. A further class of genes, the *zygotic genes*, which includes *segmentation genes*, is switched on later, after the maternal effect genes. The changes induced by the segmentation genes cause the expression of another class of selector genes, the *homeotic genes*, which encode a region of DNA called a *homeobox* (termed *Hox* genes). These subsequently regulate many downstream genes and thus act as *master control genes*.

Hox gene activation plays a critical role in the differentiation of what initially appear as identical segments in the embryo (induced by earlier expression of segmentation genes) into, for example, cervical, thoracic, abdominal and sacral regions and the segmentation of the head and neck, particularly the pharyngeal arches and subdivision of the brain.

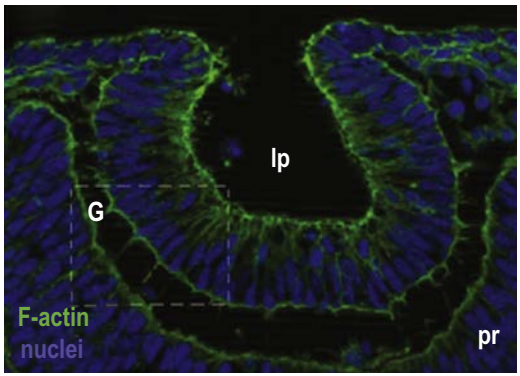
Paired-box (Pax) genes encode transcription factors involved in the regulation of several aspects of vertebrate and non-vertebrate early development and, as such, are also considered as *'master control genes'*. In vertebrates, two *Pax* genes, *Pax-6* and *Pax-2*, are important eye field transcription factors.

Eye field transcription factors

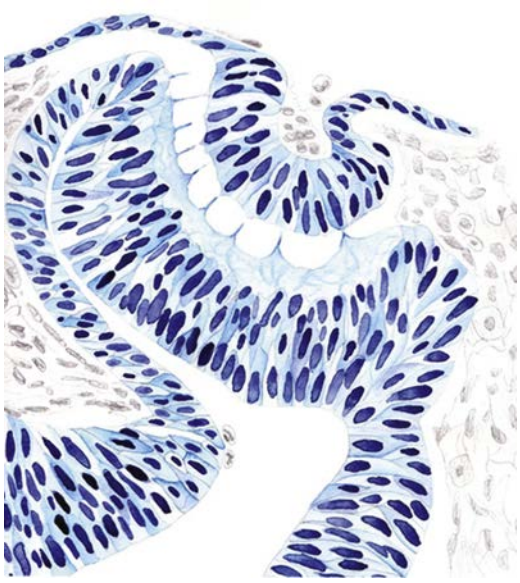
In mammals (and one must appreciate that much of this research is performed in mice) the EFTFs include *Pax-6*, *Rax, Six3* and *Lhx2* and these are expressed in an overlapping fashion in the anterior neural plate or future forebrain at the paired sites that define the site of the eye primordium or eye field (Fig. 2.4). There is evidence that *Otx2*, a transcription factor essential for forebrain development, might cooperate with *SOX2*, a neural ectoderm transcription factor, to progressively activate *Rax* expression in the eye field, which might in turn activate *Pax-6, Rax, Six3* and *Lhx2*. Evidence of the importance of these genetic transcription factors in early developmental events can be seen by considering the effects of mutations in both mouse models and humans (see Table 2.1).

The primordial eye field is initially one region of anterior neural ectoderm that has two hemispheres that later split into two respective fields, a left and a right, and at least two molecules (*sonic hedgehog (Shh)* and *Six3*) are involved in this early morphogenesis, which, if disrupted, leads to a failure in the splitting and thus to cyclopia or a single midline eye.

Pax-6 is expressed in the anterior neuroepithelium and lens placode-forming epithelium. Thus, it appears that the gene defines a field of cells *competent to differentiate into eye tissues* (neural retina, retinal pigment epithelium, iris epithelium, ciliary body epithelium, lens and corneal epithelium). It also appears to be required to *maintain growth and proliferation* of cells in the optic vesicle (and other regions of the central nervous system). Mice with mutations in *Lhx2* can generate optic vesicles but never form optic cups and it appears as if both this transcription factor and *Pax-6* are needed to form a proper eye cup and together induce *Six6* in the optic vesicle, which is essential for retinal differentiation. The boundaries between the parts of the optic vesicle that ultimately differentiate into neural retina, RPE and optic stalk are predetermined by the differential expression of cell signalling pathways along the dorsoventral and proximal–distal axes, which are in turn regulated by transcription factors (Fig. 2.4). For example, a mutually antagonistic relationship exists between expression of *Pax-2* and *Pax-6*. *Pax-2* is largely



A



B

Fig. 2.3 (**A**) Transient, F-actin-filled processes connect lens and retina during lens pit–optic cup invagination in the mouse eye. Sections from cytoplasmic protrusions interepithelial processes containing filamentous actin (F-actin, *green*) are clearly seen between the lens and future retina from E9.5 to E11.0 (embryonic days/post-conception). (**B**) Equivalent interepithelial processes were described in the early human eye many years ago by Ida Mann. *lp*, lens pit; *pr*, presumptive retina. Scale bars: 20 μm. Nuclear staining (Hoechst 33258, *blue*). (Fig. 2.3A reproduced from Chauhan et al., 2009, with permission; Fig. 2.3B redrawn from Mann, 1928.)

restricted to the future optic stalks on the ventral aspect, where it appears to be necessary for successful closure of the *choroid fissure*, whereas *Pax-6* is expressed more dorsally and is more involved with formation of the RPE and neural retina. These cell-intrinsic signalling pathways are not the only factors that determine the fate of cells or patterning in the optic vesicle. There are further extrinsic factors including members of the transforming growth factor-β (TGF-β), fibroblast growth factor (FGF) and Wnt families and *Shh* that signal to compartmentalize the optic vesicle functionally. For example, ectopic expression of FGF9 in the presumptive RPE promotes neural retina formation and inactivation of canonical Wnt signalling in the future RPE causes it to transdifferentiate into neural retina.

Periocular mesenchyme is derived from a mixture of neural crest and mesoderm

For many years it was widely held that the middle germ layer, the *mesoderm*, gave rise to most of the mesenchyme and its derivatives in the head and neck region in a similar manner to the pattern of differentiation in the trunk **(see eFig. 2.2)**

A large body of experimental evidence from avian and mammalian studies has now shown that mesenchyme in the head region is derived from two sources, neural crest (*mesectoderm*) and mesoderm. Neural crest cells originate at the neuroectoderm–surface ectoderm junction of the fore-, mid- and hindbrain regions before fusion of the neural folds. They migrate ventrally into the pharyngeal arches and rostrally around the forebrain and developing optic cup and into the facial region in a highly ordered manner, guided in their migration by components of the extracellular matrix, such as fibronectin and glycosaminoglycans (Fig. 2.5). The pattern of appearance and migration of cranial neural crest is closely associated with the expression products of the *homeobox* (*Hox*) gene family within the rhombomeres of the hindbrain. In the face region, neural crest cells contribute significantly to mesenchyme-derived tissues, such as bone, cartilage, connective tissues, meninges and ocular and periocular connective tissues (which normally do not arise from trunk neural crest), as well as giving rise to the usual crest derivatives: melanocytes, dorsal root ganglia equivalents (sensory ganglia of V, VII, IX and X) and parasympathetic ganglia (ciliary, otic, pterygopalatine and submandibular).

At the same time as cephalic neural crest cell migration commences, the optic stalks begin to constrict, thus creating a pathway between the stalk and surface ectoderm into which predominantly mesencephalic crest cells migrate. The migration of these cells ceases when they reach the choroid fissure on the ventral aspect of the optic cup.

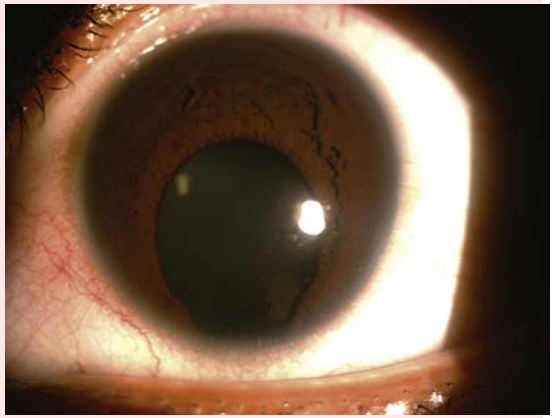
Box 2.1 **Clinical correlates**

Coloboma

The word coloboma ('a mutilation') in clinical practice is the term is used to describe a defect in the inferonasal quadrant in the iris, ciliary body or choroid. Colobomas are usually sporadic and bilateral and do not lead to significant complications. They are the result of a failure of closure of the (inferonasal) optic fissure. The consequence is an interference with the normal induction and formation of uveal tissues. A coloboma of the iris appears as an *inferonasal defect* in the stroma, the smooth muscle and the pigment epithelium, whereas a ciliary body coloboma is characterized by absence of ciliary processes and presence of a diminutive muscle. The adjacent lens is indented, owing to a failure of formation of the zonular fibres. In the colobomas associated with complex malformations and abnormalities (e.g. trisomy) there may be an ingrowth of mesodermal tissue into the retrolental space with formation of fat and cartilage.

Colobomas involving the retina may be large and may extend to the disc. In the retina adjacent to the coloboma, proliferation of neuroblastic tissue leads to the formation of rosettes. The retinal pigment epithelium and Bruch membrane and choroidal tissue are absent at the site of the coloboma, although the underlying sclera is normal. In many cases of retinal coloboma there is glial and vascular ingrowth from the retina across the bed of the coloboma. Colobomas have been described in association with mutations in the *CHD7* gene (see Table 2.1). Mutations in the *CHD7* gene are found in approximately 60%–65% of individuals with CHARGE syndrome (*C*oloboma, *H*eart disease, *A*tresia choanae, *R*etardation of growth and/or development, *G*enital hypoplasia, *E*ar malformation). Most mutations lead to the production of an abnormally short, non-functional CHD7 protein, which appears to disrupt chromatin remodelling and the regulation of gene expression.





Further details on mapping of neural crest migration are available online at https://expertconsult.inkling.com

Paraxial mesoderm (somites and the less distinct rostrally situated *somitomeres*) forms most of the walls and floor of the brain case, all voluntary muscles of the craniofacial region (including extraocular muscles), all vascular endothelial cells, the dermis and connective tissues of the dorsal region of the head and the meninges caudal to the prosencephalon. The seven rostrally situated somitomeres differentiate in a segmental manner within the pharyngeal arches and they play an important role in eye development by influencing neural crest cell migration and differentiation and by directly contributing to the periocular mesenchyme (Fig. 2.6).

Thus, mesodermally derived mesenchyme contributes more to the periocular connective tissues than experimental studies in birds had previously suggested.

The neural retina and retinal pigment epithelium are derived from neuroepithelium

The thickened portion of the optic vesicle that invaginates, the *retinal disc*, is destined to differentiate into the *neural retina*, while the thinner outermost layer of the optic vesicle is destined to form the RPE. These layers are

eBox 2.2 Experimental methods of mapping neural crest cell migration and fate

The fate of neural crest cells from various regions of the neural plate has been mapped in birds (for reviews, see Noden, 1982, 1988; Creuzet et al., 2005); it has also been mapped in mouse and rat embryos (Erickson et al., 1989; Fukiishi and Morriss-Kay, 1992; Trainor and Tam, 1995). Avian experiments have used a variety of cell-tracing methods, including transplanted radiolabelled donor crest cells, chick–quail transplant chimeras and subsequent tracking of quail cells using natural nucleolus marker or anti-quail nuclear antigenic determinant (for review, see Creuzet et al., 2005).

Tracking the fate of neural crest cells in mammals has been aided by the production of mouse mutants with neural developmental anomalies and more recently by using transgenic mice in which a transgene (bacterial *lac Z* 'reporter' gene that codes for β-galactosidase) is introduced into the mouse genome in a position where it will be co-expressed alongside proteins specific for the cell types under investigation, such as peripherin or retinoic acid receptor, which are expressed on neural crest cells during and after migration. The migration pathways of the cells carrying the transfected genome can thus be visualized with appropriate chemical substrates for bacterial β-galactosidase (Mendelson et al., 1994). Other techniques that are unveiling the destiny and differentiation of various cell types in embryonic development include micromanipulative cell grafting and labelling using fluorescent cell markers (DiI and DiO), *in situ* hybridization and chimeric and mosaic mouse models (Osumi-Yamishita et al., 1990; Trainor and Tam, 1995; Collinson et al., 2004) (see Fig. 2.6). More recently, live cell imaging of transgenic Zebrafish embryos has provided informative time lapse movies of neural crest cells migrating around the periocular region (Kwak et al., 2013).

Box 2.2 **Clinical correlates**

Malformations of the neural tube and optic vesicle and genetic factors

These occur in the first month of embryonic life and include:

- *Anophthalmia.* With an incidence of congenital anophthalmia (in England) ranging from 0.4 to 2.9 per 100,000 infants this condition is rare. It results from of a failure of formation of the optic vesicle. The orbits do not contain ocular tissue, but the extraocular muscles (mesoderm) and lacrimal gland (ectoderm) are present. Mutations in the *RAX* and *SOX2* genes have been associated with anophthalmia but a recent review (Graw, 2019) notes there are some 145 genes associated with anophthalmia (Table 2.1).
- *Nanophthalmia and microphthalmia.* Incidence of congenital microphthalmia ranging from 10.0 to 10.8 per 100,000 children. The condition results from formation of the optic vesicle without proper subsequent development which produces a rudimentary eye in the orbit—nanophthalmia (or dwarf eye). In microphthalmia there is a small but recognizable eye that contains recognizable elements, e.g. lens, choroid and retina. The mouse mutant small eye (*Sey*) has a haploinsufficiency in *Pax-6*. Human mutations in one copy of *PAX6* have aniridia (iris abnormalities) and microphthalmia but homozygous loss causes anophthalmia. Mutations in *SOX2* in humans and in *Mitf, Pitx* in mice can also cause microphthalmia.
- *Cyclopia mutations* in *SHH or SIX3* result in midline defects including cyclopia.
- *Synophthalmia.* Fusion of the two eyes may result from a malformation of the mesenchymal tissue between the optic vesicles or faulty inductive processes. Only rarely is a single eye (cyclops) formed by this mechanism and in most cases there are two recognizable corneas and lenses, and identifiable parts of the iris and ciliary body. The midline sclera and uveal tissue may be absent and the optic nerve may be single or duplicate. This malformation may be associated with a deletion of chromosome 18.

continuous at the optic cup margin, where a sharp transition in morphology is evident (Figs 2.1B,C and 2.2A–D). The *optic cup margin* will later be the site from which the neuroepithelial component of the iris and ciliary body will arise and ultimately form the *pupil margin*. Because of the invagination of the optic cup, the apical aspect of the primitive neural retina comes to lie adjacent to the apical surface of the RPE, thereby obliterating the intraretinal space (Fig. 2.1C). Just as the ependymal cells that line the developing (and adult) ventricular spaces in the brain are ciliated, the apposing surfaces of the primitive neural retina and future RPE are also ciliated. The cilia of the neural retina are important later in development, in the formation of rods and cones. The cilia of the RPE degenerate.

Axes in the neural retina

We now understand more clearly how the axes in the retina that are critical to establishing the pattern of retinotopic projection become organized during development. Mouse studies have shown that when the inner layer of the optic cup invaginates to form the neural retina it must establish three primary axes: dorsoventral, nasal–temporal and anterior–posterior. In addition, the optic stalk must form inferiorly relative to these axes (Fig. 2.7). The signalling that is important in the dorsoventral axes is mediated through the *Pax-6* (highest expression dorsally) and *Pax-2* (highest expression ventrally) transcription factors in addition to the *VAX* family of homeodomain transcription factors. *Vax2*, in particular, has a steep ventral–dorsal gradient, whereas *Vax 1* is expressed on the optic stalk. Naso-temporal patterning is regulated by the forkhead transcription factors *FOXD1* and *FOXG1* (Fig. 2.7).

Retinal morphogenesis (Fig. 2.8)

The primitive neural retina consists of an *outer nuclear zone* and an *inner acellular* or *marginal zone*. The outer nuclear zone is homologous with the proliferative neuroepithelium of the neural tube. Both the inner and outer layers of the optic cup rest on their respective basal laminae: that of the inner layer becomes the *inner limiting membrane*, the outer is incorporated into the *Bruch membrane*. Differentiation of the retinal layers commences at the posterior pole and progresses in a centrifugal manner; thus, a gradient of retinal differentiation can be seen within an individual eye. Miotic activity in the primitive neural retina is greatest in the outer part of the nuclear zone (Fig. 2.8A). Around 7 weeks of gestation (16–20 mm) newly formed cells migrate in a vitread direction into the marginal zone to form the *inner neuroblastic layer*. The outer nucleated zone is now referred to as the *outer neuroblastic layer* (Figs 2.8B and 2.9A). The two neuroblastic layers are separated by an acellular zone—the *transient layer of Chievitz*. The earliest differentiated cells, which form the inner

Table 2.1 **Critical genes in ocular development**

Mouse gene	Human gene	Expression pattern	Ocular defects in mouse mutants	Ocular defects in humans with known mutations in related gene
Sox1 Sox 2 Sox3	*SOX1, SOX2 SOX3*	Central nervous system, sensory placodes, *Sox2* in anterior neural ectoderm and lens placode	Micro-ophthalmia, cataract	Anophthalmia
Otx2	*OTX2*	Anterior neural ectoderm Future forebrain		Range of phenotypes from bilateral anophthalmia to retinal dystrophy
Rax	*RAX*	Anterior neural plate, optic vesicle, developing retina, photoreceptors	Eyeless	Anophthalmia
Pax-6	*PAX6*	Anterior neural plate, optic sulcus/cup and stalk Surface ectoderm (future lens and corneal/conjunctival epithelium) Weakly expressed in mesenchymal cells	Micro-ophthalmia, cataract, hypoplastic iris, incomplete separation of cornea and iris, corneal defects	Anophthalmia, anterior segment dysgenesis, congenital glaucoma, Peters' anomaly, Axenfeld–Rieger syndrome (aniridia)
Six3	*SIX3*	Presumptive eye field	Lack of neural retina Aphakia	Holoprosencephaly
Pitx3	*PITX3*	Developing lens vesicle	Persistent lens stalk, Aphakia or malformed lens	Congenital cataract, leucoma, Peters' anomaly
Chd7	*CHD7*	Neuroectoderm, lens vesicle	CHARGE-like features, keratoconjunctivitis sicca	CHARGE syndrome (see Box 2.1)
Maf	*MAF*	Lens placode, lens vesicle, primary lens fibres (transcription factors for α-crystallin gene along with Sox1)	Failure of lens fibres to elongate, lens vesicle fails to separate from surface ectoderm	Defects in lens, cornea and iris (coloboma), Peters' anomaly
Foxe3	*FOXE3*	Lens placode	Failure of lens to separate from surface ectoderm	Peters' anomaly, posterior embryotoxon, cataract
Pitx2 Fox1	*PITX2 FOXC1*	Periocular mesenchyme (presumptive cornea, eyelids, trabecular meshwork, extraocular muscle)	Anterior segment abnormalities	Iridogoniodysgenesis, Axenfeld–Rieger syndrome, 50% develop juvenile glaucoma
Crya Cryb Cryg	*CRYA CRYB CRYG*	Lens	Various forms of cataract	Various forms of cataract

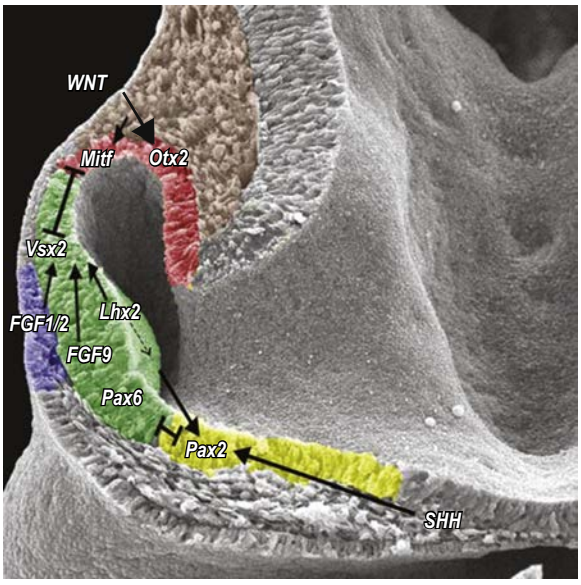


Fig. 2.4 Signalling networks establish boundaries in the optic vesicle. Dorsal is to the top and distal is to the left (note the cavity of the optic stalk communicating with the forebrain vesicle *(left)*). The optic vesicle is regionalized into prospective RPE (*red*, dorsal), neural retina (*green*, central) and optic stalk (*yellow*, ventral). Extracellular signals organize the optic vesicle in part through the activation of transcription factors that specify the tissue type in which they are expressed. These transcription factors regulate optic vesicle organization through mutual repression of one another. The lens placode, which expresses fibroblast growth factor *(FGF)* ligands important for neural retinal specification, is shown in *blue*. (Reproduced from Heavner W, Pevny L. Eye development and retinogenesis. Cold Spring Harb Perspect Biol. 2012;4:a008391, with permission.)

neuroblastic layer, are future ganglion cells, Müller cells (radial glial) and amacrine cells. Elegant studies using chimeras and mosaic mouse models have revealed that clones of cells radiate in a vitread direction and, as the retina begins to stratify, proliferation and differentiation cells that have arisen from these original 'clones' appear as columns. Subsequently some cells disperse laterally.

The nerve fibre layer becomes identifiable on the inner aspect of the inner neuroblastic layer owing to growth of ganglion cell axons that converge towards the optic stalk. A zone where the processes of cells from the inner neuroblastic layer intermingle (the *inner plexiform layer*) becomes identifiable at approximately 10.5 weeks of gestation, thereby obliterating the transient layer of Chievitz (Figs 2.8C and 2.9A). A new intermediate nucleated layer, the *inner nuclear layer*, becomes identifiable in the posterior pole retina and already contains the amacrine and Müller's cell bodies, and shortly afterwards the bipolar and horizontal cells differentiate from the outer neuroblastic layer and migrate into this new nucleated layer (Fig. 2.8C). The remaining components of the outer neuroblastic layer will form the *outer nuclear layer* containing the cell bodies of the photoreceptors (rods and cones). The zone where fibres from this layer intermingle with those of the inner nuclear layer constitutes the new *outer plexiform layer* (Fig. 2.8D). The *external limiting membrane* (not a membrane *per se*) of the retina is identifiable in the earliest stages as rows of tight junctions between adjacent neuroblasts (Fig. 2.10B).

Further important landmarks in retinal development include:

- synaptogenesis in cone pedicles at approximately 4 months, but not in rod spherules until 5 months
- photoreceptor outer segment formation commences around the fifth month
- horizontal cells become distinguishable around the fifth month
- microglia (resident tissue macrophages) invade the retina via the retinal vasculature and peripheral subretinal space (10–12 weeks onwards)
- the terminal expansions of the Müller cells beneath the inner limiting membrane mature around 4.5 months, at around the same time as their processes can be identified between the rods and cones.

The onset of cellular differentiation in the neural retina is partly dependent on Shh (sonic hedgehog) and FGF signalling. In addition, a number of bHLH (basic helix-loop-helix, transcriptional activators) genes are likely involved in determining neuronal fate. NOTCH1 signalling appears to be involved in regulating the fate of the glial cells, particularly Müller cells.

Macula and foveal development

Maculogenesis is first evident as a localized increase in ganglion cell density temporal to the optic disc at around 4.5 months. By 6 months, the ganglion cell layer may be eight or nine cells deep in this region. The thickened immature outer nuclear layer consists predominantly of immature cones. In the seventh month, there is a displacement of ganglion cells and formation of a *foveal depression*. There are approximately two layers of ganglion cells in the foveal region in the eighth month and at birth this is reduced to one. By 4 months postpartum, the inner nuclear and ganglion cell layers have receded to the margins of the fovea, leaving only cone nuclei in the foveal region. Elongation of the inner and outer segments occurs over the next few months.

Peripheral retina

Until approximately 10–12 weeks of gestation, the periphery of the retina extends to within 50–100 μm of the optic

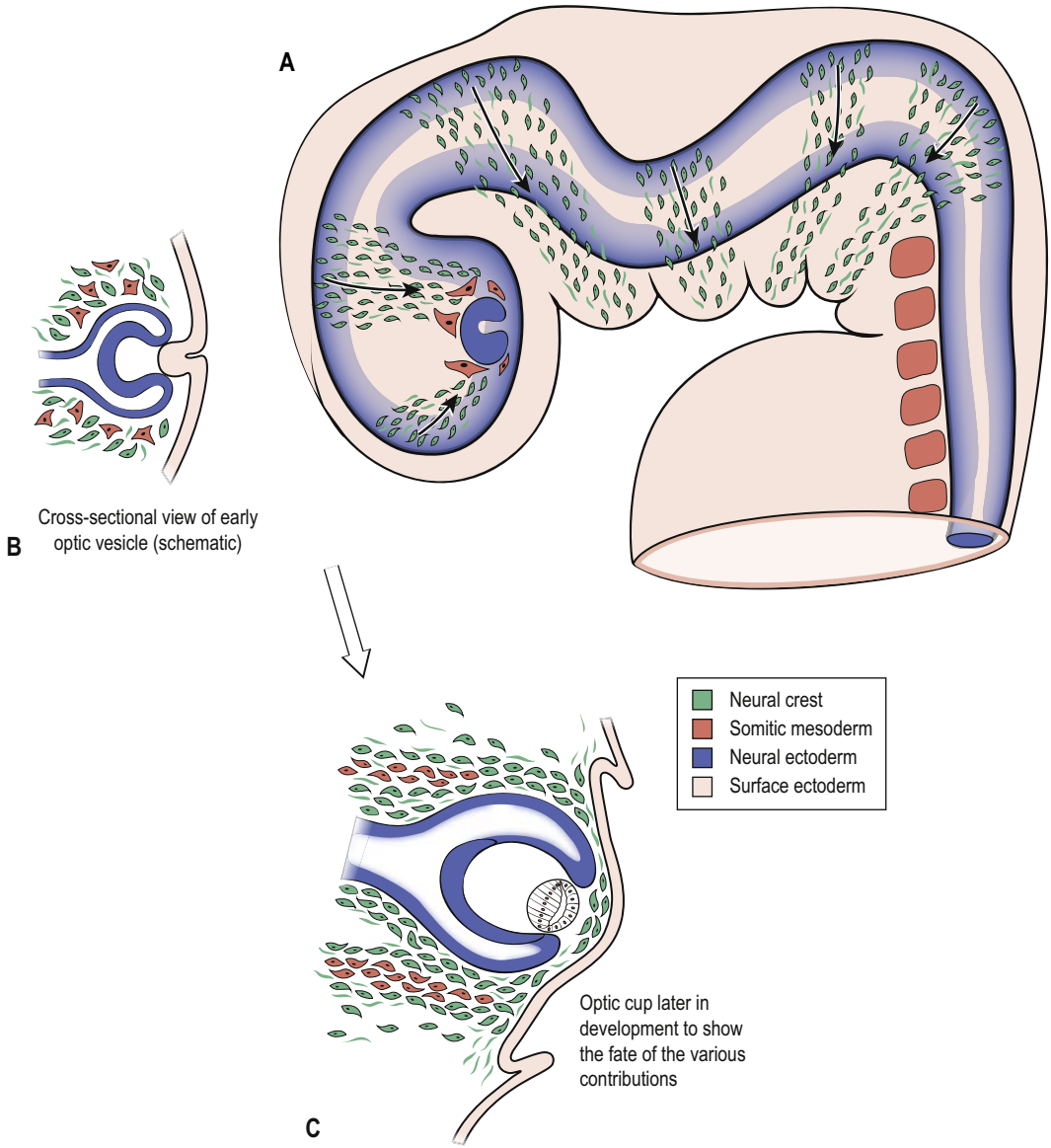


Fig. 2.5 Diagrammatic summary of the interactions between surface ectoderm, neural ectoderm, neural crest and mesoderm early in development and their final contributions to the eye and periocular tissues.

cup margin (Fig. 2.2F). By 14 weeks the retina terminates immediately posterior to the newly formed ciliary folds, with minimal pars plana. However, a definite pars plana and a poorly formed ora serrata are present by 6 months. The pars plana and the region from the ora serrata to the equator of the eye continue to grow after birth with continued growth of the eyeball, which occurs up to 2 years of age. The area of the retina is approximately 600 mm^2 at birth and reaches 800 mm^2 by 2 years.

Development of retinal vasculature

The vessel incorporated into the choroidal fissure is the *hyaloid artery*, a branch of the ophthalmic artery, itself a branch

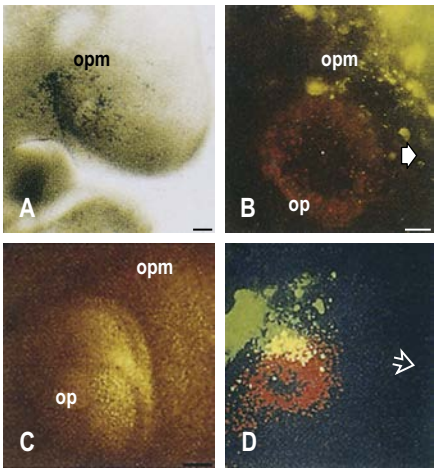


Fig. 2.6 Colonization of the periocular mesenchyme by cells derived from somitomere I in the mouse embryo. In (**A**) the cells have been labelled with X-gal *(blue)* and in (**B**) with DiO *(green fluorescence)*. In (**B**) the *red* cells are DiI-labelled neural crest cells. (**C**) A bright field view of the same specimen showing the orientation of the optic vesicle *(op)* and periocular mesenchyme *(opm)*. (**D**) A confocal image showing the co-distribution of the two cell populations showing that neural crest cells are found in both the neural epithelium of the optic vesicle and in the surrounding mesenchyme where they share territory with somitomeric-derived mesenchyme (*yellow* indicates areas of overlapping distribution but not double-labelled cells). *Arrow* points rostrally. Bar, 500 μm. (Reproduced from Trainor and Tam, 1995, with permission.)

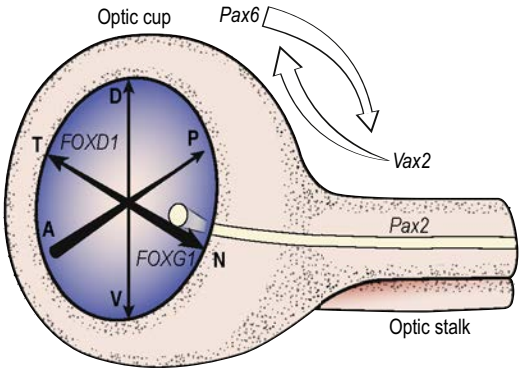


Fig. 2.7 Regulation of the axes in the optic cup and optic stalk by differential expression of transcription factors which determine the regional specificity or fate of the retina along dorsal-ventral *(D-V)*, naso-temporal *(N-T)* and antero-posterior *(A-P)* axes.

of the internal carotid artery (Fig. 2.11A). The hyaloid artery, upon emerging from the centre of the optic stalk, spreads between the lens surface and the marginal zone of the primitive neural retina (lentoretinal space). With growth of the optic cup and formation of the vitreous cavity, the hyaloid artery elongates and courses through the primitive vitreous, within the *hyaloid canal*, to reach the posterior lens surface.

Early in the fourth month of development, temporal clusters, or angiogenic buds, develop from the hyaloid vessels at the optic disc. These strands consist of endothelial cells, future glial cells and macrophages. The endothelial cells are canalized and form new vessels that course along the nerve fibre layer towards the peripheral retina at approximately 0.1 mm per day, to reach the ora serrata by the eighth month. At the same time, dividing vascular endothelial cells penetrate the depth of the neural retina to the outer border of the outer nuclear layer, a process not completed until the ninth month. Here they form a polygonal network of vessels, the outer retinal plexus. The intraneural portion of the hyaloid vessels becomes the *central retinal artery*. Developing capillaries are united by immature punctate tight and gap junctions and their basal laminae are incomplete. The pattern of retinal vessels growing during development on the retinal surface and then penetrating to form the intermediate and deeper layer (Fig. 2.9B) has been well studied in the mouse because of the popularity of this animal model in examination of the role of oxygen and vascular endothelial growth factor (VEGF) in normal retinal vessel formation during gestation (much of it postnatally in mice) and its relevance to retinopathy of prematurity. This condition, which is fortunately becoming less prevalent, occurs when the normal formation of retinal vasculature in the last trimester is disrupted if premature infants are placed in high oxygen environments to aid their breathing and increase their chances of survival. Removal of this supplemental oxygen can lead to abnormal angiogenesis which can result in blindness (Box 2.3 Clinical correlates).

Development of the retinal pigment epithelium

One of the most dramatic events in eye development is the appearance of melanin in the embryonic RPE, which occurs as early as 28 days after fertilization (6–7 mm embryo). The RPE arises from the proximal cells of the optic vesicle, which go on to form the outer layer of the optic cup (Fig. 2.1B,C). The presumptive RPE expresses the bHLH transcription factor *Mitf* (Fig. 2.4) and this appears to be dependent on canonical WNT signalling as evidenced in mice where inactivation of this signalling pathway causes the RPE to revert to neural retina. Initially, the RPE comprises a mitotically active pseudostratified columnar ciliated epithelium (Fig. 2.2B,C). The cilia disappear as melanogenesis

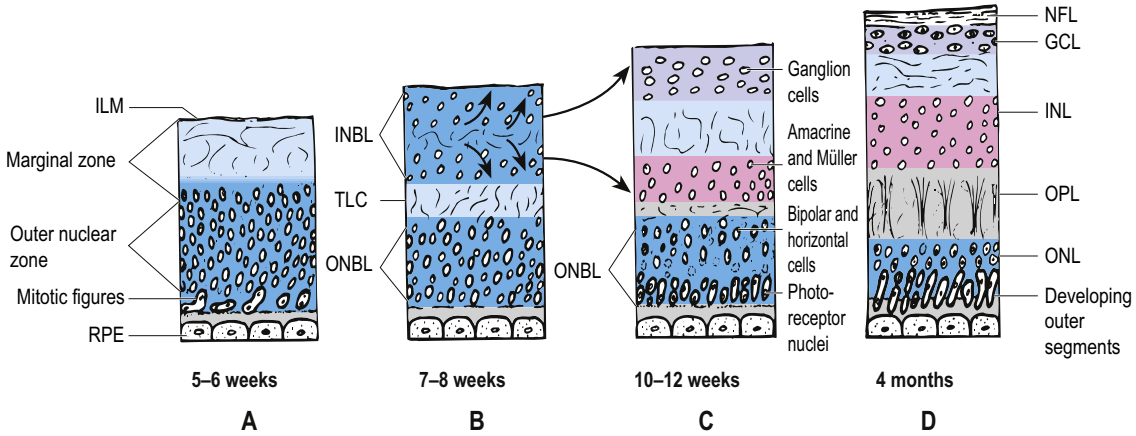


Fig. 2.8 (**A–D**) Summary of early retinal morphogenesis in the human eye. *Arrows* indicate major patterns of cell movements. *GCL*, ganglion cell layer; *ILM*, inner limiting membrane; *INBL*, inner neuroblastic layer; *INL*, inner nuclear layer; *NFL*, nerve fibre layer; *ONBL*, outer neuroblastic layer; *ONL*, outer nuclear layer; *OPL*, outer plexiform layer; *RPE*, retinal pigment epithelium; *TLC*, transient layer of Chievitz.

commences. The RPE cells become hexagonal in shape and homogeneous in size (Fig. 2.10A) and in section appear as simple cuboidal epithelium (Fig. 2.10B), although a columnar morphology is maintained in the peripheral retina for a longer period. By the fourth month the RPE has only minimal apical microvilli, few or no basal infoldings, primitive basolateral interdigitations (Fig. 2.10B), mature apical junction complexes and intracytoplasmic premelanosomes (Fig. 2.10C). Mitotic activity appears to take place early in development and is reputed to have ceased by birth; therefore, growth of the eye, and consequently of the RPE itself, is accommodated by hypertrophy or enlargement of the existing cells. A component of the future five-layered Bruch's membrane, the RPE basal lamina, is recognizable at the optic cup stage. Collagen fibrils are subsequently laid down beneath the basal lamina around 10 weeks; the first evidence of the elastic fibre layer can be detected around 3.5 months and by mid-term the elastic layer forms a fenestrated sheet.

Optic nerve and disc development

The hollow optic stalk forms a connection between the cavity of the forebrain (future third ventricle of the brain) and the cavity of the developing optic vesicles (Figs 2.1B and 2.2A). It is formed by the constriction of the proximal portion of the vesicle, particularly on the dorsal aspect concurrent with the expansion of the distal part (Fig. 2.11A). At this stage of morphogenesis (26–28 days) the central hollow fluid-filled stalk is lined by neuroectodermal cells. Invagination of the optic stalk at the choroidal fissure, on the ventral aspect, which occurs simultaneously with invagination of the optic vesicle, results in a double layer of neuroectoderm with narrowing and eventual obliteration of the intervening fluid-filled cavity (Fig. 2.11B). The invagination process in the distal and ventral portions of the stalk leads to the incorporation of the hyaloid vessels and surrounding mesenchyme (Fig. 2.11B). The lips of the optic stalk start closing over the hyaloid vessels near the forebrain (5–6 weeks) and gradually extend distally. This fusion lags behind that of the cup. The optic stalks lie at approximately 65 degrees to the mid-sagittal plane, compared with 40 degrees in the adult.

Axons from developing retinal ganglion cells grow towards the optic stalk and, upon reaching the optic disc, change direction and course towards the brain among the inner neuroectodermal cells of the developing optic nerve. The choroidal fissure closes soon after and by 6 weeks the optic nerve contains numerous axons that surround the hyaloid artery and vein. The outer neuroectodermal layer of the stalk differentiates into the peripheral glial mantle and the glial component of the lamina cribrosa. A cone-like structure, Bergmeister's papilla, consisting of glial cells and the remnants of hyaloid vessels, may persist at the optic nerve head in some individuals. An outer layer of condensed mesenchyme forms the optic nerve dura, which blends with the sclera. The glial septae surrounding the nerve bundles are composed of astroglia that differentiate from the cells of the inner layer of the optic stalk. The latter also gives origin to the oligodendroglia that surround the individual axons and are myelinated as far as the posterior margin of the lamina cribrosa. The nerve is displaced

nasally during the third month by enlargement of the temporal side of the eye. Abnormalities of optic nerve development are discussed in Box 2.4 Clinical correlates.

Development of the fibrous coat of the eye

Around weeks 6 and 7, periocular mesenchyme, derived from the neural crest and mesoderm (Fig. 2.5), begins to condense around the optic cup (Figs 2.1D,E and 2.2C,D). This mesenchyme can be differentiated into an *inner vascular layer* (uveocapillary lamina), which forms the stroma of the choroid, ciliary body and iris, and an *outer fibrous layer*, which will form the sclera and cornea.

Development of the sclera

Mesenchymal condensation is most conspicuous at the future site of insertions of the extraocular muscles (limbal–equatorial region) (Fig. 2.2D). In the third month, active fibroblasts are already embedded in an irregular matrix of collagen, elastic fibrils and glycosaminoglycans. By 12 weeks, a well-formed fibrous coat envelops the eye posteriorly as far as the optic nerve where the

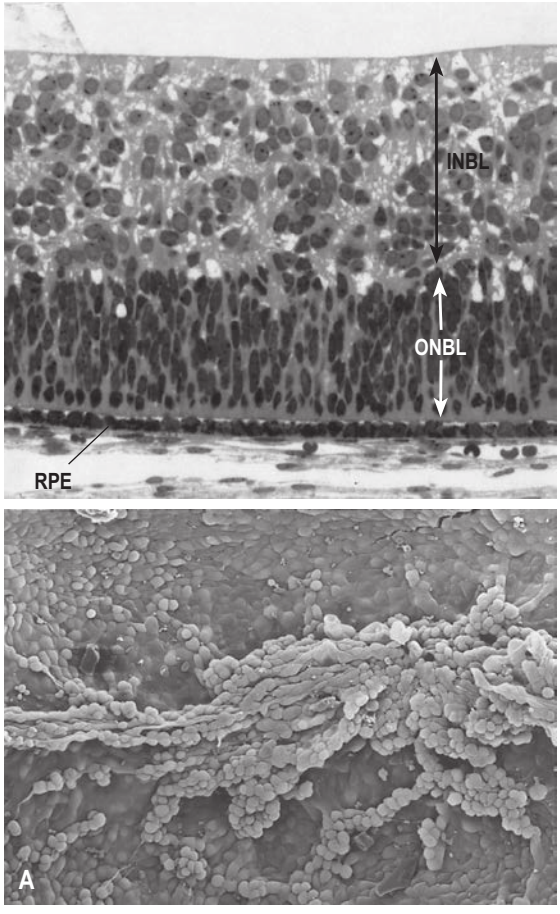


Fig. 2.9 (**A**) Human fetal retina (12–13 weeks) showing inner and outer neuroblastic layers *(INBL, ONBL).* The inner neuroblastic layer has commenced differentiation and the transient layer of Chievitz is obliterated. *RPE*, retinal pigment epithelium.

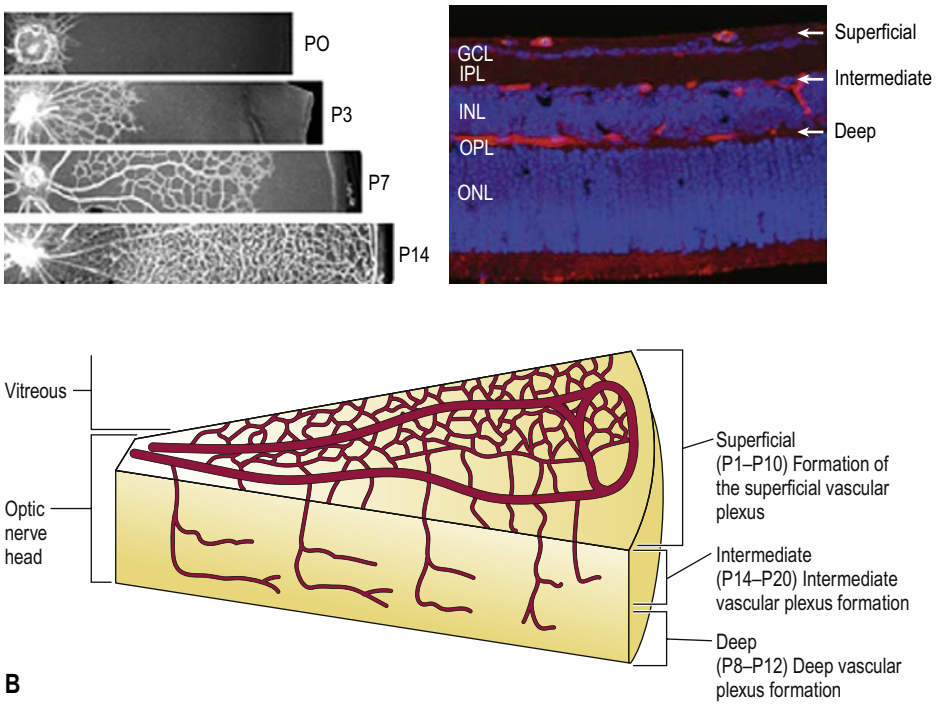


Fig. 2.9 cont'd (**B**) Diagrammatic summary of the retinal vascular development in mouse. Original magnifications: **A** × 115. Fig. 2.9B, after Dorrel and Freidlander, 2006.

connective tissue forms a perforated plate, the lamina cribrosa, through which glia-covered ganglion cell axons pass.

Development of the cornea

The *surface ectoderm* that seals over the lens pit forms the future corneal epithelium. It is a stratified squamous epithelium of three or four layers, the basal layer of which rests on a thin basal lamina. The first 'wave' of mesenchymal cells that passes over the optic cup margin migrates centripetally in the space between the anterior surface of the lens and the surface ectoderm to form the *corneal endothelium* (around 33 days) (Fig. 2.2C). Around day 49 a *second 'wave' of mesenchyme* commences migration from the optic cup margin and penetrates the space between the basal surface of the corneal epithelium and endothelium to form the corneal stroma (Fig. 2.2D). Both waves of mesenchyme are derived from the neural crest. The epithelium, which is continuous with the surface ectoderm, becomes stratified (three or four layers) and over the next few weeks the eyelids form and fuse (weeks 9–10) (Fig. 2.2D). Around 8 weeks the first evidence of loosely arranged collagen fibres can be detected amidst the actively synthetic fibroblasts, now known as *keratoblasts.* Within the corneal epithelium an intermediate layer of wing cells does not appear until the fourth or fifth month (Fig. 2.12A). The endothelium, which until now has been a double layer, becomes initially a simple cuboidal and ultimately a simple squamous layer resting on a thick basal lamina—the precursor of *Desçemet's membrane* (Fig. 2.12B). By this time all the corneal layers are present with the exception of *Bowman's membrane*, which becomes identifiable by 5 months as an acellular collagenous zone beneath the epithelium (Fig. 2.12A). The stromal collagen bundles become organized into highly oriented *lamellae* and the keratoblasts mature into long flattened *keratocytes.* This maturation process commences first in the posterior or deeper layers of the cornea and progresses more anteriorly or superficially.

Corneal thickness and diameter continue to increase throughout development by both interstitial growth (thickening of lamellae) and appositional growth (addition of new lamellae). The glycosaminoglycan constituents are known to alter during development and are thought to underlie the initial swelling of the cornea (when lids are fused) and its subsequent thinning, which occurs during eyelid opening (24 weeks). Corneal transparency is gradually attained before birth owing to maturation of the superficial lamellae and the hydration activity of the endothelial

Box 2.3 **Clinical correlates**

Retinopathy of prematurity

Altered oxygen tension has complex effects on the formation of capillary networks in the retina. Placing premature infants in high-oxygen environments (to aid in their respiration due to the immature state of their lungs) can result in a delay or reduction in retinal vascularization. Upon returning to normal oxygen tension, the retinal tissues experience a relative hypoxia which induces the release of angiogenic factors (VEGF), resulting in episodes of abnormal neovascularization within the retina and vitreous, known clinically as *retinopathy of prematurity* (ROP) or *retrolental fibroplasia.* Progress in modern neonatal care in many countries has made possible the survival of premature infants born at gestational ages and birth weights that were once considered to offer little hope of viability. The incidence of ROP increases in prematurity as both a function of reduced birth weight and lower gestational age from approximately 40% in infants with birth weights 1101–1200 g to more than 90% in those with birth weights 501–600 g (image courtesy Dr S. Tahija.)



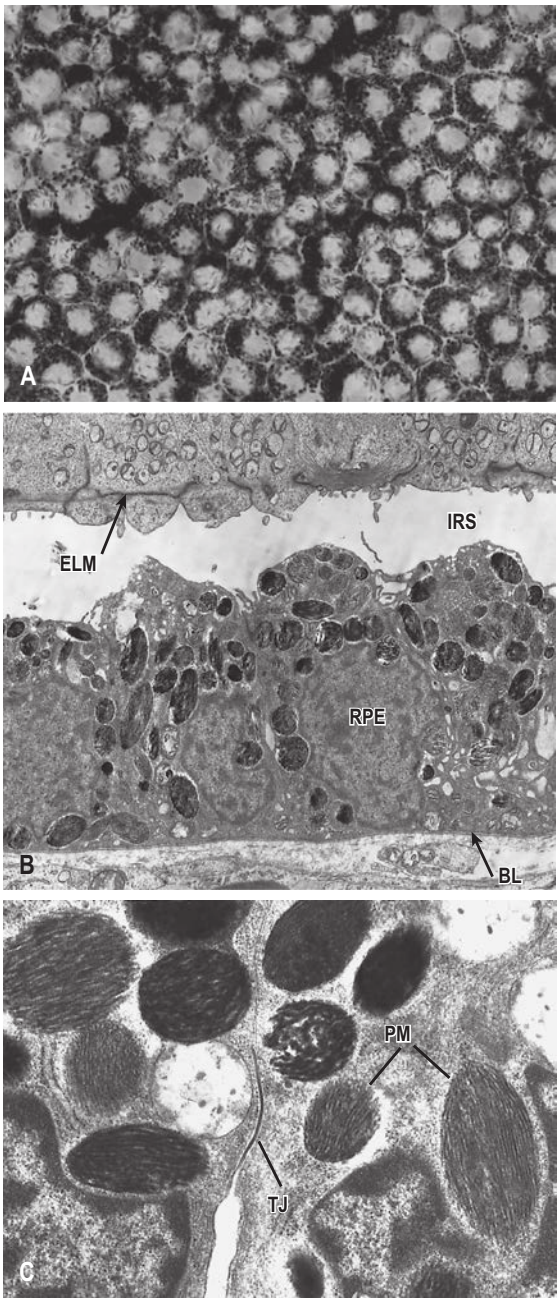


Fig. 2.10 (**A**) Human fetal RPE (15 weeks) viewed *en face* to demonstrate the regular hexagonal arrangement. (**B**) Transmission electron micrograph of human fetal RPE (12 weeks). (**C**) Premelanosomes (*PM*) and tight junctions (*TJ*) near apices of human fetal RPE (22 weeks). *BL*, basal lamina; *ELM*, external limiting membrane; *IRS*, intraretinal or subretinal space. Original magnifications: **A** × 100; **B** × 4100; **C** × 16 000.

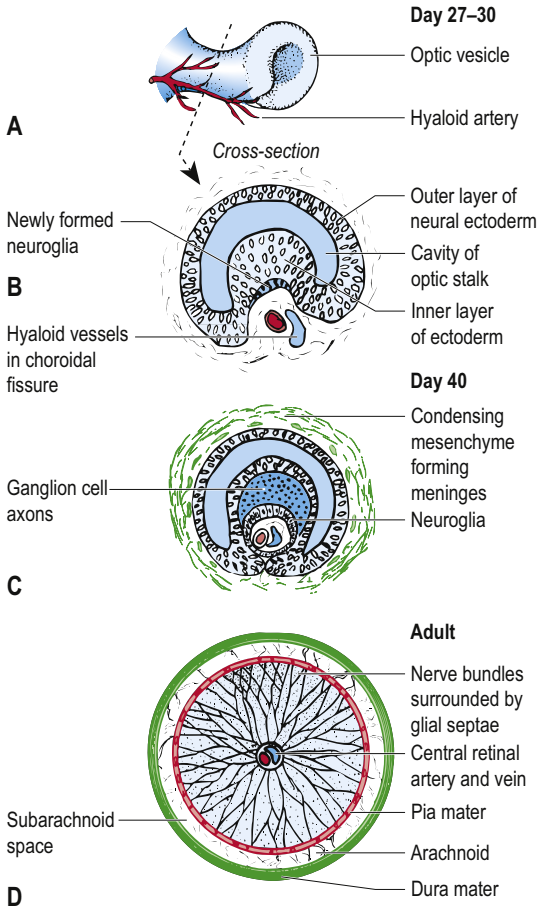


Fig. 2.11 Diagram summarizing the major events in the development of the optic stalk and optic nerve.

cells. Innervation of the cornea commences at 3 months and reaches the epithelium at 5 months.

In summary, the cornea develops from the interaction of a surface ectoderm-derived epithelium and neural crest-derived mesenchyme, which gives rise to the deeper layers including Bowman's layer, stroma, endothelium and its thick basal lamina, Desçemet's membrane. Malformations of the cornea that can occur in isolation as a feature of more generalized anterior segment abnormalities such as Peters' anomaly are discussed in Box 2.5 Clinical correlates.

Development of the intraocular contents

Lens development (Fig. 2.13)

The lens descends from a pool of a few hundred ectodermal cells that make up the thickened disc, the *lens placode*, identifiable at 27 days. Differential elongation of these cells and contraction of their apical terminal bar causes the placode to invaginate, producing a *lens vesicle* with a central depression, the *lens pit*, leading into a hollow *lens cavity* that is connected briefly to the amniotic cavity via the *lens pore* (Figs 2.1C and 2.2B). As the vesicle, surrounded by its basal lamina, detaches from the surface ectoderm (10 mm, 33 days) it sinks into the underlying rim of the optic cup (Figs 2.1D and 2.2B). Occasionally, degenerating cells (epitrichial or periderm cells) are seen within the lens cavity (Fig. 2.3). It now appears that lens induction and determination of cell fate of surface ectoderm is a multi-stage process requiring cellular and extracellular signalling by bone morphogenetic protein (Bmp) and also probably fibroblast growth factor (FGF) derived from the optic vesicle, which is regulated or inhibited by WNT signalling and retinoic acid signalling, combined with the expression of a number of local activating transcription factors, the most important of which appears to be *Pax-6* but also Six1 and Six2. More recently it has been shown that a disc-shaped 'pad' of modified extracellular matrix, consisting of fibronectin, produced by the lens placode, is critical to ensuring that the thickening of the surface ectoderm, destined to become the lens placode, is restricted to the area in close apposition to the optic vesicle and that disruption of this matrix leads to lateral spreading of the placode and failure to develop a lens placode—the so-called restricted expansion hypothesis.

Box 2.4 **Clinical correlates**

Malformations of the optic nerve head

When there is a failure of closure of the posterior part of the optic fissure, the optic nerve head is deformed by a *coloboma*, located inferonasally and associated with bulging of the sclera (scleractasia). The coloboma may take the form of a small recess (*optic pit*) at the rim of the disc where herniation of the retina occurs into the meninges and adjacent optic nerve. The clinical importance of an optic pit lies in its association with visual loss as a result of leakage from the pit and exudation of fluid beneath the macula.

Axial coloboma or 'morning glory syndrome'

There are numerous names for a symmetrically enlarged and excavated optic disc—a condition that may be unilateral or bilateral. The most extreme axial malformation is the 'morning glory syndrome', so-called because of the similarity to the American flower of the same name. This malformation is complicated by severe visual dysfunction and characterized by retrodisplacement of the optic disc into the meninges of the optic nerve. The abnormality is the result of a defect in mesodermal organization in the disc; the lamina cribrosa is not formed and there is fat and smooth muscle in the meninges.

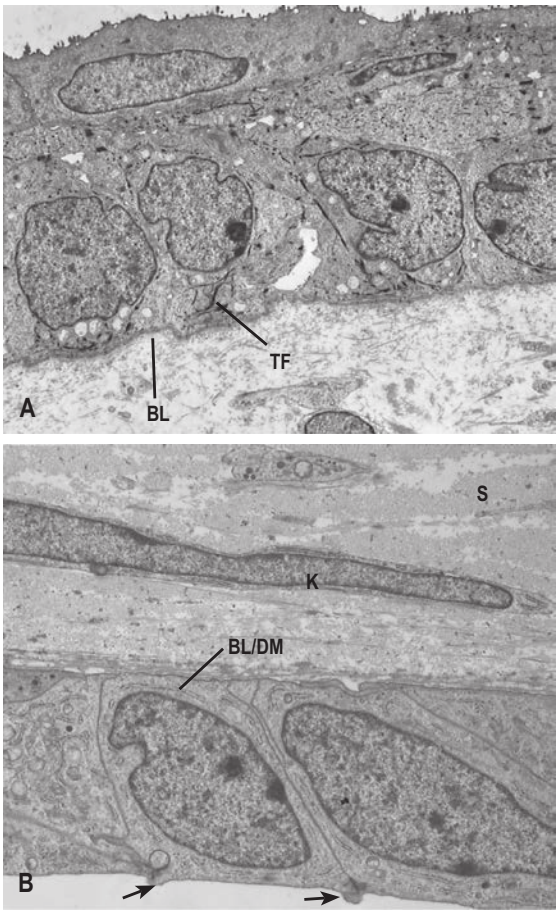


Fig. 2.12 Transmission electron micrographs of human fetal cornea (16 weeks). (**A**) Epithelium consists of three cell layers joined by desmosomes resting upon a thin basal lamina *(BL)*. Note the electron-dense intracytoplasmic tonofilaments *(TF)*. (**B**) Developing endothelial cells. Note that the cells are cuboidal with apical junctional complexes *(arrows)* and rest on the basal lamina *(BL/DM)*, which is already showing evidence of thickening to form the future Desçemet's membrane. *K*, keratoblast; *S*, collagenous stroma. Original magnifications: **A** × 2300; **B** × 5000.

The posterior cells of the lens vesicle elongate to form the *primary lens fibres* (Fig. 2.13A), and commence synthesis of a new group of intracytoplasmic proteins known as *crystallins* (see Chapter 4, page 236). In parallel they degrade their intracellular organelles to minimize light scatter. Fibroblast growth factor, as well as being an inductive signal, is likely to be involved in lens fibre differentiation. The base of each elongating lens cell remains anchored to the basal lamina posteriorly and their apices grow towards the *anterior lens epithelium*, thereby obliterating the lens cavity (Figs 2.1D,E, 2.2C and 2.13B,C). The nuclei migrate forward within the elongated cell body to produce a *lens bow, a row of cell nuclei* with a conspicuous forward convexity (Figs 2.2F and 2.13C,H). Subsequent lens fibres arise from mitotic activity within the *anterior lens epithelium* at the equatorial zone and are known as *secondary lens fibres* (Fig. 2.13D,H). The tips of the secondary fibres extend around the primary fibres and meet at the Y-shaped anterior and posterior *lens sutures* (Fig. 2.13D). Every subsequent generation of fibres throughout embryonic, and indeed later, life is added superficial to the previous layer. Early in embryonic development, the lens is nearly spherical or possibly longer in its antero-posterior axis; however, as secondary fibres are added at the equator, the lens becomes more ellipsoid, a trend that continues until birth and then into adulthood. Basal lamina material is continually deposited by the lens epithelium on its external aspect and encases the lens in a membranous non-cellular envelope, the *lens capsule.* Genetic abnormalities and environmental factors occurring during development can lead to disruption of lens organization or its growth and shape (Box 2.6 Clinical correlates).

During embryonic and fetal development, the lens receives nourishment via an intricate vascular net, the *tunica vasculosa lentis* (Fig. 2.13E–G), which completely encompasses the lens by approximately 9 weeks.

Box 2.5 **Clinical correlate**

Corneal and anterior segment malformations

Corneal opacification may be the result of a failure of the keratocytes to produce collagen fibres arranged in a lamellar array: instead, the pattern resembles sclera (scleralization of the cornea). *Peters' anomaly* is a term used to describe a posterior axial stromal defect associated with incarceration of the pupillary part of the iris at the edge of the defect. A mild form of malformation is the presence of thickening at the periphery of Desçemet's membrane (Schwalbe's line). When this is visible clinically (by gonioscopy), it takes the form of a bow: hence the term *embryotoxon* (Gk. *toxon:* a bow). Broad strands of tissue derived from the iris are sometimes seen in the chamber angle. In *Axenfeld's anomaly*, iridocorneal strands are localized to Schwalbe's line. When 'iris hypoplasia' is present, the malformation is known as *Rieger's (or Axenfeld–Rieger's) anomaly* (see Table 2.1 for gene mutations associated with these abnormalities).

Development of the vitreous and hyaloid system

At 5–5.5 weeks of gestation the lentoretinal space is narrow and occupied by the *primary vitreous* (Figs 2.1C,D, and 2.14A), which consists of the *vasa hyaloidea propria (VHP)*. Based on the elegant early work of Ida Mann, it

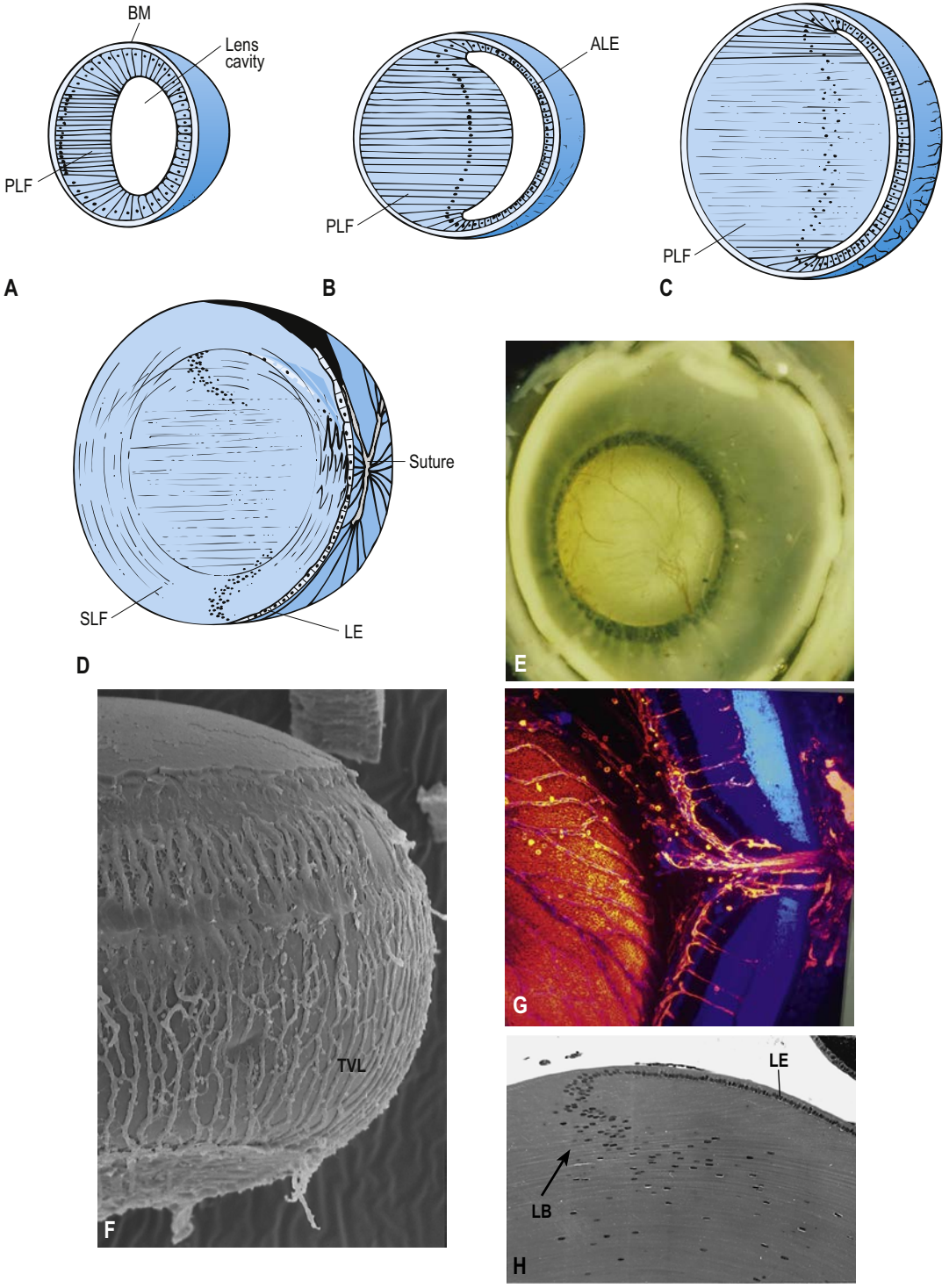


Fig. 2.13 Cont'd

Fig. 2.13 (**A–D**) Diagrammatic summary of lens morphogenesis. (**E**) Macroscopic view of human fetal eye (20 weeks) with posterior segment removed to reveal the hyaloid artery and tunica vasculosa lentis (TVL) around the lens. Original magnification: × 10. (**F**) Scanning electron micrographic view of a fetal rat lens surrounded by the fine vessels of the TVL. Note the small spherical macrophages associated with the vessels on the lens surface. Original magnification: × 95. (**G**) Frozen section of P0 mouse eye (day of birth) showing hyaloid artery emerging from the optic nerve and ramifying in the vitreous. The lens has been removed. The mouse is a Cx_3cr1-GFP transgenic mouse in which all myeloid-derived cells (monocytes, macrophages, microglia) are *fluorescent green*. Vessels are stained *red* and *blue* represent nuclei: × 30. (**H**) Fetal lens showing the lens bow *(LB)* arising from the equatorial region. Original magnification: × 90. *ALE*, anterior lens epithelium; *BM*, basement membrane of lens cells; *LE*, lens epithelium; *PLF*, primary lens fibres; *SLF*, secondary lens fibres. (Fig. 2.13G, Courtesy Dr Wai Wong.)

Box 2.6 **Clinical correlate**

Malformations of the lens

The lens is particularly susceptible, during its formation and early growth, to intrauterine toxic insults such as rubella (German measles). The disorganization that ensues causes degeneration in the fibres and visible opacities in the lens (congenital cataract). If the lens fibre cells recover, the opacities become buried in the inner part of the cortex by the newly formed lens fibres.

Congenital primary aphakia (CPA) (anterior segment dysgenesis 2; ASGD2) (absence of lens) is rare and associated with defects in FOXe3 gene. Small (microphakic) or round (spherophakic) lenses do not exhibit a strikingly abnormal histology. Axial bulges on the anterior and posterior surfaces of the lens (anterior and posterior lentiglobus) are presumed to be the result of abnormalities in the lens epithelium and the lens capsule.

was originally thought that the VHP was a branch of the hyaloid artery, which is classically described as becoming incorporated into the optic cup through the choroidal fissure and branching to form the VHP (Fig. 2.13G). Besides the branches in the vitreous, the hyaloid artery was also thought to terminate in the *tunica vasculosa lentis* (TVL) surrounding the posterior aspect of the developing lens. However, it now appears that VHP or fetal vitreous vasculature and the TVL form by haemovasculogenesis from the mesenchymal anlage that invades the open optic fissure. In this mesenchyme there are isolated islands of common precursor haemangioblasts which give rise to erythroblasts/erythrocytes in addition to endothelial cells and pericytes. Only later do these vasculogenic islands unite with the hyaloid artery when it forms at around 7 weeks of gestation.

The TVL also appears around 5 weeks of gestation as isolated islands of vasculogenic cells that eventually form a palisade-like network of vessels, the *capsulopupillary vessels*, which extend as far as the equator of the lens (Figs 2.13F and 2.14B). This capillary network eventually unites with the hyaloid vasculature posteriorly and anastomoses anteriorly with the *pupillary membrane* (*lamina iridopupillaris*), on the anterior lens surface (Fig. 2.14B). These pupillary membrane vessels are derived predominantly from branches of the long posterior ciliary arteries, which form an annular vessel close to the optic cup margin and whose branches pass over the rim of the optic cup to supply the anterior portion of the lens (Fig. 2.14B). The hyaloid system has no veins and venous drainage occurs anteriorly via the pupillary membrane and the uveal vessels. The expression of vascular endothelial growth factor by the lens is thought to be a critical molecular event in the formation of the hyaloid vascular system.

An avascular *secondary (definitive) vitreous* composed of finely fibrillar material is deposited behind the primary vitreous between 5.5 and 12 weeks. The primitive hyalocytes, widely recognized as belonging to the mononuclear phagocyte system, arise at this time and likely derive from the same population as the myeloid cells that invade the optic cup within the mesenchyme that gives rise to the hyaloid vessels and VHP and TVL (Fig. 2.13G). It is also likely that these cells give rise to the microglia of the retina and are all derived in embryonic life from yolk sac precursors, which is the source of all haematopoeitic cells prior to this role being later assumed first by the liver and then later still by the bone marrow. Mesenchymal cells in the adventitia of vitreal vessels likely contribute to the vitreous matrix. Much of the hyaluronan and collagen (type II) in the vitreous gel is added after birth.

Around the end of the third month, distinct condensations of secondary vitreous become evident in the space between the optic cup margin and the lens. The fibres, which are firmly attached to the inner limiting membrane of the retina in the developing pars plana region, are sometimes referred to as the *tertiary vitreous* and, when mature, will form the *vitreous base* (Fig. 2.14C). It seems likely that the non-pigmented ciliary epithelial cells in this region are responsible for synthesis of the tertiary vitreous and zonular fibres.

During the fourth month the remains of the primary vitreous, the hyaloid vessels and the VHP, together with the TVL, begin to atrophy (Fig. 2.14C). The regression occurs contemporaneously with the formation of the retinal vasculature. The remnant of the course these vessels took through the vitreous is evident in the adult as

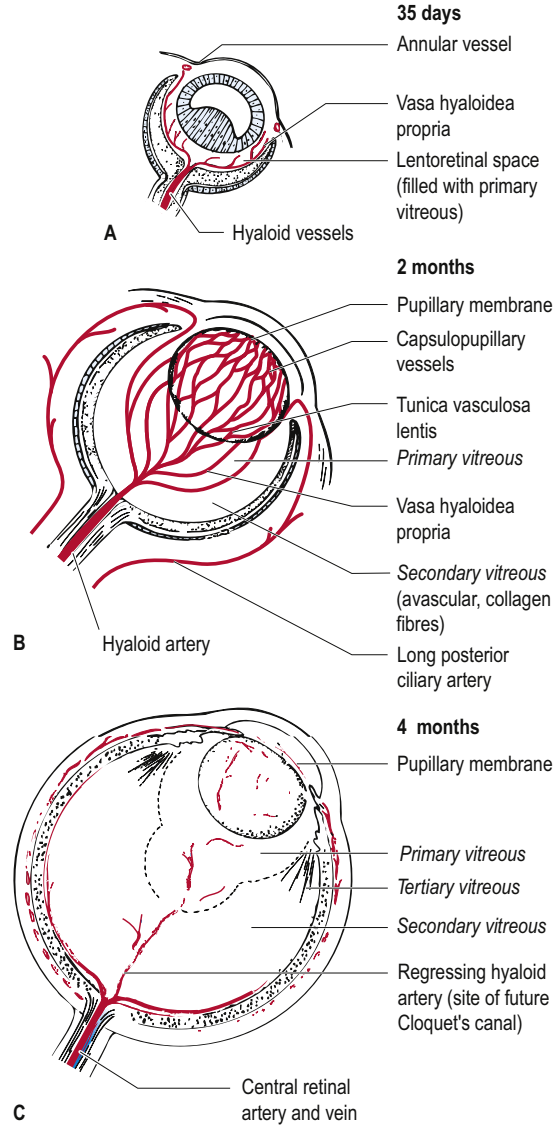


Fig. 2.14 Early development of the vitreous and hyaloid system.

a narrow fluid-filled central channel, Cloquet's canal. Macrophages play an important scavenging role in the regression of the hyaloid vessels (Fig. 2.13F,G) and may also be responsible for the induction of apoptosis of vascular endothelial cells. Small portions of the pupillary membrane may persist in otherwise normal newborn eyes. However, more severe disruptions to the normal regression of the vitreous and hyaloid vascular systems can occur and result in permanent blindness (Box 2.7 Clinical correlates).

Box 2.7 **Clinical correlates**

Malformations of the vitreous and hyaloid artery system

Persistent fetal vasculature

Normally, the hyaloid system of vessels vanishes completely but in some disorders, now grouped under the heading persistent fetal vasculature (PFV), regression does not occur.

Persistent tunica vasculosa lentis

Persistence of the anterior part of the tunica vasculosa lentis or the pupillary membrane causes deformation of the iris.

Persistent hyperplastic (anterior) primary vitreous

If the embryonic fibrovascular tissue in the anterior vitreous face persists, the ciliary processes are drawn internally, providing a valuable clinical diagnostic feature that is visible when the pupil is dilated. The lens is opaque in persistent hyperplastic anterior primary vitreous, because a retrolental fibrovascular mass erodes the posterior lens capsule and penetrates the lens cortex. The white retrolental mass (leucocoria) produced by this malformation can lead to a mistaken clinical diagnosis of retinoblastoma.

Persistent (posterior) hyperplastic primary vitreous

A persistent hyaloid artery and the condensed posterior primary vitreous (retrolental membrane) project from the optic disc and the adjacent retina. Distortion of the disc by prepapillary and preretinal fibrous membranes is associated with radial or falciform folds in the retina. Although many cases appear sporadic, there has been some evidence for mutations in *LRP5* and *FZD4* genes and also defects in apoptosis and WNT signalling indicating these processes may be critical to the normal regression of the hyaloid vascular system. In addition, mice lacking the Arf tumour suppressor protein (usually produced in response to sustained mitogenic activity and important in controlled apoptosis) fail to resorb the hyaloid system and this leads to persistence of pericyte-like cells that proliferate and destroy the lens and retina, causing blindness.

Development of the uveal tract

The choroid

The choroid arises very early in development from the loose vascular layer of mesenchyme that surrounds the optic cup (Fig. 2.1C,D). A palisade layer of vessels lies immediately external to the RPE and forms the basis of the future choriocapillaris. Fenestrations in the endothelium

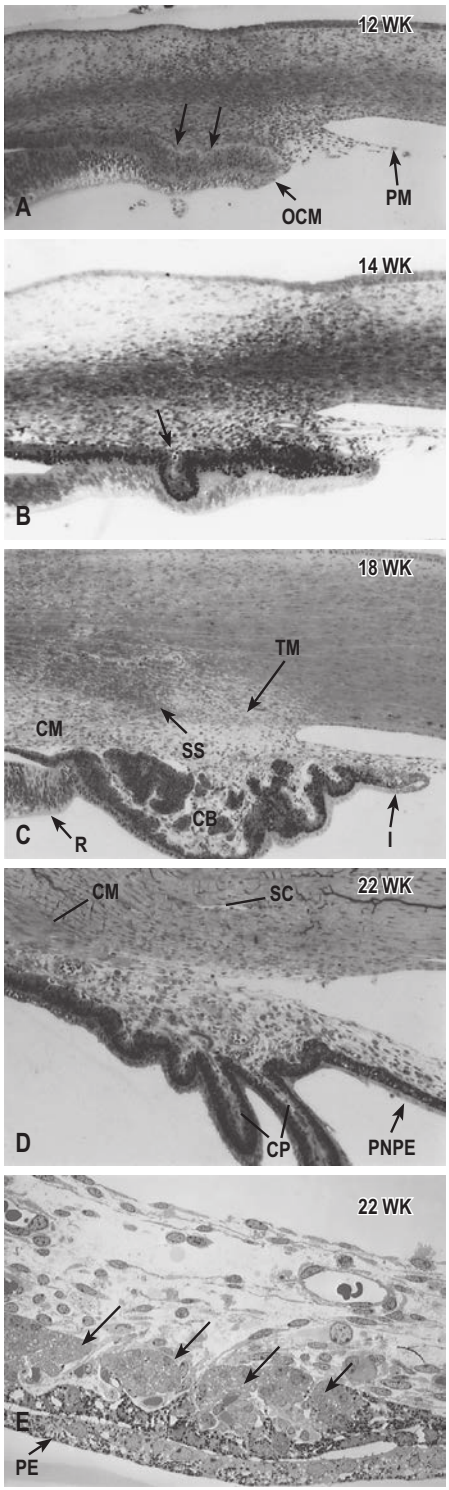


Fig. 2.15 Cont'd

Fig. 2.15 (**A–D**) Development of the ciliary body and iris in the human fetal eye from 12 to 22 weeks (the lens has been removed in all specimens). Note that the earliest evidence is the vascular mesenchyme indenting the outer neuroectoderm layer (*arrows*) near the optic cup margin *(OCM)* in 12- and 14-week specimens. (**E**) Electron micrograph of iris margin in a 22-week-old human fetal eye to illustrate sphincter pupillae smooth muscle bundles *(arrows)* differentiating from the anterior pigmented iris epithelium. Note that the posterior epithelium *(PE)* is showing early evidence of melanogenesis. Original magnifications: **A** × 75; **B** × 95; **C** × 110; **D**, × 80; **E** × 420. *CB*, ciliary body; *CM*, ciliary muscle; *CP*, ciliary processes; *I*, iris; *PM*, pupillary membrane; *PNPE*, posterior non-pigmented epithelium of the developing iris; *R*, retina; *SC*, Schlemm's canal; *SS*, scleral spur; *TM*, trabecular meshwork anlage.

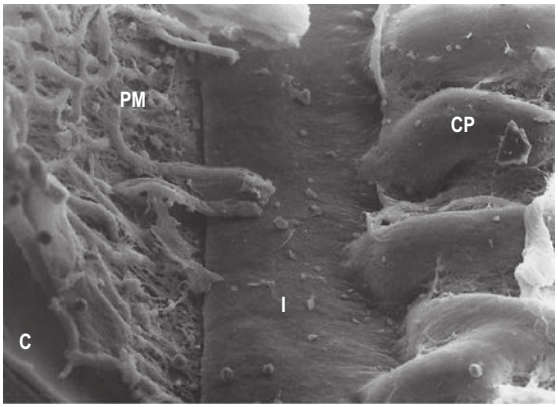


Fig. 2.16 Scanning electron micrograph of the developing human ciliary processes and iris (20-week fetus). Note the smooth outline of the ciliary processes *(CP)*, the short iris *(I)*, and vessels of the pupillary membrane *(PM)*. *C*, cornea. Original magnification: × 120.

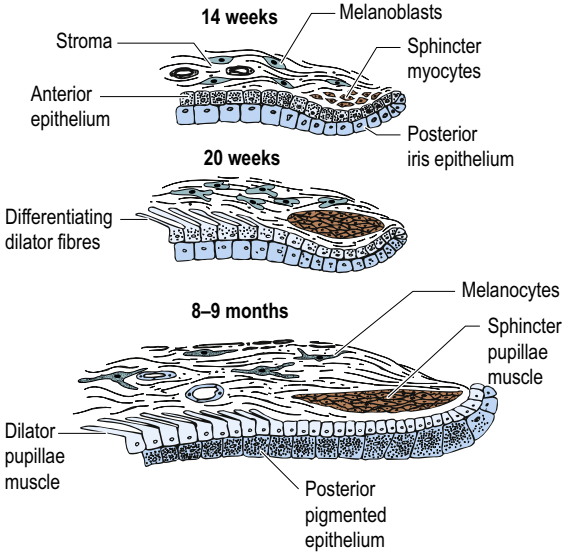


Fig. 2.17 Summary of human iris development showing the differentiation of the sphincter and dilator pupillae from the anterior iris epithelium (neuroectoderm) and the differentiation of neural crest cells into stromal fibroblasts and melanoblasts which later mature into melanocytes.

become evident very early in development. This layer of vessels forms communications with the precursors of the posterior ciliary arteries at around 2 months. A second layer of vessels forms at around 4 months external to the future choriocapillaris and consists of thin-walled venous channels that will eventually unite to form rudimentary vortex veins and branches of the long and short posterior ciliary arteries. An intermediate or middle layer (future Sattler's layer) of mainly arterioles forms between the larger vessels (Haller's layer) and the choriocapillaris. The choroidal vessels are initially embedded in a loose collagenous stroma; however, elastic fibres form later in development in the outer choroid (future lamina suprachoroidea). Studies in mice have shown that melanocytes arise from melanoblasts derived from neural crest cells which surround the optic cup early in embryological development. The appearance of discernible melanin granules within melanocytes around the seventh to eighth month of gestation perhaps misled early investigators into believing that melanocytes did not migrate into the eye till much later.

The ciliary body

The development of the ciliary body has similarities to iris development because it involves an interaction between mesenchyme and neuroectoderm. Ciliary body and iris development commence in the 11–12th week with indentation of the outer pigmented layer of the neuroectoderm (presumptive pigmented ciliary epithelium) near the optic cup rim by small capillaries in the inner vascular mesenchyme (Fig. 2.15A,B). Investigations of gene expression during early ocular development in the mouse revealed a high expression of *Pax6* and highly restricted expression of *Hox 7.1* in the non-pigmented or inner neural epithelium just behind the optic cup margin at the site of the prospective ciliary body and iris. In mice these genes are expressed 2 days before any morphological evidence of these structures is evident and thus it may be an early molecular marker for regional specialization and differentiation in the eye. Other transcription factors present in ciliary and iris epithelial progenitor cells include *Otx1* and *Msx1*, whose expression delineates the ciliary epithelial cells that do not form the neural retina. Of these, *Otx1* seems more critical as *Otx1* null mice lack a ciliary body and have a poorly developed

Box 2.8 **Clinical correlates**

Aniridia

Aniridia is a rare autosomal dominant bilateral disease in which there is an apparent absence of the iris. The term is a misnomer because histologically the abnormal iris is seen as a stump of hypercellular stroma, often with an abnormal proliferation of the pigment epithelium. Malformation or hypoplasia of the outflow system occurs in aniridia, as do anterior and posterior cortical lens opacities. The lens may dislocate (ectopia lentis) and the optic nerve may be hypoplastic. It is now well known that aniridia is caused by mutations of the *Pax-6* gene.

iris. In conjunction with transcription factor regulation, BMP (BMP4 and BMP7) and WNT signalling pathways also coordinate ciliary body and iris epithelium development.

Initially, the inner non-pigmented ciliary epithelium is flat, but as the vascular sprouts enlarge they push inwards to form primitive radial folds. This arrangement of a vascular connective tissue core overlaid by a double layer of ciliary epithelial cells forms the basis of adult ciliary process anatomy (see Chapter 1). The 70–75 radial folds that develop in this manner appear initially as smooth undulations (Figs 2.15B and 2.16); however, between weeks 14 and 22 they increase in height and complexity (Fig. 2.15C,D). Early in development the primitive neural retina terminates immediately posterior to the ciliary folds (Fig. 2.15C), but later a smooth area, the future *pars plana*, separates the two regions and continues to expand during the remainder of gestation with continued growth of the eye. The ciliary epithelium may commence aqueous production as early as 20 weeks, coinciding with concomitant changes in the iridocorneal angle.

The ciliary muscle differentiates at around 15 weeks' gestation from the mesenchyme between the neuroectoderm and the scleral condensation external to the early ciliary folds, namely the region that will form the future ciliary body stroma. The longitudinally oriented smooth muscle fibres terminate in the region of the trabecular anlage during fetal development. Circular or radial ciliary muscle fibres do not differentiate until much later in development and indeed are not fully formed until about 1 year of age.

The iris

Until 12–13 weeks of human development, there are no morphological signs of iris differentiation at the rim of the optic cup and the cup margin lies posterior to the lateral recess of the anterior chamber (Fig. 2.15A). At around 14 weeks, there is an expansion or growth of neuroepithelial cells at the cup margin anterior to the presumptive ciliary body (Fig. 2.15B,C). The optic cup neuroectoderm grows in a centripetal manner between the mesenchyme that has formed the cornea and the anterior lens surface. As it grows it incorporates some of the vessels of the lamina iridopupillaris or pupillary membrane, which lie on the anterior surface of the lens. This vascular mesenchyme is effectively split by the centrally growing neuroepithelial cells into vessels of the iridopupillary membrane, which now face the anterior chamber and forms the future iris stroma (Fig. 2.15C,D), and deep to the epithelium the vessels of the capsulopupillaris, which will regress later in development.

The smooth muscles of the iris, the *sphincter* and *dilator pupillae muscles*, are unique in embryological terms because they differentiate directly from neuroectoderm. The sphincter pupillae differentiation commences before that of the dilator pupillae. Around 13–14 weeks, anterior iris pigment epithelial cells delaminate, lose their melanin, develop intracytoplasmic microfilaments (actin) and dense bodies and deposit a basal lamina. These are classic characteristics of smooth muscle cells. Cell-to-cell contact (gap junctions) among the smooth muscle cells in this circumferential muscle band is not fully established until 7 months (Fig. 2.15E), and the muscle becomes free in the stroma at around 8 months. The dilator pupillae muscle develops much later, around 6 months, as basal extensions of the anterior or pigmented epithelial layer of the iris (Fig. 2.17), and continues to develop even after birth. These basal extensions are arranged radial to the pupil. This muscle never becomes fully independent of the epithelium because, even in adulthood, it is composed of modified basal processes of the neuroepithelial cells (Fig. 2.17).

During development, the posterior (inner) iris epithelium is largely amelanotic. It is continuous with the non-pigmented ciliary epithelium and thus the neural retina. Intracytoplasmic melanin increases in the fourth month, initially near the pupil margin (Fig. 2.15E), and by months 7–8 this layer is heavily pigmented (Fig. 2.17) and the anterior layer has lost its pigment.

Iris innervation, both adrenergic and cholinergic, is not established until late in development. In common with the choroid, pigment-bearing melanocytes are not identifiable in the iris stroma until late in development, around birth or later; however, studies in mice have revealed melanoblasts populating the iris primordium (Fig. 2.17). The thickness of the stroma and degree of melanogenesis are determining factors in eye colour at birth, and indeed full pigmentation and the pattern of the anterior surface are not complete until a few years postpartum. Abnormalities of iris development can occur in isolation or as part of a spectrum of disorders affecting the anterior portion of the eye (Boxes 2.1 and 2.8 Clinical correlates).

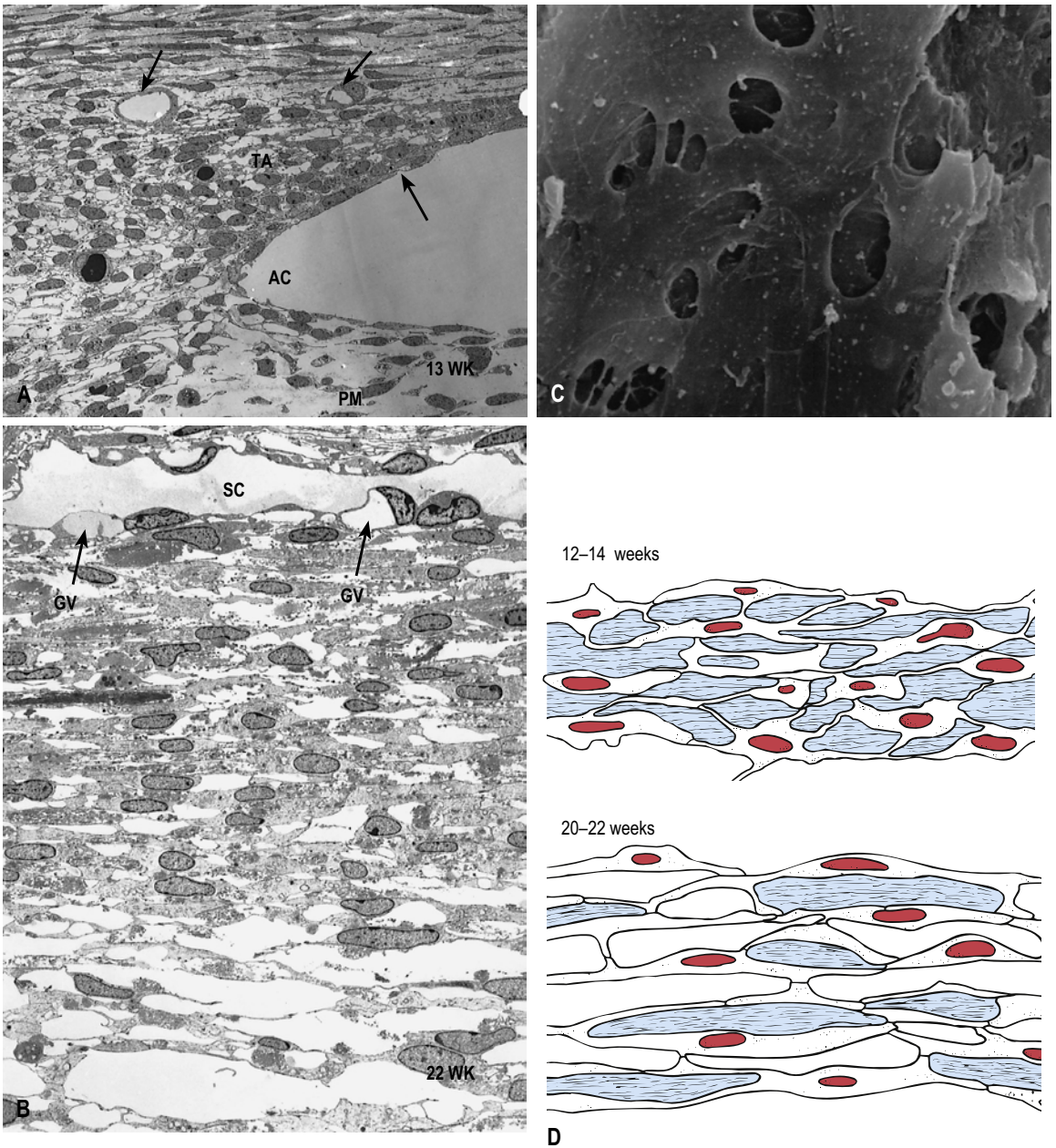


Fig. 2.18 Development of the iridocorneal or chamber angle (see also Fig. 2.16). (**A**) Electron micrograph of the trabecular anlage *(TA)* in a 13-week-old human fetus. Note the two small capillaries *(arrows)* on the deep aspect of the anlage, the high density of trabecular cells and the poorly developed extracellular matrix. *AC*, anterior chamber; *PM*, pupillary membrane (future iris stroma). Original magnification: × 500. (**B**) Electron micrograph of the trabecular meshwork in a 22-week-old human fetus. Note the enlarged intratrabecular spaces separated by well-formed connective tissue trabeculae and the size of Schlemm's canal *(SC)*, which possesses giant vacuoles *(GV)* in its inner wall. Original magnification: × 100. (**C**) Scanning electron micrograph of the inner surface of the trabecular anlage in a 13-week-old human fetus revealing the incomplete nature of the endothelial cells facing the anterior chamber. These perforations most likely allow free passage of cells and fluids from the anterior chamber to the developing meshwork from this early stage onwards. *Large arrow* in (**A**) indicates the perspective from which the scanning micrograph was obtained. Original magnification: × 1600. (**D**) Summary of the morphogenetic changes that occur during remodelling of the loose mesenchyme of the trabecular anlage to form the trabecular meshwork. (Fig. 2.18A–C from McMenamin, 1989, 1991; with permission.)

Development of the anterior chamber angle and aqueous outflow pathways

As early as the 12th week, a roughly wedged-shaped distinctive mass of mesenchyme, the trabecular anlage, can be identified at the junction of the pupillary membrane and lateral margins of the cornea, namely the future anterior chamber angle (Figs 2.15A and 2.18). The trabecular anlage consists of a dense collection of stellate mesenchymal cells (neural crest-derived) and some loosely arranged extracellular matrix (Fig. 2.18A,D). The deep aspect of the wedge-shaped anlage is characterized by a row of small capillaries (Fig. 2.18A), which most probably have grown in from the capillary plexus on the external surface of the eye (future episcleral plexus) and are thus lined by mesoderm-derived vascular endothelial cells. By weeks 20–22 (fifth month) the connective tissue matrix of the trabecular anlage consists of flattened trabecular endothelial-lined sheets and cords (early trabeculae) separated by intervening spaces (Fig. 2.18). On the deep aspect of the fetal trabecular meshwork the collection of small capillaries fuses to form a single elongated slit-like vessel, the *canal of Schlemm*, lined by endothelial cells that are continuous with those of the collector channels and episcleral veins. The characteristic 'giant vacuoles' in the canal endothelium that are responsible for the passage of aqueous across the inner wall of the canal (see Chapter 1) appear at around 18–20 weeks of gestation. Recent research has found that Schlemm's canal represents a new intermediate vessel type, which expresses lymphatic endothelial cell markers Prox1, VEGFR3, and integrin α9, but not LYVE1 and podoplanin, which are often used to identify lymphatic vessels in immunohistochemical studies. The primitive Schlemm's canal is formed by vascular endothelial cells that are postnatally respecified to acquire lymphatic phenotypes through up-regulation of Prox1. Thus, although they originate from Tie2+ venous endothelial cell progenitors, they assume molecular characteristics in part typical of a lymph vessel. *Tie2* mutations have been identified in patients with primary congenital glaucoma. Nevertheless, although the incidence rate of primary congenital glaucoma is low, primary open-angle glaucoma is frequently observed in the elderly. A recent study showed that double *Angpt1*/*Angpt2* deletions or *Tie2* deletion in adult mice alters Schlemm's canal integrity and function leading to raised intraocular pressure and impaired retinal ganglion cell function, features of glaucoma.

During the remainder of fetal development, the trabecular meshwork becomes further specialized into cord-like inner uveal trabeculae, numerous intermediate layers of lamellar corneoscleral trabeculae and a deep loosely arranged cribriform meshwork (Fig. 2.18B). The scleral spur is formed by month 4–5 (Fig. 2.15C). Scanning electron microscopy of the developing angle reveals that the inner aspect of the developing uveal meshwork is incomplete and numerous perforations allow communication between the anterior chamber and the spaces of the developing meshwork from 15 weeks onwards (Fig. 2.18C). Abnormalities in the development of the outflow system either in isolation or as part of a spectrum of disorders can lead to a failure of normal aqueous drainage and congenital glaucoma (Box 2.9 Clinical correlates).

Box 2.9 **Congenital glaucoma**

Although there have been many theories on the aetiology of congenital glaucoma, it appears to be a failure of differentiation or alterations in differential growth rates. This is a strong candidate because the few pathological accounts of congenital glaucoma describe the tissue as undifferentiated or lacking the typical organized trabeculae and intertrabecular spaces, especially in the outer or cribriform zone. Mutations in *CYP1B1*, *LTBP2* and *MYOC* have been found in a cohort of primary congenital glaucoma patients and the *CYP1B1* gene, which encodes cytochrome P450 1B1, a member of the cytochrome P450 superfamily of enzymes. It may metabolize a signalling molecule involved in eye development, possibly a steroid; hence, mutations may lead to congenital glaucoma.

Chamber angle malformation (goniodysgenesis)

A failure of remodelling of the progenitor tissue in the angle leaves very obvious strands between the iris stroma and the trabecular meshwork, or between the iris and cornea. Abnormalities in the outflow system, goniodysgenesis, include hypoplasia of the scleral spur with extension of the ciliary muscle into the outer part of the trabecular meshwork and an excessive amount of trabecular tissue. In addition, there are a number of more generalized abnormalities of the mesenchyme and mesoderm, which display altered iridocorneal angle structure and may manifest as infantile glaucoma. These include posterior embryotoxon, Axenfeld syndrome, Rieger's anomaly and Peters' anomaly. A number of mutations or deletions have been associated with iridogoniodysgenesis (Table 2.1).

Development of extraocular muscles

The extraocular muscles are some of the few periocular tissues that have been shown not to be of neural crest origin (Fig. 2.6). They are thought to arise from presumptive

myocytes in the *preotic region (paraxial mesoderm)* in the area of the prochordal plate.

They migrate ventrally and caudally around the developing eye. The presumptive myocytes concentrate particularly in the equatorial zone external to the mesenchymal condensation, which forms the sclera. Here they proliferate and differentiate. Their flattened connective tissue tendons, which are of neural crest origin, eventually fuse with the sclera.

Investigations of *MyoD* gene expression in transgenic mice (using the *LacZ* reporter gene) have shown evidence of myogenesis *in situ* around the developing eye as early as E10.5 (day 10.5 of embryogenesis). The extraocular muscles appear in approximately the following sequence: lateral rectus, superior rectus and levator palpebrae superioris (week 5), superior oblique and medial rectus (week 6), followed by inferior oblique and inferior rectus (common primordium).

Congenital Cranial Dysinnervation Disorders (CCDDs) are a group of congenital, non-progressive ophthalmoplegias with restriction of globe movement in one or more fields of gaze. This group of strabismus syndromes was assumed to be a primary pathologic process starting in the extraocular muscles. However, recent evidence suggests that the primary pathologic process is of neuropathic origin. It was believed that for normal development of extraocular muscles normal development innervation to these muscles was essential and that congenital dysinnervation to these extraocular muscles led to abnormal muscle structure depending upon the stage and the extent of such innervational defects. However, mouse studies in which extraocular muscles are absent have shown that extraocular muscles are not necessary for initial outgrowth and guidance of ocular motor axons from the brainstem to the orbit but are required for their terminal branching and survival. Thus, intermediate targets in the mesenchyme migrating around the developing eye provide cues for specific targeting of ocular motor axons to the orbit, while the cues from the developing muscles themselves are responsible for the appropriate terminal branching and motor neuron survival.

Development of the eyes and surrounding structures is influenced by the pattern of development of the skull, pharyngeal arches and face. A detailed account of the embryology of the head, skull and face pertinent to understanding the development of the eye and orbit can be found at the following eLink:

Further details on development of the head and neck with special reference to the face and orbit are available online at https://expertconsult.inkling.com

A full reading list is available online at https://expertconsult.inkling.com/.

eBox 2.3 Craniofacial abnormalities

Skeletal development can be considered to have three essential steps: (1) classic induction as the result of mesenchyme–epithelium interactions mediated by highly conserved signalling molecules such as Bmp (bone morphogenetic protein) and FGF (fibroblast growth factor); (2) condensation of mesenchyme; and (3) overt differentiation. Interestingly, 65% of skeletal abnormalities of the head and neck are the result of defects in this first signalling step (Hall, 1998).

It is now recognized that a group of craniofacial abnormalities, the mandibulofacial dysostoses, including Treacher Collins and Hallerman–Streiff syndromes, are caused by deficits in neural crest cell migration and differentiation in the first and second pharyngeal arches. These are manifest as abnormal ear development, hypoplasia of the maxilla and mandible, and lower lid defects. These and more generalized disturbances of neural crest migration, e.g. Rieger syndrome, Pierre Robin syndrome, and conditions affecting primarily the periocular region such as Peters' anomaly, are increasingly being classified as *neurocristopathies* because of their proposed link to disturbances in neural crest cell migration, proliferation and differentiation.

eBox 2.4 Developmental anomalies of the nasolacrimal duct

Disturbances in morphogenetic processes may lead to multiple canaliculi and punctae, abnormal diverticulae and blockage of the nasolacrimal duct, possibly because of debris from the degenerating central cells producing a mucocoele (not uncommon in the first few weeks after birth).

There is a range of *congenital facial defects*, of which cleft lip and palate are the most common. These are a consequence of complete or partial failure of fusion of the various processes and swellings in the face. Oblique facial clefts occur along the course of the nasolacrimal duct (eFig. 2.5E). The aetiology of these conditions is multifactorial, although maternal exposure to anticonvulsant drugs (phenytoin) and vitamin A are known teratogenic agents that produce these defects. The spatial and temporal expression of retinoic acid receptors during embryological development is currently the subject of much research. While details are still emerging, it is clear that the patterns of expression partially explain the effects of retinoic acid (a vitamin A derivative) as a potent morphogen. Retinoic acid appears sequentially to activate genes of the *Hox* cluster, thereby influencing the normal segmental pattern of hindbrain development (rhombomeres), neural crest migration pathways into pharyngeal arches, and development of the face (Maden et al., 1992).

eBox 2.5 Trisomy 13 (Patau syndrome)

The ocular pathology in trisomy 13 illustrates various forms of malformation. The cornea and chamber angle are malformed and persistent hyperplastic primary vitreous is common. An anterior coloboma is present and is characterized by a fibrous ingrowth that contains nodules of cartilage. Retinal dysplasia is extensive. Optic nerve malformation is limited to hypoplasia. The systemic malformations are not compatible with survival and are extreme forms of brain malformation (arrhinencephaly) with cardiac and renal malformation.

eBox 2.6 Trisomy 21 (Down syndrome)

The systemic disturbances in this disorder are well known. In ophthalmology the important components are a high incidence of keratoconus and cataract. Small nodules are formed by spindle cells on the iris (Brushfield's spots); myopia and the attendant complication of retinal detachment may require surgical intervention.

Development of the skull

The orbit lies at the junction of the neurocranium (skull vault) and viscerocranium (facial skeleton) and has also evolved partly from the primitive sensory capsules around the eyes. Therefore, to understand the embryology and development of the orbit, it is essential to appreciate that the entire craniofacial skeleton is formed by a combination of several components.

Chondrocranium

The *chondrocranium* forms the skull base initially as cartilaginous precursors that develop in a rostral to caudal sequence, namely the prechordal plate, hypophyseal and parachordal cartilages. These subsequently ossify to form the midline bones in the base of the skull from the interorbital region (body of the sphenoid) to the occipital region. The *sensory capsules* that evolved to support the olfactory organs, eyes and inner ear develop separately alongside this midline basal cartilage. These capsules develop initially as cartilage and in humans are represented by bones in the nasal cavity (ethmoid from prechordal cartilages), orbits (body of the sphenoid—hypophyseal cartilages, lesser wing and medial part of the greater wing of the sphenoid) and part of the temporal bone.

Membranous bones

The *membranous bones*, evolved from dermal bones such as those seen in fossils of primitive placoderm fishes, ossify directly from mesenchyme (in membrane). They are derived from neural crest and parachordal (head) mesoderm and form the calvaria or cranial vault in the human skull. They are represented in the orbit by the orbital plate of the frontal bone.

Viscerocranium

The *viscerocranium* evolved to support the *branchial (gill) arches* in fish. In humans there are five pairs of *pharyngeal arches*, either side of the foregut tube, whose derivatives contribute to the viscerocranium (see eBox 2.2).

The pharyngeal arches

The pharyngeal arches play an important role in the morphogenesis of the head and neck; the first and second arches in particular are important to the development of the periocular region. Pharyngeal arches form in a craniocaudal sequence and are not all at similar stages of development at any point in time. In humans, five arches develop that correspond to arches 1, 2, 3, 4 and 6 of their evolutionary precursors in fish. The first or mandibular arch has two components, one forming the upper jaw the other the lower jaw; the second or hyoid arch has evolved to support the jaw, tongue and larynx (eFig. 2.3). The third arch also contributes to the hyoid. The fourth and sixth arches help form the larynx. Each arch has an inner covering of endoderm (separated by endodermal *pharyngeal pouches*) and an outer covering of ectoderm (separated by ectodermal *pharyngeal clefts*), a cartilaginous component (e.g. Meckel's cartilage), an arch nerve and an arch artery, together with a core of mesenchyme (eFig. 2.4). This mesenchyme is a mixture of somatic mesoderm and neural crest-derived mesectoderm: derived from rhombomere 2 (first arch), rhombomere 4 (second arch) and rhombomere 6 (third arch) (Hunt et al., 1991; Noden, 1991; Maden et al., 1992; Hall, 2005). eTable 2.1 and eFig. 2.3 summarize the skeletal, neuronal and muscular derivatives from the pharyngeal arches. The migration of neural crest cells into each of the pharyngeal arches is like all axial organization (i.e. neural tube, somitic mesoderm and endoderm) controlled by the *Hox* code. Only a few elements of the viscerocranium, such as the medial part of the greater wing of the sphenoid (alisphenoid), incus, malleus and stapes, arise directly from the cartilaginous component (eFig. 2.3A); the majority of the cartilaginous elements of the pharyngeal arches regress and become encased within membrane bones that ossify directly from neural crest-derived mesenchyme. Such membranous bones include the maxilla, zygoma, squamous temporal bone and dentary (described in most textbooks as the mandible) (eFig. 2.3A).

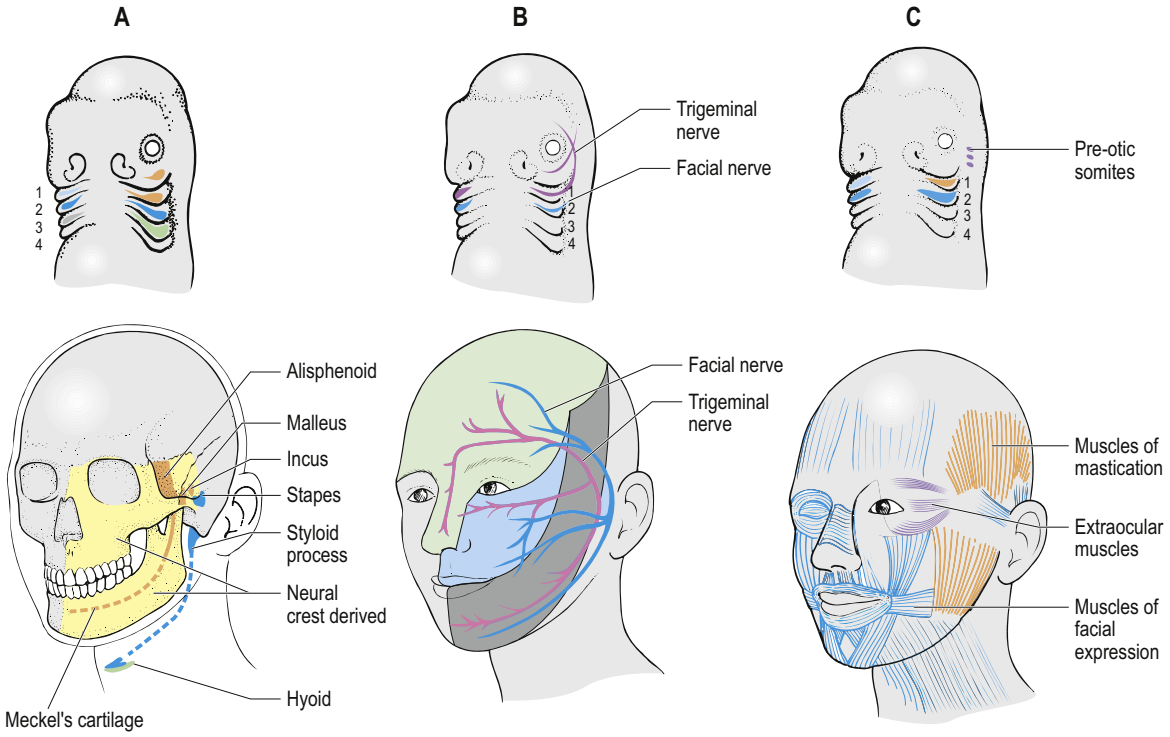
Some of the cranial nerve sensory neurones (part of V, VII, IX and X), such as the dorsal root ganglia of the trunk, arise from the neural crest, as do the four parasympathetic ganglia of the head and neck.

Development of the face

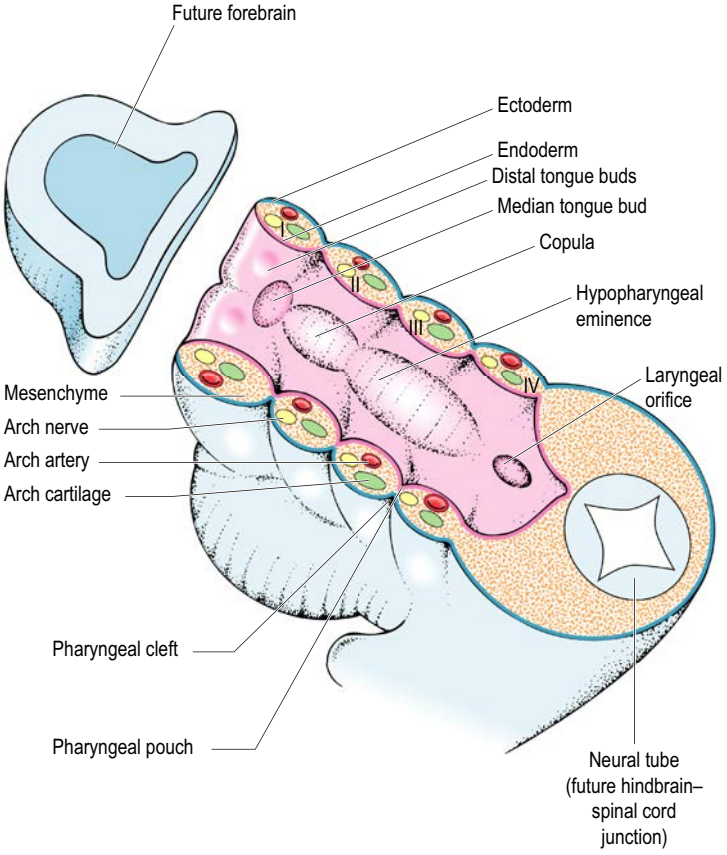
(eFig. 2.5)

Development of the face commences around the fourth week and is largely complete by week 10. The basic arrangement of the face is the result of fusion of five swellings around the stomodeum (primitive mouth), nasal pits and eyes. There are paired maxillary and mandibular processes plus an unpaired fronto-nasal process (eFig. 2.5A). The mesenchyme of the fronto-nasal process, a conspicuous swelling over the developing forebrain vesicles, arises from the neural crest and does not appear to have any association with pharyngeal arch development. The fronto-nasal process is innervated by the ophthalmic division of the trigeminal nerve and forms the tissues above the eye as far back as the vertex, thus explaining the sensory cutaneous distribution in the adult. The fronto-nasal process also gives rise to medial and lateral nasal swellings around the nasal pits (eFig. 2.5B). The maxillary processes grow medially beneath the developing eyes and form the lower eyelids. Concomitant with this growth, the eyes also move from their position on the lateral aspect of the embryonic head to the front of the face. The *nasolacrimal groove* on each side lies along the line of fusion of the maxillary process with the lateral nasal swelling (eFig. 2.5C). The ectoderm of the groove invaginates (10 mm stage)

Continued...



eFig. 2.3 Summary of the embryonic pharyngeal arches and their derivatives in the adult. (**A**) Arch cartilages (*yellow* shaded areas represent bones derived from the neural crest); (**B**) arch nerves and the sensory territories of the three divisions of the trigeminal nerve shaded *green* (V_1), *blue* (B_2) and *grey* (V_3) (see also Fig. 1.59); and (**C**) arch musculature.

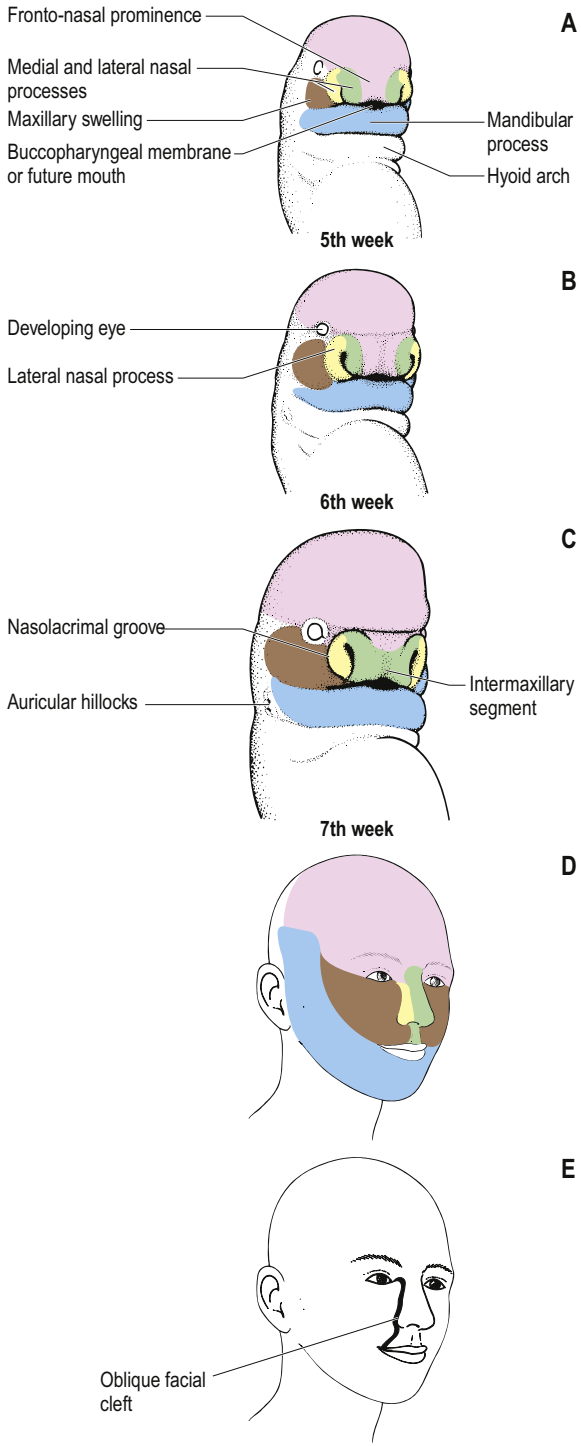


eFig. 2.4 Schematic diagram of the floor of the embryonic mouth as viewed from above in a horizontal section of the head. Note the pharyngeal arches (I–IV) have been cut to expose their neural *(yellow)*, arterial *(red)*, cartilaginous *(green)* and mesenchymal *(orange stipple)* contents. The arches are lined externally by ectoderm *(blue)* and internally by endoderm *(purple)*.

eTable 2.1 **Pharyngeal arch derivatives (arch arteries and their derivatives not included)**

Pharyngeal arch	Pouch endoderm	SKELETAL ELEMENTS: Ossify in cartilage	SKELETAL ELEMENTS: Ossify in membrane	Cranial nerve	Muscles
1	Tympanic cavity Pharyngotympanic tube (*cleft ectoderm—external auditory meatus*)	Arch cartilage—palatoquadrate bar: alisphenoid, incus Meckel's cartilage—malleus	Arch mesenchyme: *upper part* maxilla, zygoma, squamous temporal bone *lower part* mandible	Maxillary division of trigeminal (V) (upper part of arch 1) Mandibular division of trigeminal (V) (lower part of arch 1)	*From cranial somitomere 4*: Muscles of mastication (temporalis, masseter, pterygoids) *plus* mylohyoid, anterior belly of digastric, tensor tympani, tensor veli, palatini
2	Epithelial lining of tonsillar crypts and tonsillar fossa	Reichert's cartilage: stapes, stylohyoid ligament, upper part of hyoid		Facial nerve (VII)	*From somitomere 6*: Muscles of facial expression *plus* posterior belly of digastric, stylohyoid, stapedius
3	Dorsal wing: inferior parathyroid Ventral wing: epithelioid cells of thymus (Hassall's corpuscles and epithelial reticulum)	Lower part of hyoid		Glossopharyngeal nerve (IX)	*From somitomere 7*: Stylopharyngeus
4	Dorsal wing: superior parathyroid Ventral wing: ultimobranchial body: C cells in thyroid gland		Upper laryngeal cartilages (mesoderm)	Superior laryngeal branch of vagus nerve (X)	*From occipital somites 2–4*: Pharyngeal constrictors, cricothyroid, levator veli palatini
6			Lower laryngeal cartilages (mesoderm)	Recurrent laryngeal nerve (X)	Intrinsic muscles of the larynx

The accessory nerve has all the hallmarks of a branchial arch nerve as its cell bodies in the brainstem and cervical spinal cord are in line with other branchial efferent cell bodies (branchial efferent column); however, its evolutionary history, or more specifically the muscles it supplies (sternocleidomastoid and trapezius), is controversial, as are many of the musculoskeletal elements of the neck and shoulder (Matsuoka et al., 2005).



eFig. 2.5 Early development (**A–C**) of the face, (**D**) the adult form, and one of the many congenital abnormalities, (**E**) oblique facial cleft, that results from incomplete fusion of the facial swellings.

into the surrounding mesenchyme to form the *nasolacrimal duct*, which at this stage consists of a solid cord of ectodermal cells that grows upwards into the lids to form the canaliculi and downwards into the nose. Canaliculization of the solid columns of cells commences first in the vicinity of the lacrimal sac (3 months). Mesenchyme condenses and ossifies around this cord of cells to form the bony walls of the nasolacrimal canal see eBox 2.4 for discussion of abnormalities..

The upper eyelid develops from the fronto-nasal process and both lids are visible early in the second month (Fig. 2.1E). The primitive lid folds fuse in weeks 9–10 of gestation, enclosing a surface ectoderm-lined cavity, the conjunctival sac (Fig. 2.2D). Myocytes differentiate in the mesenchyme of both lids and eventually form the orbicularis oculi muscle (second arch derived; nerve supply, facial nerve). Meibomian glands, sebaceous glands and eyelashes develop as invaginations of the conjunctival epithelium or epidermis and are therefore ectodermal derivatives. In the seventh week the future *lacrimal gland* arises as a bud of epithelial cells from the region of the upper temporal conjunctiva sac.

In humans, at birth and in the infant, the eyes are in a comparatively advanced state of development relative to the rest of the face; the orbits are therefore large, although the remainder of the facial skeleton is small by comparison with the adult. During infancy and childhood, as a consequence of the development of teeth and growth of the paranasal sinuses, the facial skeleton steadily increases in size relative to the neurocranium. With the exception of the frontal sinuses, the paranasal sinuses develop as diverticulae from the nasal passages around the fifth month of gestation and are rudimentary at birth. The frontal sinuses appear in the fifth to sixth postnatal year as evaginations from the ethmoid sinuses or middle meatus.

Congenital malformations

Classification of congenital malformations or abnormalities is difficult for several reasons. First, the aetiology is often unknown and, even where a single genetic or environmental cause is suspected, it is often difficult to ascribe full responsibility to such agents or events because exposure to widely different teratogenic agents, such as drugs or trauma, can result in identical developmental defects. Second, defects may be strongly associated with chromosomal abnormality (e.g. trisomy 13).

Furthermore, disturbances in basic cellular events such as neural tube closure may lead to multisystem pathologies. In common with more generalized developmental defects, the consequences of exposure to teratogenic agents are highly dependent on the timing of major embryonic or fetal developmental events occurring at the time of exposure: there are important known *periods of vulnerability*, or critical (sensitive) periods, during which morphogenesis of particular systems and organs may be at risk. In this respect the eye is highly vulnerable for a long period of embryonic and fetal development because there are crucial developmental landmarks from the point of formation of the optic pits (day 22) until late in gestation, with such events as retinal vascularization and pupillary membrane regression occurring close to birth.

Congenital malformations affecting ocular tissues may have the following causes:

Genetic causes

- *Chromosomal anomalies:* deletions (*cri-du-chat* syndrome, Turner syndrome), trisomies (trisomy 13 or Patau syndrome, Down syndrome, Klinefelter syndrome) and triploidy (see Table 2.1 and eBoxes 2.5 and 2.6)
- *Hereditary:* either sporadic mutations (aniridia with Wilms' tumour) or dominant/recessive inheritance.

Environmental causes

- *Drugs:* alcohol, tobacco, anticonvulsants, thalidomide
- *Vitamins and minerals:* excess (e.g. hypervitaminosis A) or deficiency (e.g. folic acid, zinc)
- *Infection:* e.g. rubella, syphilis, toxoplasmosis
- *Radiation:* X-rays.

Maternal age

There is an increased incidence of genetic abnormalities in the oocyte with advancing maternal age.

Further reading

Bartelmez, G.W., Blount, M.P., Gage, P.J., Rhoades, W., Prucka, S.K., Hjalt, T., 1954. The formation of neural crest from the primary optic vesicle in man. Contr. Embryol. 35, 55–71.

Blankenship, T., Peterson, P.E., Hendricks, A.G., 1996. Emigration of neural crest cells from macaque optic vesicles is correlated with discontinuities in its basement membrane. J. Anat. 188, 473–483.

Bohnsack, B.L., Gallina, D., Thompson, H., Kasprick, D.S., Lucarelli, M.J., Dootz, G., et al, 2011. Development of extraocular muscles requires early signals from periocular neural crest and the developing eye. Arch. Ophthalmol. 129, 1030–1041.

Chan-Ling, T., 1997. Glial, vascular, and neuronal cytogenesis in whole-mounted cat retina. Microsc. Res. Tech. 36, 1–16.

Chauhan, B.K., Disanza, A., Choi, S.Y., Faber, S.C., Lou, M., Beggs, H.E., et al, 2009. Cdc42- and IRSp53-dependent contractile filopodia tether presumptive lens and retina to coordinate epithelial invagination. Development 136, 3657–3667.

Collinson, J.M., Hill, R.E., West, J., 2004. Analysis of mouse eye development with chimeras and mosaics. Int. J. Dev. Biol. 48, 793–804.

Creuzet, S., Vincent, C., Couly, G., 2005. Neural crest derivatives in ocular and periocular structures. Int. J. Dev. Biol. 49, 161–171.

Cvekl, A., Tamm, E.R., 2004. Anterior eye development and ocular mesenchyme: new insights from mouse models and human diseases. Bioessays 26, 374–386.

Dorrell, M.I., Friedlander, M., 2006. Mechanisms of endothelial cell guidance and vascular patterning in the developing mouse retina. Prog. Retin. Eye Res. 25, 277–295.

Erickson, C.A., Loring, J.F., Lester, S.M., 1989. Migratory pathways of HNK-1-immunoreactive neural crest cells in the rat embryo. Dev. Biol. 134, 112–118.

Fukiishi, Y., Morriss-Kay, G.M., 1992. Migration of neural crest cells to the pharyngeal arches and heart in rat embryos. Cell Tissue Res. 268, 1–8.

Gage, P.J., Rhoades, W., Prucka, S.K., Hjalt, T., 2005. Fate maps of neural crest and mesoderm in the mammalian eye. Invest. Ophthalmol. Vis. Sci. 46, 4200–4208.

Gilland, E., Baker, R., 1993. Conservation of neuroepithelial and mesodermal segments in the embryonic vertebrate head. Acta Anat. 148, 110–123.

Graw, J., 2019. Mouse models for microphthalmia, anophthalmia and cataracts. 2019 Mar 27. Hum Genet. https://doi.org/10.1007/s00439-019-01995-w. [Epub ahead of print].

Hall, B.K., 1998. Evolutionary Developmental Biology. Kluwer Academic Publishers, Dordrecht.

Hall, B.K., 2005. Bones and Cartilage: Developmental and Evolutionary Skeletal Biology. Elsevier Academic Press, London.

Heavner, W., Pevny, L., 2012. Eye development and retinogenesis. Cold Spring Harb. Perspect. Biol. 4 pii: a008391.

Hunt, P., Wilkinson, D., Krumlauf, R., 1991. Patterning the vertebrate head: murine *Hox2* genes mark distinct subpopulations of premigratory and migratory cranial neural crest. Development 112, 43–50.

Kwak, J., Park, O.K., Jung, Y.J., Hwang, B.J., Kwon, S.H., Kee, Y., 2013. Live image profiling of neural crest lineages in zebrafish transgenic lines. Mol. Cells 35 (3), 255–260. https://doi.org/10.1007/s10059-013-0001-5. Epub 2013 Feb 26.

Lang, R.A., 2004. Pathways regulating lens induction in the mouse. Int. J. Dev. Biol. 48, 783–791.

Lang, R.A., Bishop, J.M., 1993. Macrophages are required for cell death and tissue remodelling in the developing mouse eye. Cell 74, 453–462.

Larsen, W.J., 1997. Human Embryology, second ed. Churchill Livingstone, New York.

Lutty, G.A., McLeod, D.S. 2018. Development of the hyaloid, choroidal and retinal vasculatures in the fetal human eye. Prog. Retin. Eye Res. 62, 58–76.

Maden, M., Horton, C., Graham, A., Leonard, L., Pizzey, J., Siegenthaler, G., et al, 1992. Domains of cellular retinoic acid-binding protein I (CRABP I) expression in the hindbrain and neural crest of the mouse embryo. Mech. Dev. 37, 13–23.

Mann, I., 1928. The Development of the Human Eye. Grune and Stratton, New York.

Mann, I., 1964. The Development of the Human Eye, third ed. Grune and Stratton, New York.

Matsuoka, T., Ahlberg, P.E., Kessaris, N., Iannarelli, P., Dennehy, U., Richardson, W.D., et al, 2005. Neural crest origins of the neck and shoulder. Nature 436, 347–355.

McAvoy, J.W., Chamberlain, C.G., de Jongh, R.U., 1999. Lens development. Eye 13, 425–437.

McMenamin, P.G., 1989. A morphological study of the inner surface of the anterior chamber angle in pre- and post-natal human eyes. Curr. Eye Res. 8, 727–739.

McMenamin, P.G., 1989. The human fetal iridocorneal angle: a scanning electron microscopic study. Br. J. Ophthalmol. 73, 871–879.

McMenamin, P.G., 1991. A quantitative study of the prenatal development of aqueous outflow system in the human eye. Exp. Eye Res. 53, 507–517.

McMenamin, P.G., Djano, J., Wealthall, R., Griffin, B.J., 2002. Characterisation of the macrophages associated with the tunica vasculosa lentis of the rat eye. Invest. Ophthalmol. Vis. Sci. 43, 2076–2082.

McMenamin, P.G., Krause, W., 1993. Development of the eye in the North American opossum (*Didelphis virginiana*). J. Anat. 183, 343–358.

Mendelson, C., Larkin, S., Mark, M., LeMeur, M., Clifford, J., Zelent, A., et al, 1994. RARβ isoforms: distinct transcriptional control by retinoic acid and specific spatial patterns of promotor activity during mouse embryonic development. Mech. Dev. 45, 227–241.

Michalak, S.M., Whitman, M.C., Park, J.G., Tischfield, M.A., Nguyen, E.H., Engle, E.C., 2017. Ocular motor nerve development in the presence and absence of extraocular muscle. Invest. Ophthalmol. Vis. Sci. 58 (4), 2388–2396.

Miesfeld, J.B., Brown, N.L., 2019. Eye organogenesis: a hierarchical view of ocular development. Curr. Top. Dev. Biol. 2019 (132), 351–393.

Monaghan, A.P., Davidson, D.R., Sime, C., 1991. The Msh-like homeobox genes define domains in the developing vertebrate eye. Development 112, 1053–1061.
Noden, D., 1982. Periocular mesenchyme: neural crest and mesoderm interactions. In: Jakobiec, F.A. (Ed.), Ocular Anatomy, Embryology and Teratology. Harper and Row, Philadelphia, pp. 97–119.
Noden, D.M., 1988. Interactions and fate of avian craniofacial mesenchyme. Development 103 (Suppl), 121–140.
Noden, D.M., 1991. Vertebrate craniofacial development: the relation between ontogenetic process and morphological outcome. Brain. Behav. Evol. 38, 190–225.
O'Rahilly, R., 1966. The prenatal development of the human eye. Exp. Eye Res. 21, 93–112.
Osumi-Yamashita, N.O., Noji, S., Nohno, T., Koyama, E., Doi, H., Eto, K., et al, 1990. Expression of retinoic acid receptor genes in neural crest-derived cells during mouse facial development. FEBS Lett. 264, 71–74.
Ozanics, V., Jakobiec, F.A., 1982. Prenatal development of the eye and its adnexa. In: Jakobiec, F.A. (Ed.), Ocular Anatomy, Embryology and Teratology. Harper and Row, Philadelphia, pp. 11–96.
Patapoutian, A., Miner, J., Lyons, G.E., Wold, B., 1993. Isolated sequences from the linked *Myf-5* and *MRF4* genes drive distinct patterns of muscle-specific expression in transgenic mice. Development 118, 61–69.
Reese, B.E., Necessary, B.D., Tam, P.P.L., Faulker-Jones, B., Tan, S.S., 1999. Clonal expansion and cell dispersion in the developing mouse retina. Eur. J. Neurosci. 11, 2965–2978.
Robinson, S.R., 1991. Development of the mammalian retina. In: Dreher, B., Robinson, S.R. (Eds.), Neuroanatomy of the visual pathways and their development; Cronly-Dillon JR, series ed. Vision and visual dysfunction, vol. 3. Macmillan, Basingstoke, pp. 69–128.
Saha, M.S., Spann, C., Grainger, R.M., 1989. Embryonic lens induction: more than meets the optic vesicle. Cell Differ. Dev. 28, 153–171.
Saint-Geniez, M., D'Amore, P., 2004. Development and pathology of the hyaloid, choroidal and retinal vasculature. Int. J. Dev. Biol. 48, 1045–1058.
Schulz, M.W., Chamberlain, C.G., de Longh, R.U., McAvoy, J.W., 1993. Acidic and basic FGF in ocular media and lens: implications for lens polarity and growth patterns. Development 118, 117–126.
Spaeth, G., Nelson, L.B., Beaudoin, A.R., 1982. Ocular teratology. In: Jakobiec, F.A. (Ed.), Ocular Anatomy, Embryology and Teratology. Harper and Row, Philadelphia, pp. 955–1080.
Sullivan, C.H., Braunstein, L., Hazard-Leonards, R.M., Holen, A.L., Samaha, F., Stephens, L., et al, 2004. A re-examination of lens induction in chicken embryos: in vitro studies of early tissue interactions. Int. J. Dev. Biol. 48, 771–782.
Tatiana V. Petrova, Gou Young Koh (2018). Organ-specific lymphatic vasculature: From development to pathophysiology. J. Exp. Med. 2018 2; 215(1): 35–49.
Trainor, P.A., Tam, P.P.L., 1995. Cranial paraxial mesoderm and neural crest cells of the mouse embryo: co-distribution in the craniofacial mesenchyme but distinct segregation in the branchial arches. Development 121, 2569–2582.
Zieske, J.D., 2004. Corneal development associated with eyelid opening. Int. J. Dev. Biol. 48, 903–911.

Chapter 3

Genetics

CONTENTS

Introduction
Chromosomes and cell division
- Chromosomes
- Cell division

Molecular genetics (DNA and genes)
- How genes work: decoding DNA
- Regulating gene expression is constantly active, has multiple pathways, and is influenced by environment (epigenetics)

Chromosome defects and gene mutations
- Structural chromosomal abnormalities
- Gene mutations

Clinical genetics
- Autosomal inheritance
- Mitochondrial inheritance
- Multifactorial inheritance
- Disease associations

Population genetics
- Gene frequency
- Genetic linkage and linkage analysis
- Epigenetics

Understanding the human genome: DNA analysis
- Cloning
- Refining technology: identifying and manipulating the gene
- Polymerase chain reaction
- Genomics, transcriptomics and proteomics
- Finding and tracking the gene through families

Molecular biology and clinical medicine
- Gene probes

Molecular and cell biology: controlling cell destiny
- Using molecular biological technology to detect apoptosis
- Gene therapy
- Genes, cell differentiation and cell-based therapies
- Gene editing and repair

Molecular genetics and ophthalmology
- The X chromosome
- Retinitis pigmentosa
- Choroideraemia
- Norrie disease
- Retinoblastoma
- Albinism
- Myotonic dystrophy
- Mitochondrial inheritance

Introduction

Genetic disorders are often thought of as interesting but rare. However, in a population of 100,000, there will be 10 children born with a genetic disorder, a further 10 who will develop genetic disorders in later life, and more than 10% of adults will have a chronic illness with a strong genetic basis. The advances in medical and molecular genetics have increased our understanding of disease processes, which can be illustrated particularly with examples of inherited ophthalmic disorders. Moreover, the increasing ability to define individually, or within disease populations, the genome, polymorphisms in gene expression, control of translation and the impact of environment on genetic expression or heritable as well non-heritable non-DNA influence on gene expression (epigenetics) has profoundly influenced our interrogation of disease pathogenesis and targeted therapeutic intervention. This chapter will further highlight particular advances in genome-wide association studies (GWAS), functional genomics leading to advances in therapeutics (gene therapy, RNA therapies and clustered regularly interspaced short palindromic repeats (CRISPR) technology), in particular describing advances in gene therapy for monogenetic disorders and expanding with secreted

proteins for non-mendelian complex diseases, RNA therapies including anti-sense oligonucleotide therapies and the potential of gene editing utilizing CRISPR technology.

Chromosomes and cell division

Chromosomes

Human cells contain 46 chromosomes (diploid), consisting of 22 almost identical pairs (homologous autosomes) and a pair of sex chromosomes, which constitute the karyotype of an individual. Chromosomes are numbered 1–22 as they decrease in size. The chromosome consists of two arms: the short arm is designated 'p' and the long arm 'q'. During somatic division, or mitosis, each chromosome, and therefore gene, replicates precisely. Gamete formation involves a special reduction division (meiosis), in which the homologous chromosomes separate from each other so that each cell contains 23 chromosomes (haploid). Since chance determines which chromosome of a pair ends up in a particular gamete, there is a possibility of 2^{23} combinations in the gametes. The different stages of the two forms of cell division are summarized later. The genes are located in a linear order along the chromosome, each gene having a precise location or *locus*. Each of a pair of chromosomes (homologous chromosomes) carries matching genetic information, i.e. they have the same sequence of gene loci. However, at any locus they may have slightly different forms, which are called *alleles*. Cytogenetic analysis of the chromosomes can be performed by looking at the banding pattern of individual chromosomes. Giemsa staining highlights the horizontal banding of the chromosome by the characteristic pattern of light and dark staining (G-banding). More specific patterns of staining can be performed by looking at *high-resolution banding* with G-banding patterns at the early stages of mitosis (prophase; see later section). *Fragile sites* are non-staining gaps that can be demonstrated only when the chromosomes are cultured in specific conditions (e.g. thymidine and folic acid deprivation) and then stained.

Cell division

Mitosis

During mitosis each chromosome and every gene along its length replicate to produce two identical daughter cells, so that every nucleated cell (except gametes) has 46 chromosomes (diploid). In culture, the cell cycle of most mammalian cells varies but is usually about 24 hours. Mitosis itself occupies approximately 1 hour of the total time, whereas the time taken to synthesize DNA required for replication is 6–8 hours (Fig. 3.1) (Box 3.1). The gap phases G_1, between mitosis and synthesis of DNA, and to a lesser extent G_2, between synthesis and mitosis, govern the turnover of the cells because some cells can rest in G_1 for days. It appears

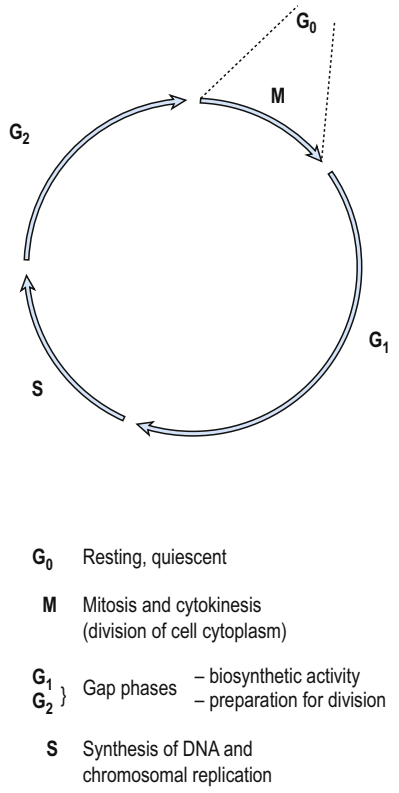


Fig. 3.1 The cell cycle. *S*, Synthesis of DNA; *M*, mitosis and cytokinesis; G_1 and G_2, gap phases.

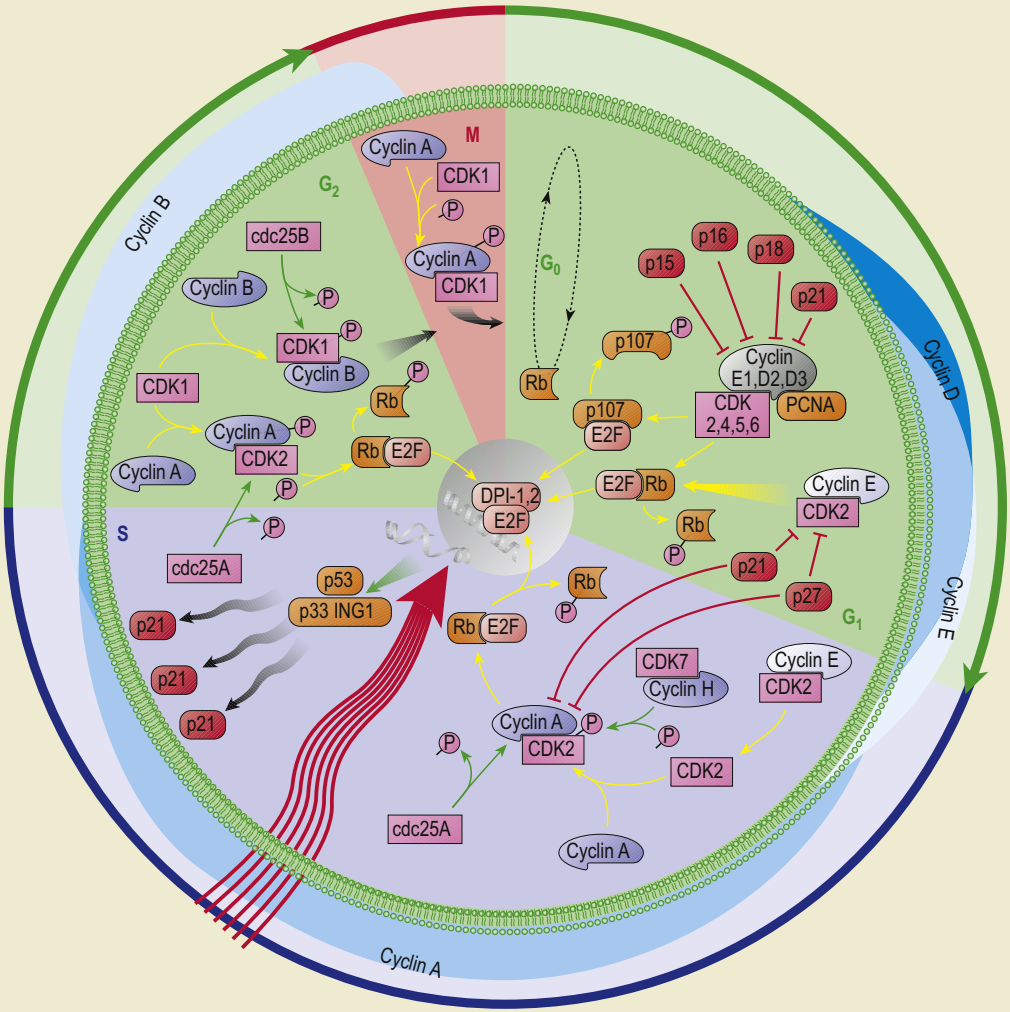
Box 3.1 **Stages of mitosis**

Interphase	Resting phase
Prophase	Chromosomes identifiable in nucleus
Metaphase	Chromosomes aligned in centre of nucleus
Anaphase	Centromere of chromosome divides with chromatid separation
Telophase	Daughter chromosomes separate

that late in G_1 the cell passes a *restriction point* after which it will proceed through the rest of the cell cycle at a standard rate. During the S phase each chromosome replicates, each with its own individual pattern, although homologous pairs of chromosomes replicate synchronously. During G_2, the cell gradually enlarges, doubling its total mass before the next mitosis. Some cells, however, do not divide once they are fully differentiated (e.g. neurones and erythrocytes) and are permanently arrested in a phase known as G_0. Control of cell cycle is represented as a schematic in Box 3.2.

Box 3.2 **Control of cell cycle**

Cyclins are regulatory subunits within the cell that are periodically synthesized and degraded, regulating cyclin-dependent kinases (CDK) essential for cell cycle control.

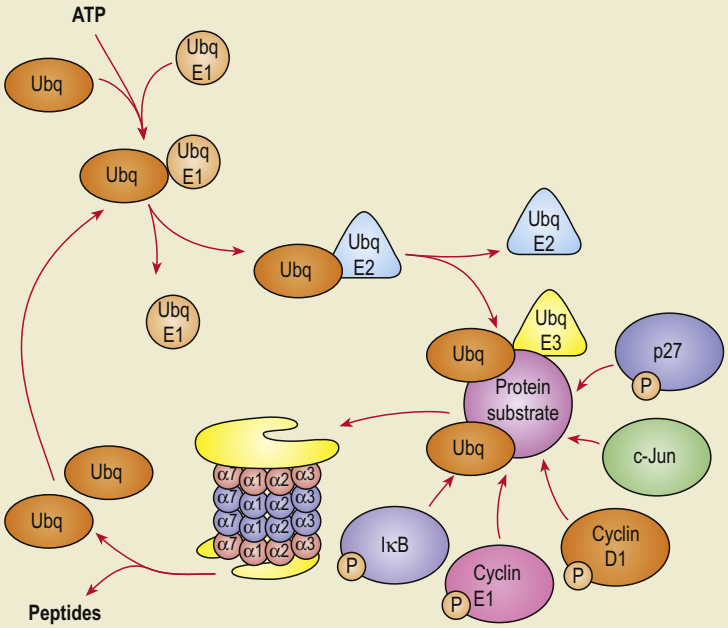


Cyclins (of which to date there are nine; A–I) bind to CDKs, forming an active complex that in a controlled fashion drives the cell through its stages via phosphorylation of unique protein substrates, such as retinoblastoma protein, Rb, allowing progression into the next phase of the cycle. Interaction of cyclin–CDK complexes with the retinoblastoma (Rb) protein family (Rb, p107, p130) is regulated by proteins that block cyclin–CDK activity (e.g. p21). p53, a cell cycle regulator, leads to increased p21 expression.

Continued

Box 3.2 **Control of cell cycle—Cont'd**

Ubiquitinylation of the various proteins (see Chapter 4, page 169) ensures regulated control of the various stages of protein interactions.



Meiosis: formation of gametes

Meiosis takes place in two stages, both having the same stages as mitosis. In the first stage of meiosis there is a prolonged prophase, in which cell division of only the chromosomes occurs so that each subsequent daughter nucleus contains half the number of chromosomes (23), resembling mitosis but in the haploid state. The prophase of the first meiotic division is complex and may be divided into a further five stages.

- *Leptotene.* Starts with the first appearance of the chromosome.
- *Zygotene.* The homologous chromosomes pair and bind closely to form bivalents.
- *Pachytene.* The main stage of chromosomal thickening; each chromosome consists of two chromatids and, because they are bivalent, has four strands. During this stage, chromatids of the chromosomes exchange material (crossovers), further ensuring a random assortment of paternal and maternal homologues.
- *Diplotene.* The bivalents start to separate (the centromere of each remains intact).
- *Diakinesis.* The bivalent chromosome formations separate and coil tightly.

The chromosomes then undergo *metaphase* and *anaphase*, where the bivalents disjoin, going to each equatorial plane of the cell and as the cytoplasm divides, each cell has 23 chromosomes, each with a pair of chromatids. The second meiotic division follows the first without an interphase and resembles mitosis. Thus, meiosis gives rise to gametes that contain only one representative of each homologous pair of chromosomes and there is random selection of paternal and maternal homologues.

Genetic makeup

The genetic makeup of each individual or of a specific patient population is a result of mutations and natural selection in the past. During cell division, the division of chromosomes may be imperfect. A mutation is a sudden change in the gross or fine structure of DNA, such as might be caused by a replication error or a crossover defect between misaligned chromosomes at meiosis, giving rise to deletion or duplication of DNA. The mechanisms by which chromosomal abnormalities may occur are discussed later. A mutation may affect a specific gene and thus the protein structure it codes for. Equally, the mutation may alter the structure of the protein in such a way that the function of the protein is changed only slightly but the metabolism remains unaffected. If the gene is destroyed altogether, including the nucleotide sequence, which is critically important for the protein that it encodes,

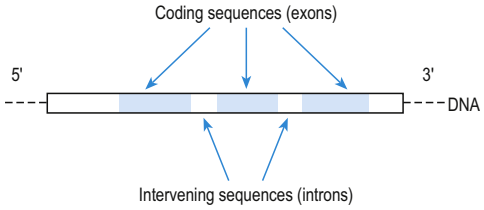


Fig. 3.2 Components of a gene.

then the individual will be at an immediate disadvantage. It is therefore not surprising that genetic diseases occur as a result of harmful mutations as opposed to diseases with multifactorial inheritance patterns, which have a predictable mode of inheritance.

Molecular genetics (DNA and genes)

The basis of inherited disease can be understood from information about chromosomal makeup, specific genes and how proteins are encoded. The human haploid genome consists of 3×10^9 base pairs of double-stranded DNA (dsDNA). The genetic information that determines the sequence of amino acids in peptide chains is stored in the DNA by the order of the nucleotide bases: adenine (A), cytosine (C), guanine (G) and thymine (T). Three bases (a codon) code for a particular amino acid. A gene is part of the DNA that directs the sequencing of polypeptide chains. The gene itself consists of both coding regions (*exons*) and variable-length intervening regions (*introns*). There are also enhancer regions of the gene, which carry out regulatory activities, particularly the expression of genes in tissues. Similarly, there is a family of regulatory molecules on the gene that activate or suppress both pairs of homologous genes (Fig. 3.2).

How genes work: decoding DNA

All proteins are encoded in DNA and, as already mentioned, each amino acid is represented by a triplet code of bases (codon). Since the number of possible codons is greater than the number of amino acids, most amino acids, except methionine and tryptophan, may be encoded by more than one codon. The code AUG (methionine) acts as a starter signal for protein synthesis (Box 3.3) and there are also codes that act as polypeptide chain terminators (stop codons).

Polypeptide chain synthesis occurs in two processes: *transcription* and *translation* (Fig. 3.3). During transcription one strand of DNA forms a template for the synthesis of messenger RNA (mRNA), catalysed by the enzyme RNA polymerase. RNA polymerase binds to a specific DNA sequence known as the *promoter*, which signals the unwinding of the DNA helix, which acts as a template for the complementary base pairs. The process moves along the DNA molecule, extending the RNA molecule in the 5′ to 3′ direction. This process continues until it meets the termination signal on the DNA molecule. The first transcript is an exact replica of the single strand of DNA. Before leaving the nucleus, the introns are excised and the exons are spliced together. On leaving the nucleus, the mRNA acts as a template on cytoplasmic ribosomes for protein synthesis. The mRNA 5′ end is blocked before the whole molecule is transcribed with 7-methylguanine joined 5′ to 5′ and a 3′ poly-A tail, which aids transport into the cytoplasm. The poly-A tail is added after transcription by *poly-A polymerase*. Translation of nucleotide sequences is performed by specific transfer RNA (tRNA): amino acids, each specific for three nucleotide bases, are attached before polymerization into polypeptide chains. They are attached by their C-terminal ends and are transformed into high-energy molecules, from which peptide bonds form to produce polypeptides. Each tRNA (anticodons) is linked with the appropriate codons on the mRNA. This process of translation is controlled by a set of enzymes called *aminoacyl-tRNA synthetases*, which couple each amino acid to the appropriate tRNA molecule. The mRNA sequence is read three nucleotides at a time from the 5′ to 3′ direction along the mRNA. The polypeptide chain is constructed from its N-terminal end to its C-terminal end and the growing C-terminal end of the protein remains activated by covalent attachment to the tRNA (peptidyl-tRNA molecule). The mRNA is coded with specific start and stop codons, which thus determine the length of the polypeptide chain. The initiation of polypeptide chain synthesis is under the control of initiation factor 2 (IF-2), which catalyses the reaction between a specific initiator tRNA molecule and the mRNA. The growing polypeptide chain must be exactly three nucleotides along the mRNA

Box 3.3 Summary of protein synthesis

Messenger RNA is produced from DNA transcription under control of the enzyme 5′-3′ RNA polymerase. Translation begins in the cell cytoplasm by the binding of mRNA to ribosomal subunits, to which an initiator tRNA molecule binds. Amino acid-specific tRNAs have anticodons specific for a triplet of nucleotides (codon), forming the aminoacyl-tRNA molecule, which binds to the A site on the ribosome under the control of an enzyme, aminoacyl-tRNA synthetase. Each amino acid is added to the C-terminal of the growing polypeptide chain, progressing from codon to codon in the 5′ to 3′ direction of the mRNA. The length of the polypeptide chain is regulated by specific start and stop codons.

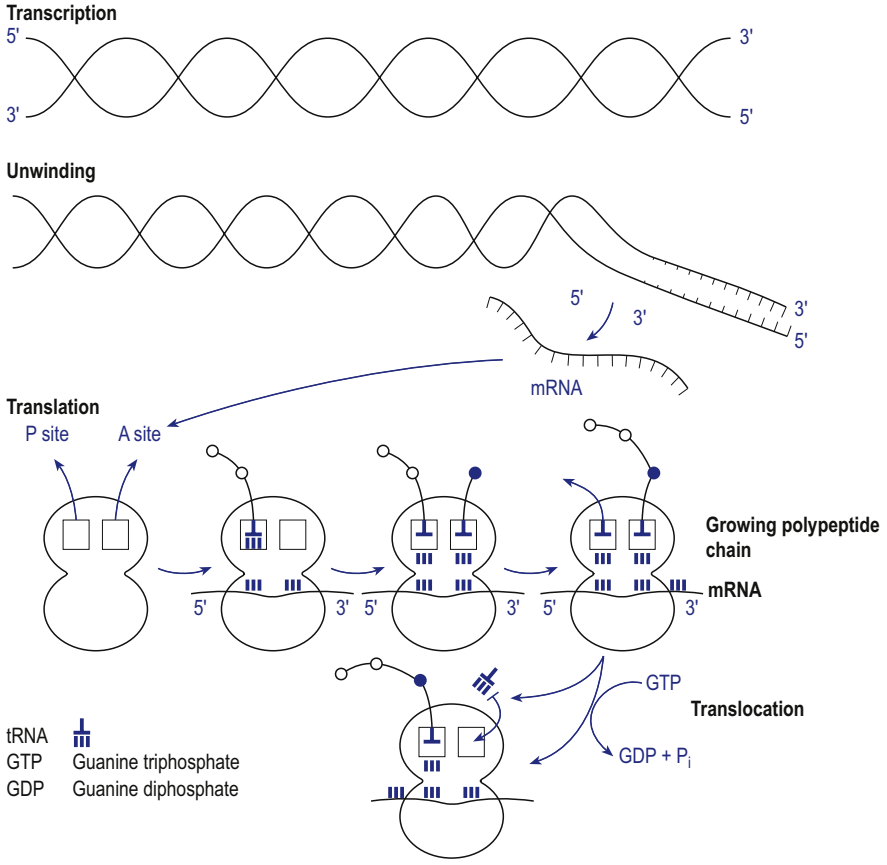


Fig. 3.3 Protein synthesis: transcription and translation.

after the addition of each amino acid, which requires a multi-enzyme system incorporated within the ribosome. Ribosomes are half RNA by weight and have a groove that accommodates a growing polypeptide chain and a groove that accommodates each mRNA molecule. On the ribosome there are two different binding sites for tRNA. One site holds the tRNA and growing polypeptide chain, the peptidyl-tRNA-binding site (P site), and the other site holds the incoming tRNA molecule, the aminoacyl-tRNA-binding site (A site). The process of polypeptide chain elongation involves two steps. The first step involves an aminoacyl-tRNA molecule becoming bound to the A site adjacent to an occupied P site. The second step involves uncoupling the C-terminal end of the polypeptide chain from the tRNA in the P site and joining it to the tRNA molecule in the A site. The reaction is catalysed by *peptidyl transferase*, which is present on the ribosome. The new peptidyl-tRNA molecule is then transferred to the P site, releasing the now unoccupied tRNA molecule and the whole process is repeated. The stop codon on the mRNA initiates a protein called *release factor* to bind to the A site, causing hydrolysation of the peptidyl-tRNA molecule on the P site and subsequent release of the polypeptide chain into the cell cytoplasm.

Genes are regulated by adjacent sequences on the genome. Upstream of the gene is the *promoter*, which is involved in the attachment of RNA polymerase to the DNA strand. Promoters have an element called a TATA box, which specifies the correct 5′ end of the RNA precursor. Several promoter-specific transcription factors have now been isolated, each specific for a gene promoter which, when bound, activates transcription. TATA-binding protein is a general transcription factor for RNA polymerases I, II and III. *RNA polymerase III* is responsible for the synthesis of a variety of small RNA molecules including tRNA and small subunits of ribosomal RNA. *RNA polymerase II* carries out the transcription of genes, and *RNA polymerase I* synthesizes the large subunits of ribosomal RNA.

Regulating gene expression is constantly active, has multiple pathways, and is influenced by environment (epigenetics)

Ordering of DNA and accessibility for transcription

Histones are proteins that package DNA within nucleosomes. They act to form a structure that DNA spools around, are key to gene regulation, and are influenced by environmental change (*epigenetics—changes in gene expression not caused by changes in DNA sequence*). Several families of histones exist and the formation and structure of the nucleosome, which in turn designates which areas of DNA are open for transcription, is influenced by key enzymatic modification, such as methylation, acetylation, phosphorylation, ubiquitination and SUMOlyation (see also Chapter 4 page 169). For example, methylation of histone 3 at lysine residue 4 (H3K4Me3) at the promoter **of active genes is associated with active transcription, as opposed to methylation at lysine 27, H3K27Me3**, which results in gene repression.

Similar mechanisms may occur at DNA level. DNA methylation, in particular *CpG* repeats, confers an 'off-switch' where genes remain repressed. Methylation is essential for cell cycle and cell differentiation and alteration (which is part of epigenetic target) in methylation profile or sites of methylation can alter cell differentiation, fate and gene function.

Non-coding RNA and RNA-binding proteins constantly regulate gene expression

Gene expression can be further regulated during or after transcription (post-transcriptional regulation of gene expression). For example, encoded by nuclear DNA are microRNA (miRNA), which are small non-coding RNA molecules (around 22 nucleotides) that have complementary sequences with mRNA and usually result in regulation via gene silencing or via RNA-binding proteins, which are nuclear or cytoplasmic proteins that regulate RNA transcripts (for example, within the spliceosome).

More work is still required on the understanding of RNA-binding proteins and biological and disease relevance; however, we do appreciate that miRNA have multiple mRNA targets, while each mRNA may also be targeted by several miRNA. Given that our genome codes for over 1000 miRNA, there is a rich network by which regulation of gene expression occurs, while aberrant expression of miRNA has been implicated in disease, in particular inflammation and cancer. Such miRNA has become very topical in the pathogenesis of degenerative disorders such as age-related macular degeneration. The formation of RISC (RNA-induced silencing complex) incorporates miRNA and activates enzyme RNA endonuclease III Dicer 1 that aids formation of small interference RNA (siRNA) within RISC and inhibition of gene expression (Fig. 3.4). Non-coding RNA do influence translation, transcription and chromatin remodelling and therefore engage and regulate networks for normal physiological processes and disease, including ageing of cells. The findings that *C. elegans* lifespan can be positively or negatively regulated by specific miRNA has opened understanding of regulation in animal models and how miRNA are altered in age-related diseases and cancers. miRNA are also secreted in extracellular vesicles and exosomes into extracellular matrix, for example, secreted by senescent cells.

Spliceosomes

When the gene is transcribed (Fig. 3.3), the precursor mRNA is then further modified in spliceosomes where non-coding sequences are removed and coding sequences (exons) are ligated. Recognizing this process is important, as some proteins undergo 'alternative' splicing and thus form other active mRNA with the consequence of altered protein action and cell response and behaviour. Understanding mechanisms of splicing and the spliceosome therefore remain important, as altered activity is implicated in disease and targets for therapeutics.

Chromosome defects and gene mutations

During cell division, the division of chromosomes may be imperfect. The extent, however, of genetically determined disease remains undetermined. The risk of genetic abnormality detectable at birth or in infancy is approximately 1 in 40. However, most fetuses with chromosomal aberrations are spontaneously aborted and it has been estimated that the frequency of genetic defects in live births is 0.6% and in stillbirths 5%. Chronic diseases with a significant genetic contribution occur in about 10% of the adult population. Abnormalities of the chromosomes are usually classified as numeric abnormalities, where somatic cells contain an abnormal number of normal chromosomes, or as structural abnormalities, where somatic cells contain one or more abnormal chromosomes, which in turn may affect either sex or autosomal chromosome.

Further information on numerical chromosomal abnormalities is available online at https://expertconsult.inkling.com

Numerical chromosomal abnormalities

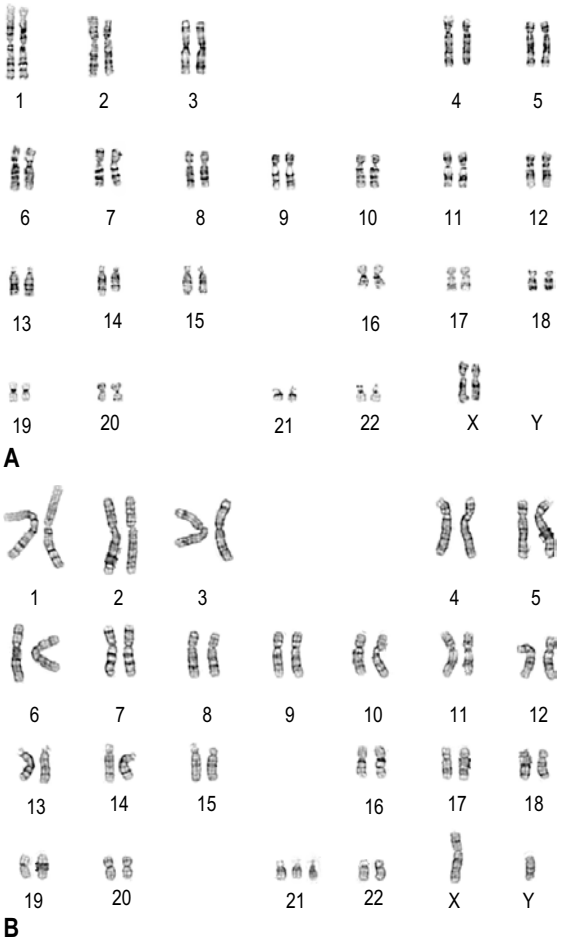
Aneuploidy arises from failure of paired chromosomes to disjoin or as a result of delayed movement during anaphase, thus producing an extra copy of the chromosome (trisomy) or a missing copy of the chromosome (monosomy). Meiotic aneuploidy can occur during either of the stages of meiosis. Polyploidy describes a complete extra set of chromosomes (69 in triploidy), which usually results from fertilization of the ovum by two sperm cells. This usually results in miscarriage.

Autosomal trisomies

Trisomy describes the individual chromosome that is somatically triplicated in the cell. Triplication of individual chromosomes gives rise to characteristic syndromes with multisystem disorders. Examples of autosomal trisomies include trisomy of chromosome 21 (Down syndrome), chromosome 18 (Edwards syndrome) and chromosome 13 (Patau syndrome). In each of the autosomal trisomies, mosaicisms (two or more cell lineages derived from a single zygote) may influence the clinical picture markedly, particularly when the normal cell line is present in a critical tissue. Trisomy, as mentioned previously, may result from a failure of the separation of chromosome pairs during cell division (non-disjunction) or (in approximately 4% of cases) from translocation of one chromosome onto another inherited from one of the parents, thus giving the offspring a normal number of chromosomes but with a translocated copy of, for example, chromosome 21 (Robertsonian translocation) (eFig. 3.1).

Meiotic aneuploidy

Similar chromosomal anomalies exist with the sex chromosomes, whereby an extra X or Y chromosome gives rise to a phenotype not distinguished in the normal population. Loss of an X chromosome gives rise to Turner syndrome (45 X0), which again may not present with the characteristic phenotype because of the prevalence of mosaic forms and therefore normal sexual development may be present. This is also true for Klinefelter syndrome, in which there is an extra X chromosome (47 XXY).



eFig. 3.1 (**A**) Normal human karyotype: 22 autosome pairs and sex chromosomes. (**B**) Trisomy of chromosome 21.

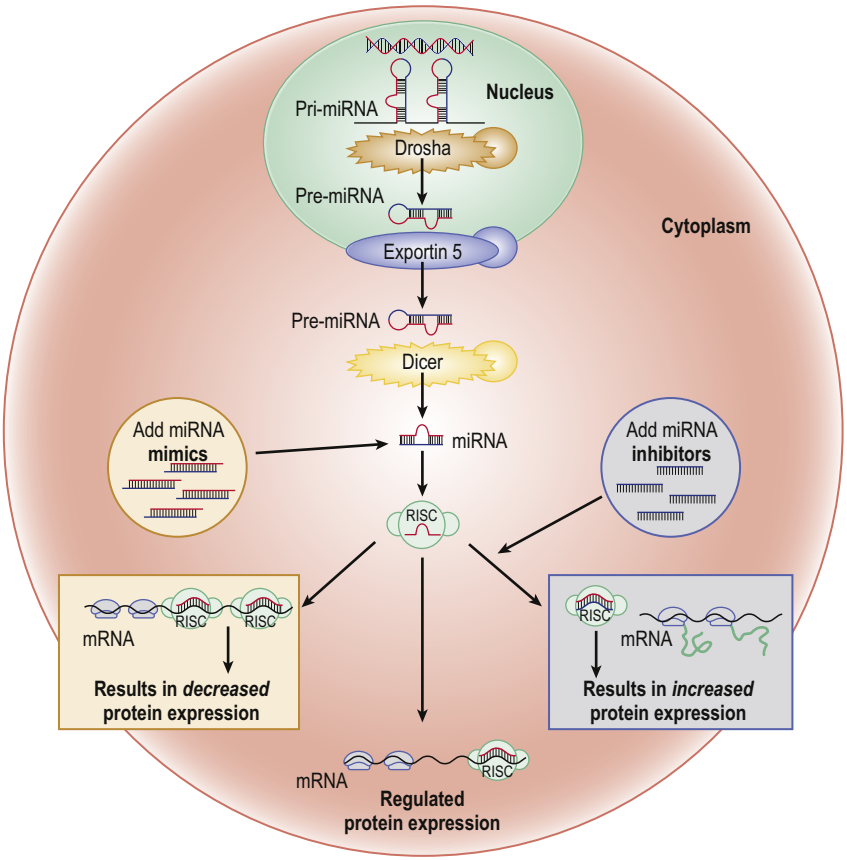


Fig. 3.4 Micro-RNA pathway (adapted from an illustration by Michael Hamers, Lightspeed Design and Branding, Boulder, CO, USA, with permission from miRagen Therapeutics http://www.miragentherapeutics.com.)

Structural chromosomal abnormalities

Structural chromosomal abnormalities result from chromosomal breakage. Normally, single breaks are repaired quickly, but if more than one break occurs, repair mechanisms may cause random re-joining of the wrong ends. Spontaneous breakage increases with exposure to mutagenic chemicals and ionizing radiation. The following structural abnormalities might occur:

- translocation—chromosomes break and exchange segments
- inversion—segment of chromosome is inverted in sequence
- deletion—section of chromosome is lost
- mutation—point mutation occurs with a change in a single base of a triplet code of a gene.

Translocation usually results in no loss of DNA so that individuals might appear clinically normal. Translocation may be reciprocal if exchange of chromosomal segments distal to the breaks occurs. *Robertsonian* translocation arises from breaks at or near the centromere in two acrocentric chromosomes (where centromere is located nearer one end of the chromosome). For example, translocation of segments on chromosomes 14 and 21 may occur and, during meiosis, a trivalent is formed which will then lead to a mosaic of normal and abnormal karyotype during anaphase. Insertional translocation requires the occurrence of three breaks in one or two chromosomes, resulting in deletion of one segment and insertion of another into the gap in the first chromosome. Inversions arise from two chromosomal breaks and the segment being inverted through 180 degrees between the breaks. Inversions interfere with the pairing of chromosomes during meiosis and crossovers are suppressed, generating unbalanced gametes. Deletion may also occur when only part of the chromosome is lost. This can occur between two breakpoints or as a result of breaks and loss of segments

Box 3.4 **Mutations within the human genome**

Point mutation (nucleotide substitution)

A single nucleotide change in the transcribed codon leading to a different amino acid substitution in the polypeptide chain—***missense mutation.***

A single nucleotide change that results in a stop codon, giving rise to a shortened polypeptide chain, or vice versa, resulting in an elongated polypeptide chain—***nonsense mutation.***

A single nucleotide base change that alters a critical splice junction and leads to abnormal RNA processing or a reduction in the normal gene product, or prevents the efficient addition of the poly-A chain for effective transcription—***splice site mutation.***

Deletions and insertions (frameshifts)

Deletion of one or two nucleotide bases leading to incorrect reading during translation and resulting in an inappropriate amino acid sequence and premature termination of the polypeptide chain.

Codon deletions andiInsertion (repeat expansions)

Insertion of a repeated codon (three nucleotide bases) that, as a consequence, interrupts the coding sequence. These mutations, known as ***triplet repeats***, are found in myotonic dystrophy, the fragile X syndrome (X-linked mental retardation), Huntington's disease and spinocerebellar ataxia.

in both arms of the chromosome. Point mutations, on the other hand, occur when a single nucleotide base is replaced by another, which may in turn alter the amino acid coding for that protein.

Gene mutations

Many types of mutation may occur throughout the alleles at each locus. Within the normal population and in examples of inherited disease, mutations can occur that range from a single base pair to deletion of large gene segments involving many millions of base pairs. The description of these mutations has led to increasing availability of diagnostic and screening tools for genetic disease. Box 3.4 details the diversity of gene mutations that occur within the human genome.

Clinical genetics

There are several main types of genetic disorders: chromosomal (see earlier section), mitochondrial, multifactorial, somatic cell genetic and single gene (autosomal and X-linked inheritance) disorders (Box 3.5).

Box 3.5 **Characteristics of Mendelian inheritance**

Autosomal dominant

- Vertical transmission
- 50% of offspring affected
- Males and females affected equally
- Variable expressivity
- Unaffected persons do not pass on trait
- Homozygotes are more severely affected

Autosomal recessive

- Recessive gene does not cause disease in heterozygotes
- Disease expressed in homozygotes
- Males and females affected equally
- Constant expressivity
- Affected individuals have children who are carriers

X-linked recessive

- Vertical transmission
- Usually only males are affected
- All daughters of affected males are carriers
- Heterozygous females may be affected because of 'lyonization'
- Variable expressivity

Autosomal inheritance

A trait that is determined by a gene on an autosome may be *dominant* or *recessive*. Heterozygotes are individuals with different alleles at the corresponding locus on the pair of homologous chromosomes. Homozygotes have the same allele. As such, autosomal dominant refers to the situation where a monogenetic disorder manifests clinically in the heterozygous state and inheritance is usually from one parent only. Autosomal recessive traits refer to disorders that, while they can affect both sexes, require both abnormal genes for the disease to be clinically expressed.

Further information on mode of inheritance of gene disorders is available online at https://expertconsult.inkling.com

X-linked disorders

X-linked inheritance refers to the pattern of inheritance carried by the genes on the sex chromosomes. This inheritance therefore carries a characteristic family pedigree. An X-linked recessive trait carried on the X chromosome

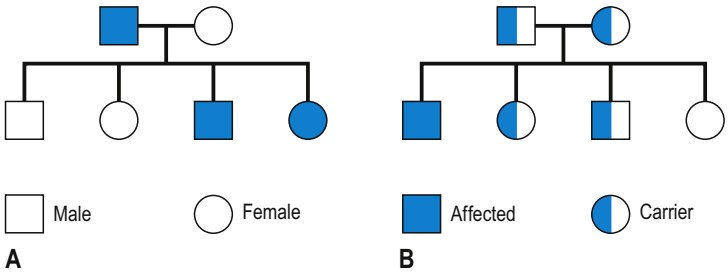
Inheritance of genetic disease

(A) Autosomal dominant disorders

The overall incidence of autosomal dominant inheritance (eFig. 3.2A) of disease in the UK is 7 per 1000 live births. Usually there is one affected parent, unless the trait occurs as a new mutation, in which case neither parent expresses the disease. Affected individuals have a 1 in 2 chance of passing it on to their child. This form of inheritance is not related to sex chromosomes and male-to-male transmission may occur. In contrast, X-linked disorders cannot be transmitted by males (see later section). In clinical practice the detection of autosomal dominantly inherited disease is complicated by variation in expression of the gene and by new mutations. As such the clinical manifestation in an affected individual may not be apparent, giving rise to the view that the condition has skipped a generation. Most autosomal dominant disorders present clinically with features in later life, often after the carriers of the gene have completed their families. This raises difficulties in genetic counselling and, for this reason, consideration is now being given to developing predictive tests based on gene tracking with DNA probes. The more severe the disorder, the less likely the individual is to reproduce, so that the proportion of individuals affected as a result of new mutations will increase. Both new mutations and variable gene expression can combine to create difficult clinical decisions for the affected individuals and the clinical geneticist.

(B) Autosomal recessive disorders

Autosomal recessive traits affect both sexes but the trait is manifest only if both abnormal genes are present (eFig. 3.2B), i.e. the patient has no normal allele at the affected locus. Usually both parents are heterozygous carriers of the gene in question and are clinically normal. Thus, there is a 1 in 4 chance of any offspring being affected, a 1 in 2 chance of producing a heterozygous carrier and a 1 in 4 chance of being normal. In rare recessive traits, there is usually a strong family history of consanguinity between first cousins, who share 1 in 8 of their genes by virtue of their common ancestry. In practical terms the incidence of severe autosomal recessive disorders is low, making the risk to the sibling of first-cousin parents small. In the UK the overall incidence of autosomal recessive disorders is 2.5 per 1000 live births, one of the commonest being cystic fibrosis.



eFig. 3.2 Pedigree of (**A**) autosomal dominant inheritance, (**B**) recessive inheritance.

Box 3.6 **Lyonization**

Lyonization is the process of inactivation of one member of a pair of X chromosomes in every female somatic cell. The Barr body (densely stained mass of chromatin, which is the inactivated X chromosome) occurs in the trophoblast by day 12 after fertilization. Inactivation occurs only in somatic cells; it is random and then fixed for all subsequent generations of that cell line. The inactive X chromosome is not transcribed, except for a small region at the tip of the short arm. If loss of material occurs in one X chromosome, then the structurally abnormal chromosome is inactivated. About 30% of cells from buccal smears show a Barr body and 1%–10% of female neutrophils will also show an inactivated X chromosome (drumstick). Females have random expression of either paternal or maternal X chromosome, which clinically gives rise to patchy expression of mutant X-linked genes in carrier females (e.g. retinitis pigmentosa and choroideraemia). Lyonization is separate to, for example, allelic exclusion as observed with Immunoglobulin (Ig) genes during B cell development. (See Chapter 7).

is manifest only in females when the homozygous state exists, because in heterozygotes the normal dominant gene would be expressed. Such individuals may express disease traits. Traits are thus transmitted from healthy female carriers or affected males. An affected male would pass the trait to all his daughters, who would then be heterozygous carriers, but he would be unable to pass the disorder on to his sons, who receive only his Y chromosome. In some X-linked recessive disorders, a proportion of female heterozygotes are affected more as a trait. The explanation lies in the fact that only one of the X chromosomes in a cell is active (Lyon hypothesis; see Box 3.6). This inactivation occurs early in ontogeny and thereafter the descendants of the cell have the same inactive X chromosome. By chance, therefore, females may inactivate the healthy chromosome and thus manifest the disease.

In the UK the incidence of X-linked inherited disorders is around 1 per 1000 live births. The term X-linked dominant has been used to describe X-linked inheritance where the heterozygous female is regularly affected. In these disorders, males are frequently more severely affected and do not survive gestation (e.g. Aicardi syndrome: agenesis of the corpus callosum and chorioretinopathy). If the affected male does reproduce, none of the sons will be affected and all the daughters will be. Occasionally these pedigree patterns may be mimicked by autosomal traits that display sex limitation. If the affected males of an autosomal dominant trait with sex limitation are infertile, the pedigree pattern is identical to that of X-linked recessive traits, except that carrier females exhibit lyonization. The family pedigree will also show a smaller proportion of affected females than would be expected in X-linked dominant traits.

Mitochondrial inheritance

Cells have multiple copies of an extranuclear chromosome, contained in mitochondria. The mitochondria contain about 10 single circular chromosomes, which replicate independently of the nuclear genome. Within their genome there are the genes for tRNA and ribosomal RNA necessary for mitochondrial protein synthesis and peptides involved in cellular oxidative phosphorylation. As the acrosome of the sperm is lost during fertilization, the inheritance of mitochondrial chromosomes is exclusively through the female ova, which gives rise to a characteristic form of cytoplasmic inheritance with the following features: there is no transmission from males to their offspring, although both males and females can receive the defective gene from their mother; if all the mitochondria carried the gene, then all offspring of the affected woman would also be affected. However, the precise pattern of the disease may not be recognized if only a proportion of the mitochondrial genes are affected. More recently, further challenge to dogma is through the discovery of biparental inheritance of mitochondrial gene mutations, although the molecular mechanism is still to be revealed (Luo, 2018).

Multifactorial inheritance

There are many conditions that have a familial incidence, yet the incidence of affected siblings is less than can be accounted for by unifactorial Mendelian inheritance. This is because the clinical presentation is influenced by several genes and by environmental factors. In multifactorial inheritance both genetic and environmental factors combine in varying proportions in different individuals to permit the disease to be manifest. Each trait is determined by the interaction of a number of genes at different loci. Multifactorial inheritance may be *continuous* or *discontinuous*. In discontinuous multifactorial traits, the risk within affected families is above that in the general population but is low compared with that of single-gene defects. In continuous traits, there is a range of disease expression, with individuals falling between two extremes (for example, blood pressure).

Multifactorial discontinuous traits have an increased incidence in affected families compared with that in the general population. If multifactorial inheritance is suspected, twin concordance and family correlation studies are undertaken. For instance, in multifactorial inheritance the incidence of the trait in monozygotic twins will exceed that in dizygotic twins. Some discontinuous multifactorial

Table 3.1 **HLA-linked disease associations**

Disease	HLA allele	Relative risk
Ankylosing spondylitis	B27	90–100
Acute anterior uveitis	B25	8
Reiter syndrome	B27	40
Behçet syndrome	B51	4–10
Birdshot chorioretinopathy	A29	50–224
Intermediate uveitis	DR15	6
Sympathetic ophthalmia	DR4	14
Vogt–Koyanagi–Harada syndrome	DR4	12

traits will also show unequal sex distribution, indicating that for one of the sexes the threshold for disease penetrance is higher than for the other sex, i.e. a higher proportion of underactive genes is required before the disease is expressed.

Disease associations

One approach to identifying disease associations with a specific genotype is to examine major genetic polymorphisms in a population and search for associations. Strong associations have been obtained between ankylosing spondylitis and *HLA-B27* and between birdshot chorioretinopathy and *HLA-A29*. In practice it is perhaps the absence of such allelic correlations that is more useful in predicting disease occurrence. Disease associations can be expressed as relative risk by comparing the incidence of the allele within a control population with allelic expression in patients. Table 3.1 gives a list of some of the commoner associations. Currently, high-resolution typing that characterizes the nucleic acid sequence of major histocompatibility complex alleles has meant that, for example, classic DR4 associations are now further subtyped, e.g. the association of *DRB1 *0404* with sympathetic ophthalmia and Vogt–Koyanagi–Harada syndrome.

Understanding and accounting for genetic polymorphisms association with disease have further progressed with the ability to undertake whole genome sequencing (see later section) and study of populations to identify potential causality (see later section; Mendelian randomization).

Genetic polymorphisms

Within the human genome many gene loci possess commonly occurring alleles, which can be used to distinguish populations into discrete phenotypes. This was first described for components of human blood, particularly ABO and MNS phenotypes, where allelic differences in the former determine terminal sugar residues on membrane glycoproteins and those in the latter determine the amino acid sequence of membrane-bound glycoproteins. This is also true for the Rhesus system where polymorphisms exist for the Rhesus polypeptide on the red blood cell membrane. Genetic polymorphisms are defined as the occurrence of multiple alleles at a locus, where at least two alleles occur with a frequency greater than 1%. The existence of different versions of the same genetic material in different people has prompted their use as genetic markers. This is particularly the case for human enzymes and other proteins in which polymorphisms are detectable in population groups. The use of polymorphisms in genetics will be highlighted later when we discuss the types and uses of molecular polymorphisms. To summarize at this point, their practical use in medical genetics is seen for mapping genes to individual chromosomes by linkage analysis (see following section), presymptomatic and prenatal diagnosis of genetic disease, evaluation of high-risk and low-risk persons with a predisposition to common adult disorders (e.g. diabetes mellitus) and tissue typing for both tissue and organ transplantation.

Population genetics

Population genetics is the study of the distribution of genes within populations and how these gene frequencies are changed or maintained. This form of study is important for the understanding of *linkage analysis* (linked genes that have their loci within measurable distances of one another on the same chromosome) and *linkage disequilibrium* (the association of two linked alleles more frequently than would be expected by chance), which is described later.

With our increased understanding of the regulation of gene expression that we have discussed previously, it is now appreciated that gene expression and cell function (differentiation) is influenced by *epigenetics*.

Gene frequency

When contemplating the genetic makeup of whole populations, in particular when relating the frequency of people who are heterozygotes at a particular locus to the frequency of homozygotes, the *Hardy–Weinberg equilibrium* is employed. This calculation is based on the assumption that in a given population where random mating occurs, in the absence of mutation and selection, the genetic constitution remains the same from one generation to the next.

It is important to remember that genes are paired in alleles, so that within the Hardy–Weinberg equilibrium, p is the proportion of normal alleles and q is the proportion of abnormal alleles, so that $p + q = 1$. Since a pair of alleles occurs at each gene locus, the relative proportions of normal homozygotes, heterozygotes and abnormal homozygotes are given by the equation:

$$p^2 + 2pq + q^2 = 1.$$

A few important points can be derived from this equation.

- Although the gene frequency determines the chance of any one chromosome carrying the abnormal allele, individuals have two alleles for each gene, which thus doubles the chance of having the abnormal allele on one or other chromosome.
- The heterozygote frequency is often twice the gene frequency.
- In autosomal recessive inherited disorders, the gene frequency is the square root of the abnormal homozygote frequency.

It is also important to note that the Hardy–Weinberg equilibrium is concerned only with alleles at a single gene locus; in practical terms the relationship of gene loci to each other along the length of the chromosomes is of greater importance.

Genetic linkage and linkage analysis

Two genes may be transmitted more frequently than independent assortment would suggest.

During meiosis, a crossing over between homologous chromosomes may occur at any point along the length of the chromosome. Therefore, the closer two genes are to each other, the more likely they are to be transmitted together. Gene loci on the same chromosome are therefore said to be linked when their alleles do not show independent segregation during meiosis (i.e. their loci are within measurable distance of one another on the same chromosome). Although genes may be mapped on the chromosome by techniques such as in situ hybridization (see following section), genetic linkage may also be shown by family studies, confirming a genetic background to the disorder. In a family linkage study, two loci are considered: one for the disease, the other for the trait.

One way of performing linkage analysis is to study family members who are heterozygous at each of the gene loci. One must first establish the linkage phase, determining which allele at locus 1 to be studied is on the same chromosome (of a pair) as a particular allele at locus 2 (marker locus). The family must of course be informative for the loci being considered (i.e. express the phenotype). The loci most informative for linkage analysis are those that are highly polymorphic and thus heterozygous in a large proportion of the population. The probability of meiotic crossing over between them is in effect a means of describing the distance between linked loci. Gene loci that are sufficiently well separated on the same chromosome will always have several crossovers between them and thus exhibit a 50% recombination rate (as if they were situated on different chromosomes and segregating independently). To establish the probability of two gene loci being linked, the number of opportunities for recombination and the proportion in which they have occurred are counted. This may be given as a ratio of the probability of observing the data in the sibship, based on the assumption that there is no linkage between the two loci. As large sibships are required to assess probability of linkage, it is calculated by computer and given as a logarithmic ratio, known as the Lod score (derived from log odds). In most situations there is significant linkage if the Lod score is greater than 3 and linkage can be ruled out if it reaches −2.

Linkage equilibrium and disequilibrium

Linkage equilibrium exists when, in the presence of random mating and given enough generations, the various combinations of alleles that could occur are expressed with equal proportion. *Linkage disequilibrium*, on the other hand, describes the association of two alleles more frequently than would occur by chance. This linkage may arise from mutation of one allele that has not yet established equilibrium because of a selective advantage or disadvantage. The same principles of equilibrium and disequilibrium can apply to alleles of any linked gene loci. At this point it should be emphasized that close linkage between a disease-specific gene locus and a gene showing considerable polymorphism does not necessarily mean association between the two particular alleles of the gene loci. The association between disease and human leucocyte antigens (HLA), for example, arises as a result of both genetic linkage between *HLA* allele and the disease susceptibility gene and linkage disequilibrium involving the particular allele at the *HLA* locus. Linkage disequilibrium is important clinically as a means of identifying disease-causing genes and identifying the origin and spread of mutations. As mentioned before, genetic polymorphism is the occurrence in the same population of two or more discontinuous traits at a frequency where the rarest trait would not be maintained by recurrent mutation alone. The most commonly accounted genetic polymorphisms exist within blood groups (ABO) and cell surface antigens (HLA) (see Chapter 7, page) but more recently, with the advent of molecular biological techniques, DNA sequences of

Box 3.7 **Epigenetics**

Genetic imprinting on offspring results from heritable patterns and is different from both mother and father.
Epigenetic regulation of gene expression and cancer may result from epigenetic carcinogens or alteration in DNA methylation patterns (including DNA repair mechanisms–BRCA1 and breast cancer).
Cancer therapy is currently identifying histone deacetylases or methyltransferases that, as therapies, can alter the cell differentiation and transcriptome profile of malignant cells.
Study of the epigenome is currently ongoing in understanding of methylation status in tumours. We are able to interrogate changes by assays such as:

- chromatin immunoprecipitation sequencing (CHiP-Seq)
- RNA-Seq (identifying RNA expression or coding levels and can include non-coding siRNA).

genes have shown that many are polymorphic (see later section).

Mendelian randomization identifies causal modifiable environmental risk to disease

Studying the influence of environmental risk factors influenced by gene mutations for complex multifactorial disorders such as cardiovascular disease to move from just association (as may be determined through genome-wide association studies, GWAS) towards causality involves *Mendelian randomization*. Mendelian randomization uses known variation in genes of known function to examine the causal effect of a modifiable exposure on disease. The studies rely on data from genetic association studies, as misleading conclusions on causality can occur because of linkage disequilibrium, genetic heterogeneity, pleiotropy or population stratification. For example, to test for supposed causality, the method takes common genetic polymorphisms which have indirect biological consequences on environmental exposure patterns (e.g. nicotine addiction) or consequences generated from modifiable risks (e.g. raised blood pressure).

Epigenetics (Box 3.7)

The term *epigenetics* is now largely understood as a heritable phenotype resulting from changes in a chromosome without alterations in DNA sequence. Mechanisms that give rise to epigenetic inheritance in which a stably inherited cell memory exists and altered gene expression without altering DNA sequence include: DNA methylation and chromatin remodelling, RNA-binding proteins and miRNA (as discussed previously).

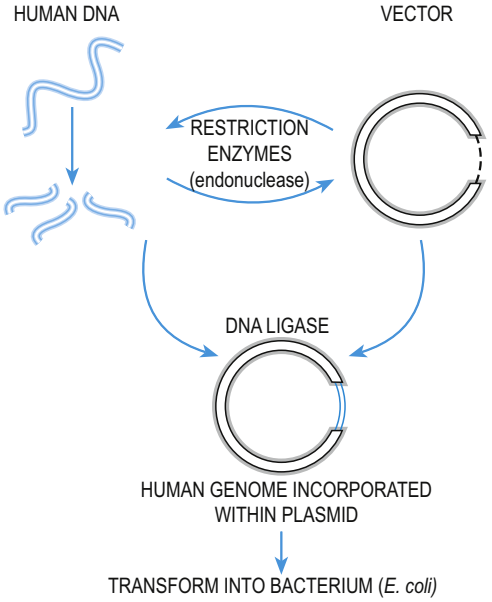


Fig. 3.5 Recombinant DNA technology: cloning.

Understanding the human genome: DNA analysis

The advent of modern molecular biological techniques has benefited many branches of medicine, including ophthalmology. In particular, the techniques described later have increased our fundamental understanding of such conditions as inherited retinal degenerations, X-linked ophthalmic disorders, retinoblastoma and Leber's hereditary optic neuropathy. Some of the techniques used to examine genes and genetic linkage are shown, with examples of how this has helped in the identification and understanding of the molecular and genetic basis of ophthalmic diseases.

Cloning

Cloning (Fig. 3.5) is an in vivo technique that produces identical copies of DNA sequences, which can then be used as gene probes. Cloning is achieved by inserting the specific fragment of DNA to be studied into another DNA molecule, which can then be replicated quickly in bacteria (recombination). Fragments of DNA can consistently be obtained by cleaving the DNA with naturally occurring enzymes, known as *restriction endonucleases*. These restriction enzymes cut DNA at specific recognition sites, usually into lengths of four to six nucleotide base sequences. Reproducing large

quantities of the same DNA fragment (i.e. cloning) can, however, be performed only by transferring the genetic material into a bacterium via a vector. The common vectors are bacteriophages or modified bacterial plasmids (extrachromosomal, self-replicating, circular DNA). The choice of vector is dependent on the size of DNA fragment to be cloned: bacteriophages have a capacity of 5–20 kilobases (kb) and plasmids a smaller capacity of up to 10 kb. These vectors replicate their own DNA independently of the host DNA and also have single restriction enzyme sites, which allow insertion of recombinant molecules. After slicing the DNA molecule, the recombinant DNA is prepared by incubation of the plasmid with the fragment in the presence of an enzyme, *DNA ligase*, which ligates or binds the DNA to the cut ends of the plasmid. To date, over 200 restriction enzyme sites have been recognized. Restriction enzymes are named according to the organism of origin, e.g. E-cor is derived from *Escherichia coli*.

Refining technology: identifying and manipulating the gene

DNA polymorphisms

More recent molecular biological techniques (polymerase chain reaction, whole genome sequencing, exome sequencing) to detect single nucleotide polymorphisms have largely replaced the need for traditional techniques such as restriction fragment length polymorphisms (RFLP) and Southern blotting although variable number of tandem repeats (VNTR) polymorphism is still used for DNA fingerprinting with PCR technology.

information on restriction fragment length polymorphism and Southern blotting is available online at https://expertconsult.inkling.com

Polymerase chain reaction

The polymerase chain reaction (PCR) is a technique that allows small amounts of DNA to be amplified for detection and analysis. PCR requires two flanking sequences to the sequence that is to be amplified. DNA primers are formed from short sequences of DNA with these flanking sequences. Amplification of the DNA sequence uses heat-resistant DNA polymerase to extend the short DNA primers of about 20 base pairs, which are hybridized at either end. Cycles of denaturation with heat and polymerization with cooling occur up to 50 times. As a result, the target DNA is amplified 10^5–10^6 times. This process, which is largely automated, is proving a great asset in detecting specific DNA where the amount of starting material is very small, e.g. neonatal diagnosis and the identification of infective organisms in tissue samples, and as a research tool. However, quality control of false-positive rates must be established before its full clinical efficacy can be ascertained.

Genomics, transcriptomics and proteomics

The genome contains genes that, as we have already described in this chapter, are transcribed to produce RNA and ultimately generate proteins. What has been realized since the findings of the Human Genome Project is that there are very many more proteins in the *human proteome* than there are genes in the *human genome* (*c.* 400,000 proteins and 22,000 genes). The study of the *genome* (*genomics*) has been made possible by the development of *gene arrays* (Fig. 3.6).These are DNA microarrays (small solid platform devices with thousands of DNA sequences corresponding to genes, immobilized on a variety of surfaces—nylon membranes or glass) which give us the ability to look at many genes at one time. As such, we can get a picture of the possible interactions among thousands of genes. The array is an orderly arrangement of gene-specific DNA that is hybridized to mixtures of labelled RNA in samples. Therefore, based on the amount of signal from successful hybridization, we can determine the relative expression of the corresponding gene. With such a capability we can now compare the extent of gene expression within a sample or of specific genes between samples (see Box 3.8). For population studies with the advent of high-throughput sequencing, programs of work have undertaken genome-wide association studies (GWAS) and now increasingly commonplace for diagnosis of inherited disorders. This platform allows for examination of genetics variants from study of single nucleotide polymorphisms in major diseases to be undertaken on a large scale (Box 3.9). Currently there is also exome sequencing, a process where sequencing of coding regions of the genome can be targeted instead of whole genome sequencing. All this has become more accessible with the advent of next-generation sequencing (NGS) technology and ability to study DNA and RNA sequences (RNA seq).

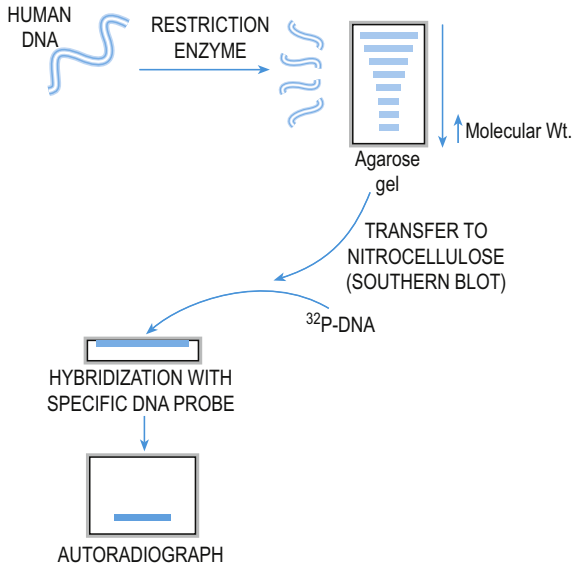
The genome is only a source of information and the initial step to provide proteins involves the transcription to RNA, where the *transcriptome* is the complete set of RNA transcripts. *Transcriptomics* is the ability to study the dynamic *transcriptome*, which varies depending upon the cell, the stage of cellular development and activation and environmental influences. *Transcriptomics* allows us to investigate the control of RNA transcripts by activating or inactivating transcription factors that control RNA transcription and looking at gene expression patterns under various conditions or

Restriction fragment length polymorphism (RFLP)

RFLP occurs because of point mutations in the human genome, which normally occur in the non-coding regions of the DNA molecule. As a result of these acquired polymorphisms, recognition sites for the restriction endonucleases are created or abolished. The consequence is that restriction enzymes will produce variable lengths of DNA fragments, which give characteristic banding on Southern blotting (see eFig 3.3). If an individual is heterozygous for an RFLP, one restriction band on the electrophoretic gel corresponds to one chromosome and the other band to the other chromosome of the pair. This allows one to track the transmission of a single chromosome region through the family and perform classic linkage studies. The discovery of DNA polymorphisms has greatly increased the extent to which individual copies of particular genes are recognized as being truly unique. In fact, DNA polymorphisms are simply the molecular manifestations of variations in the genome, which have been recognized for a long time from studies of protein polymorphisms.

Variable number of tandem repeats (VNTR) polymorphism

Also within the genome are many tandem repeats of sequences that vary for each individual. The sequences act as spacers between restriction enzyme sites. VNTR polymorphisms are short sequences of DNA and the distance between the VNTR depends on the number of repeats. VNTR are heterozygous and can be detected using probes for the core sequence that gives rise to these repeats. VNTR are highly polymorphic and almost unique for each individual (genetic fingerprinting). VNTR markers are highly informative for genetic linkage analysis and for individual identification (have been used prior to gene sequencing for paternity and forensic testing). Within the genome there are several classes of repetitive DNA, whose nucleotide sequence is repeated several hundred times within the genome. These tandem repeats are again highly polymorphic and can be used for genetic linkage studies.



eFig. 3.3 Recombinant DNA technology: Southern blotting.

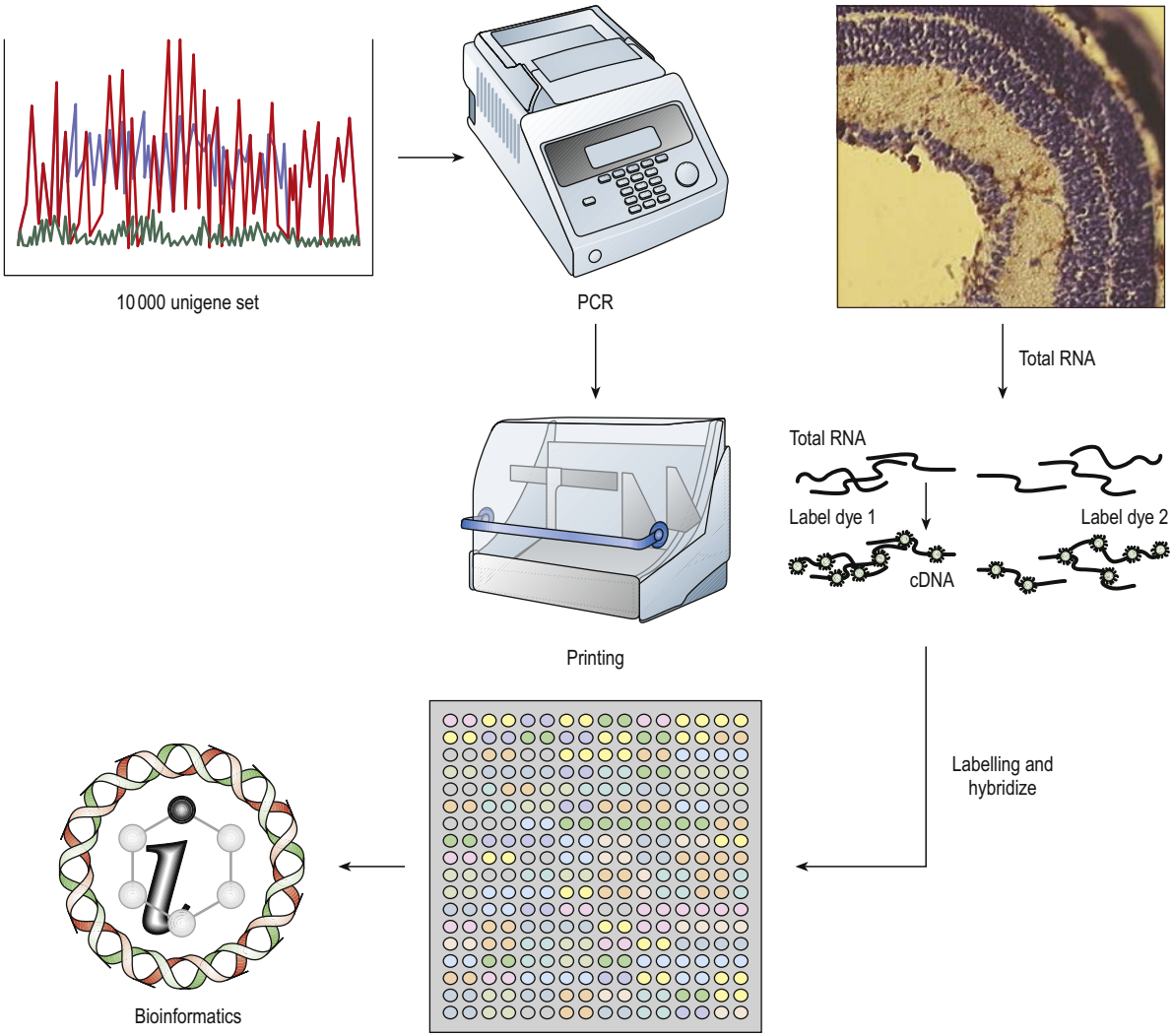


Fig. 3.6 Gene array and bioinformatics. The array hybridization generates many thousands of computational parameters describing upregulation and downregulation of genes that have been targeted and require quantification. (Courtesy UOB Transcriptomics Facility and Dr Chungui Lu, School of Biological Sciences, University of Bristol, UK.)

Box 3.8 **Gene arrays (Affymettrix GeneChips®)**

- Isolate RNA from sample or samples to be compared
- Convert to DNA with reverse transcriptase–complementary or cDNA
- Hybridize the labelled cDNA (e.g. ^{32}P) to arrays with fixed DNA sequences (spot sizes on plate of 300 μm)
- Remove unhybridized cDNA
- Detect and quantify hybridized cDNA.

comparing normal with pathology. The dynamics of the system are further exemplified by the fact that there are very many more proteins than genes. Also, ultimately, cell and tissue behaviour are defined by protein interactions both intracellularly and extracellularly. *Proteomics* studies these dynamic consequences of gene expression. The study relies on an ability to separate and to identify proteins by, for example, mass spectrometry, gel quantification (two-dimensional electrophoresis) and protein sequence analysis. *Structural proteomics* determines

Box 3.9 **GWAS has illuminated susceptibility of disease, potential causal relationships and targets for therapies***

- Polymorphisms of the complement gene *CFH* have a high association in age-related macular degeneration (AMD)
  GWAS has generated multiple gene associations with AMD with highest risk in polymorphisms in chromosome 1 and 10.
- IL-23R-IL12RB2 and IL-10 linked as Behçet disease susceptibility loci
- Multiple sclerosis risk following anti-TNF therapy mirrors TNFR1 genetic variant
- Genome-wide meta-analysis identified new associated genetic susceptibility loci for refractive error and myopia and glaucoma.

* See www.genome.gov

three-dimensional structure by X-ray crystallography and nuclear magnetic resonance spectroscopy. How proteins interact can be assessed by affinity chromatography and fluorescent resonance energy transfer (FRET). Finally, given the diversity of proteins observed but not accounted for at the gene level, their post-translational modifications can be studied by looking at the extent of phosphorylation or glycosylation (Fig. 3.7)

Finding and tracking the gene through families

The use of restriction endonuclease mapping (gene mapping) is a tool for analysing genetic disease. DNA obtained from tissues or cells (usually peripheral blood leucocytes) is treated with restriction enzymes and analysed by Southern blotting using a radioactively labelled gene probe. By using different restriction enzymes and orientation of the fragments, it is possible to build up restriction enzyme maps of sections of the human genome. Any normal or abnormal gene for which a specific probe has been generated can be analysed by this technique. Potential probes are chosen from a central library and screened for their ability to detect RFLPs by hybridizing them to Southern blots of DNA from a panel of unrelated individuals and by searching for restriction enzyme sites that give variable patterns. A disease gene can be mapped by collecting large pedigrees and comparing the segregation of the disease with the segregation of the RFLP. Once linkage has been clearly established, the probe is assumed to map to a chromosome region close to the disease gene and can be used to track the gene through families.

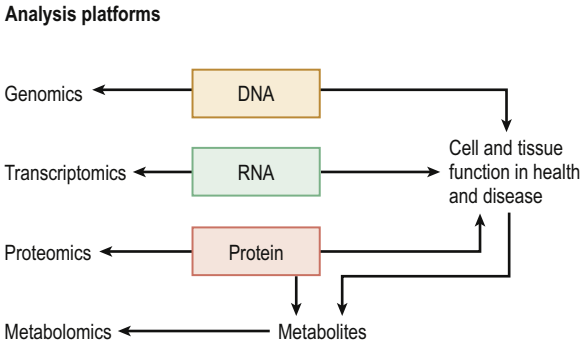


Fig. 3.7 The era of the 'OMICS' analysis. Analysis platforms for detection of gene defects and relation to cell and tissue functional consequences.

Genetic analysis has largely been superseded by techniques such as high-throughput genetic sequencing and exome analysis. There is a high demand for high-throughput sequencing (next-generation sequencing) that generates hundreds of thousands of sequences at once and is increasingly cost-effective. *Exome* sequencing is an efficient strategy that sequences the coding regions, although there is a risk of missing possible disease-causing mutations because changes in coding regions are estimated to account for around 85% of disease-causing mutations.

Molecular biology and clinical medicine

Gene probes

Gene probes can be produced in several ways and fall broadly into three types: gene-specific probes, oligonucleotide probes and polymorphic probes. Gene-specific probes are produced from specific mRNA by the enzyme reverse transcriptase, which synthesizes a complementary DNA copy (cDNA) from mRNA. If radioactive bases are added to the reaction mixture, the cDNA will be labelled and can thus be used as a hybridization probe to look for the complementary sequences. Probes can be used in *dot–blot hybridization*, where serial dilutions of DNA samples are held on DNA-binding membranes and the complementary radioactively labelled probes are hybridized in vitro, so that the amount of radioactive signal is proportional to the amount of target DNA present. The cDNA can also be cloned by synthesizing a second DNA strand from cDNA using a bacterial DNA polymerase, which is incorporated into a plasmid and grown in bacterial cells. These oligonucleotide probes recognize short sequences of DNA that

Box 3.10 **Gene regulation of apoptosis**

- p53: tumour suppressor protein which functions to activate DNA fragmentation and apoptosis, thus regulating tumour development
- *Bcl-2*: gene for family of proteins which *inhibit* apoptosis
- *BAX*: gene for family of proteins which *activate* apoptosis
- Fas and TNFRI: death domain-containing proteins which when cross-linked with ligand (Fas ligand and tumour necrosis factor-α, respectively) induce apoptosis.

correspond to the sequence known to occur in the gene. With a probe of this length, a single mismatched base pair is sufficient to impair hybridization and can be used to detect changes to a single base (point mutations). Similarly, probes can be developed that recognize the various DNA polymorphisms within the non-coding sequences, for example, RFLP and VNTR. DNA probes can also be used to identify abnormal genes or gene products at the molecular level within cell cytoplasm or nucleus, by *in situ hybridization*. This technique uses labelled DNA or mRNA probes, which hybridize to the expressed genes in the cell in a manner similar to that used for immunohistochemistry. In situ hybridization can thus establish whether the genomic material of interest is present in the DNA of the cell in vitro.

Molecular and cell biology: controlling cell destiny

Cell death can occur by either necrosis or apoptosis (programmed cell death). Apoptosis is a process whereby developmental or exogenous environmental signals trigger specific intracellular genes, which results in cell death. Ligation of cell surface receptors such as Fas ligand (Box 3.10), which is commonly associated with death domains that signal and stimulate caspases, disrupts mitochondrial membrane channel permeability (see also Chapter 7 page). Apoptosis is essential for normal development and many of the genes that control apoptosis (Box 3.10) have been highly conserved throughout evolution. Histologically, apoptosis is associated with chromatin condensation and nuclear DNA fragmentation. Only finally when caspase enzyme activation has occurred is membrane integrity affected and the cell eliminated via phagocytosis and the reticuloendothelial system without the secondary inflammatory response that occurs when cells undergo necrosis. Understanding control of cell homeostasis—or where it goes wrong—is important in understanding disease pathogenesis and cancers and the development of therapies (Box 3.11).

Box 3.11 **Cell destiny and function, mutations and drug development**

Some cutaneous melanomas have an activating mutation in the serine-threonine protein kinase *B-RAF* oncogene (V600E mutation), which regulates the MAP kinase pathway that controls cell proliferation. By inhibiting BRAF with new-targeted drugs inhibiting the enzyme B-RAF (vemurafenib) benefits in outcomes of stage IV melanomas are now possible.

In cystic fibrosis there is abnormal or complete loss of function of cystic fibrosis transmembrane conductance regulator gene (*CFTR*). CFTR is an ABC transporter-class ion channel regulating the constituents of sweat, mucus and digestive fluids. In a small number of cytstic fibrosis patients (5%) there is a specific mutation in the *CFTR* gene (G551D) that causes impaired ion transport but CFTR remains expressed on epithelial surfaces. Drugs have been developed (ivacaftor) to restore or enhance function to the CFTR in this targeted group of patients, improving their outcome.

Further information on role of the Ubiquitin-proteosome pathway in homeostasis is available online at https://expertconsult.inkling.com

Using molecular biological technology to detect apoptosis

Observed and therefore measurable changes in apoptotic cells include intranucleosomal cleavage of DNA, which generates both monomers and multimers of DNA that can be detected by DNA electrophoresis (DNA laddering). Early changes can be detected by labelling DNA fragments at the 3′ OH ends with terminal deoxynucleotidyl transferase either immunohistochemically (TUNEL staining) or by flow cytometric analysis. Changes in apoptotic cell membranes can be detected by increased flip-flopping of the plasma membrane and exposure of phosphatidylserine residues on its outer leaflet. Phosphatidylserine is detected by annexin V via either in situ histochemistry or flow cytometry. Apoptosis ultimately increases intracellular caspase activity and cell lysates can be tested for levels of caspase enzymatic activity by simple fluorometric analysis.

Gene therapy

Gene therapy is the transfer of selected genes into a host with the intention of alleviating or curing disease. There are many gene therapy strategies that can be used; which one to choose depends upon the pathogenesis.

Ubiquitin (Ubq)–proteosome pathway maintains cellular homeostasis

Proteosomes are nuclear and cytosolic protease complexes that degrade proteins, which become covalently linked to Ubq via a cascade of enzymatic reactions (Ubq 1–3). Ubq is a highly conserved 76-amino-acid protein that has been implicated in the pathogenesis of genetic disease and malignancies. It is integral in the destruction of phosphorylated cyclins and cyclin inhibitors, such as p27, an inducible inhibitor of cyclin-dependent kinase activity (see Box 3.2); transcriptional proteins such as STAT 1; and p53, a tumour suppressor gene product. Viruses (e.g. human papillomavirus) can upregulate p53 degradation in the Ubq–proteosome pathway, a process implicated in tumour formation. Proteosome inhibitors such as lactacystin can arrest cycle–cycle progression and therefore offer potential as chemotherapeutic agents.

Box 3.12 **Vectors**

Viral vectors

Retroviruses e.g. lentiviruses (human immunodeficiency virus; simian immunodeficiency virus). Although RNA viruses, they possess reverse transcriptase and can therefore generate complementary DNA.

Adenoviruses have been engineered to remove replicatory ability and also antigenic coats, particularly to reduce the immune response to these vectors.

Adeno-associated viruses (AAV): a single strand of DNA virus without viral genes and therefore reduced immunogenicity. They only accommodate smaller DNA inserts but do establish long-term expression.

Non-viral vectors

Liposomes: spherical lipid bilayer vesicles. No limit to size of DNA to be carried but efficiency of transfer is low.

- *Gene augmentation therapy* For diseases caused by loss of gene function and which will supply more copies of normal gene in the hope of restoring normal phenotype, e.g. cystic fibrosis, haemophilia, severe combined immunodeficiency syndrome (SCID), retinitis pigmentosa.
- *Targeted killing of specific cells* Genes are directed to specific cell types and incorporated into the genome, expressed so that protein interferes with cell cycling and survival, thus killing cells.
- *Targeted mutation correction* Can be performed using *ribozymes* (which cleave and repair mRNA), *triple helix oligonucleotides* (block gene transcription) and *anti-sense oligonucleotides* (that block mRNA translation).
- *Targeting inhibition of gene expression* When diseases display novel gene products or excessive expression of gene product, then blocking at a single gene level (either DNA or RNA) or blocking protein can be possible to attain specific inhibition of expression.

Achieving gene therapy depends upon the size of the DNA fragments to be transferred (*transfection/transduction*). Techniques can include injection of naked DNA, which is not limited by size or number of genes but is inefficient (*gene gun*). Often, therefore, the gene to be transferred is not conventional and the coding DNA is engineered to be flanked by regulatory sequences to ensure high expression dependent upon the tissue. By whatever method, the aim is to insert the DNA into the chromosome of the cell. To assist entry of DNA into the cell, *vectors* are used (Box 3.12). These are currently attenuated viruses that have been engineered so that the replication and disease-causing components have been removed. Even with this manipulation they retain an efficient ability to enter cells—so-called *transduction* (Box 3.13).

Box 3.13 **Not all gene therapy requires replacement of the gene: *optogenetics***

Recent developments to try and restore vision in the degenerate retina uses viral transduction of, for example, channel rhodopsins to excite neurones, creating a rudimentary but significant and potential functional non-photoreceptor voltage-sensitive system to create artificial vision. Such systems could be coupled with a functional prosthetic (optobionics). Additionally, the use of optogenetic approaches is very useful to determine disease mechanisms in models, including understanding of circadian rhythms and non-visual transducing photopigments.

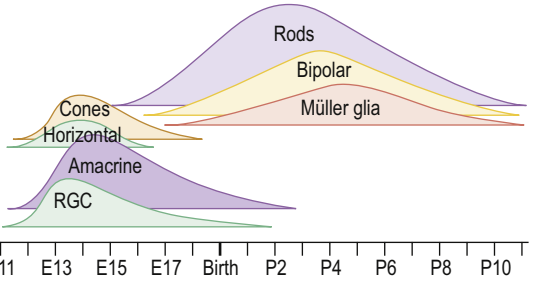


Fig. 3.8 Two waves of retinal cell-type differentiation. The curves depict the relative number of retinal progenitors that exit the cell cycle and commit to a specific fate over time. Retinal ganglion cells *(RGCs)* appear at E11 and peak at E13.5, as do amacrine cells, horizontal cells and cone photoreceptor cells. The other cell types, bipolar cells, Müller glia and rod photoreceptor cells, appear later as indicated. (Adapted from Cepko et-al., 1996, and Wang et-al., 2002.)

Genes, cell differentiation and cell-based therapies

Stem cells have the capacity to self-renew, proliferate and have no limitation to their potential differentiation. Progenitor cells can divide but have restricted differentiation potential. In the adult vertebrate central nervous system, neural progenitor cells (NPC) have been identified and shown to generate neurones and glia. Developments from molecular and cellular understanding of cell development in addition to understanding the control of gene expression during cell differentiation and tissue organization in rodents have opened opportunities to develop cell-based therapies for degenerative disorders. Cepko's fate mapping of retinal cells in rodents determined the timing of differentiation from which many researchers since have ascribed specificities of genes, transcription factors and growth factors for cell fate determination (Fig. 3.8). During

development there are two waves of progenitor cells that exit their cell cycle and commit to their post-mitotic mature cell fate. Together with increasing understanding of pivotal gene expression during cell differentiation and retinal development, this allows recapitulation experimentally to generate cell sources for cell transplantation.

How do you generate cells for retinal reconstruction? The opportunities come from several sources. First, embryonic stem cells (ES cells), which are derived from the inner cell mass of the blastocyst, can be conditioned in culture to generate stem cells that can be driven toward neural differentiation and then photoreceptors. Second, the recent Nobel Prize-winning creation of inducible pluripotential stem cells (iPS cells) from non-pluriopotential cells (including adult somatic cells such as fibroblasts) creates great possibilities to derive cells from the host cells (Fig. 3.9).

However, the developing and the postnatal vertebrate animal retina contain NPC, which divide, generate *neurospheres* and undergo neuronal and glial differentiation. In humans, central nervous system-derived NPC have been successfully cultured from the adult brain and more recently adult retina; NPC have also been isolated from the human retina while it is still immature and undergoing development. One phenotypic marker of stem cells and NPC is nestin, which is an intermediate filament. The adult human retina contains nestin-positive neuronal and glial cells, as does epiretinal scar tissue. This evidence suggests that NPC exist throughout life or are inducible in the human retina. Human retinal NPC are currently under extensive investigation to look at their capacity to renew damaged retina via, for example, transplantation with potentially gene-modifying cells that have undergone gene therapy before

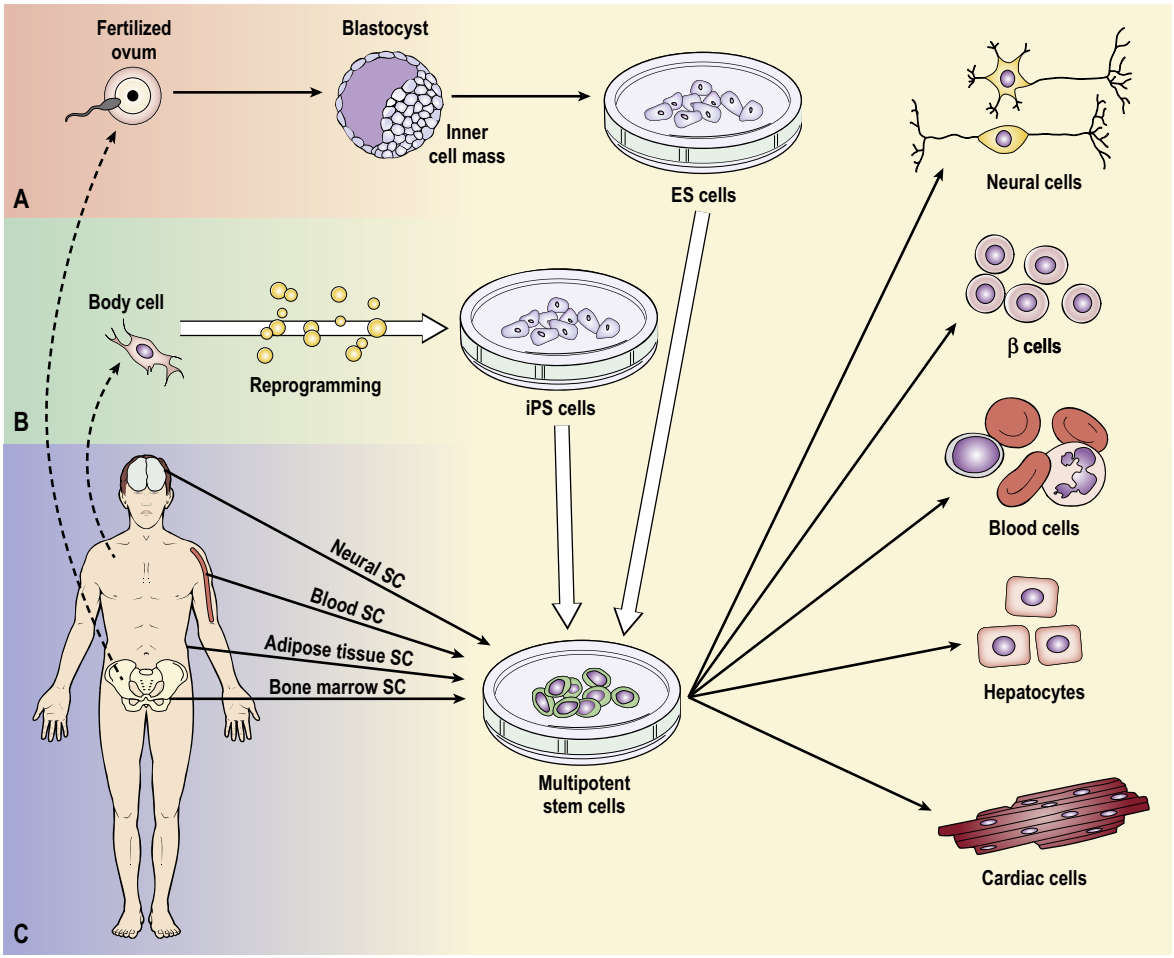


Fig. 3.9 Pluripotential stem cell sources for delivery of cell-based therapies. (A) Generation of embryonic stem (*ES*) cells. (B) Generation of inducible pluripotent stem cells (*iPS*). (C) Harvesting of lineage precursor cells. *SC,* Stem cells. (Reproduced from Power and Rasko, 2011.)

administration. To date the research has determined an ability to create retinal pigment epithelial cells from human ES cells, which have now been developed and are currently used in several clinical trials for age-related macular degeneration (see www.clinicaltrials.gov). Alongside this, there are great advances in the ability, at least experimentally, to develop ways of transplanting photoreceptor precursors that are able to integrate into the retina and develop synapses and generate action potential. Further advances have been made in developing 3-D retinal organoids delivering an in vitro system that can be applied for understanding development and pharmacological testing (see also Chapter 4,). Fig. 3.10 demonstrates how both ES cells and iPS cells can be adapted for elucidating mechanisms of disease, both complex and inherited retinopathies, and can assess suitability for tractable genomic and cellular based therapies.

With all the developments discussed, we can see that there are many exciting current and potential ways forward for the treatment of inherited retinal degeneration. Fig. 3.11 shows possible routes for intervention, depending on the stage at which treatment might be required.

Gene editing and repair

Clustered regularly interspaced short palindromic repeats (CRISPR) or CRISPR-Cas9 is fundamentally a bacterial defence mechanism to prevent against viral invasion into the bacterial genome: an evolutionary immune defence mechanism for survival. CRISPR sequences are transcribed to RNA that guide (gRNA) to matching sequences of DNA and then using the enzyme Cas-9 to cut the DNA and therefore, for example, remove inserted DNA of the virus. With modification of the Cas-9 system, CRISPR is now able to delete and to activate gene expression. CRISPR-Cas9 can be customized. Customizable forms are very useful for experimental work and can be adapted for several applications in the laboratory (see Fig. 3.12) and are being developed for potential clinical use.

RNA treatments include development of antisense oligonucleotides (ASO), which are synthetic and can restore protein expression through various methods. One most recent clinical trial for CEP290 Leber Congenital Amaurosis is to correct splicing targeted to skip the abnormal intronic site. There are other mechanisms of ASO including: (i)

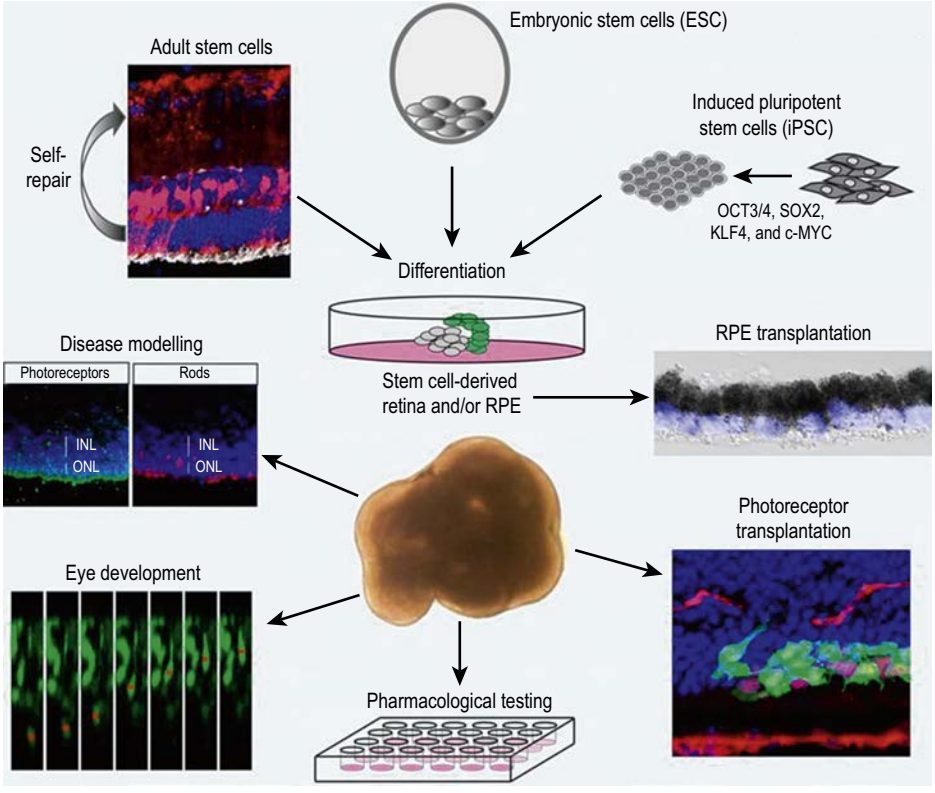


Fig. 3.10 Schematic showing how endogenous retinal stem cells, embryonic stem cells and induced pluripotent stem cells can be differentiated into retinal cells with a wide range of applications. *INL*, inner nuclear layer; *ONL,* outer nuclear layer. (Courtesy Nature Publications)

RNA-DNA hybrid to increase target mRNA degradation; (ii) block RNA binding proteins; (iii) inhibiting upstream open reading frames allowing increased translation of proteins.

Molecular genetics and ophthalmology

With the advent of modern molecular biological techniques, which have improved and will continue to improve the understanding of the pathology of disease at both the cellular and molecular levels, considerable progress has been made, in particular in the study of X-linked disorders, several of which are of interest to the ophthalmologist.

The problem of how to find and to identify an abnormal gene among the millions of genes in the human genome has already been discussed. To summarize briefly, the problem can be tackled principally by two approaches. The first is to find a gene marker in the human genome that is close to the causative gene defect. This approach, as described previously, is the basis of genetic linkage and requires the analysis of DNA from affected families. Using the genetic marker, linkage analysis will locate the gene to loci on the same chromosome. Fragments of DNA from that region of the chromosome are cloned and sequenced to identify mutations in the gene.

The second approach is to identify candidate genes specifically expressed in the tissue, or which are known to code for proteins important in that tissue. Patients are then screened for mutation in these genes. By this method the discovery of mutant genes gives insight into the underlying pathophysiology of the condition, if the functions of the proteins from the genes studied are already known. This method of analysis is known as *candidate gene analysis*. Both of these methods have been used to study ophthalmic disorders and enhanced in current studies by the new platforms of next-generation sequencing discussed previously.

The ability to detect mutations has led to several consortia for 'new gene detection' but also genetic testing. There is now increasing opportunity to develop full gene testing with NGS or a restricted panel of 235 exome panel, such as supplied through UK Genetic Testing Network (for example; see https://digital.nhs.uk/about-nhs-digital/our-work/keeping-patient-data-safe/gdpr/gdpr-register/

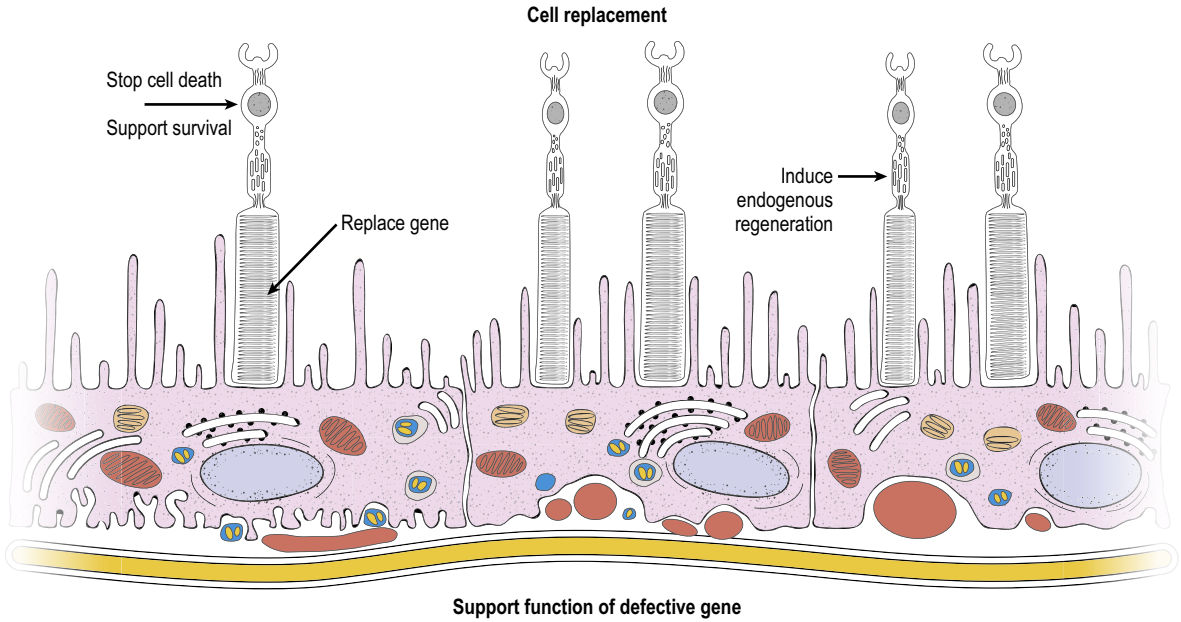


Fig. 3.11 The schematic represents possible approaches either to maintain function or to restore function in face of photoreceptor degeneration. The approach is derived from information that from early in diagnosis therapeutic support may be developed to prevent the cell death that occurs via apoptosis (see the explanation of the exponential cell death theory where the risk of cell death remains constant throughout life in Clarke et al., 2000). Similarly, it is assumed that until the cells die they receive survival support via growth factors to maintain function and therefore pharmacological approaches may support function for longer. Ultimately, gene therapy can be applied to replace the mutant, defective or lost gene. However, late in the disease process, when loss of cells via apoptosis has occurred, an approach using cell replacement via stem cells (see Fig. 3.10) or one stimulating endogenous retinal precursor cells to replace cells are being pursued. Function may be partially restored by optogenetics (using gene therapy approaches to deliver channel rhodopsins) to substitute photoreceptor function.

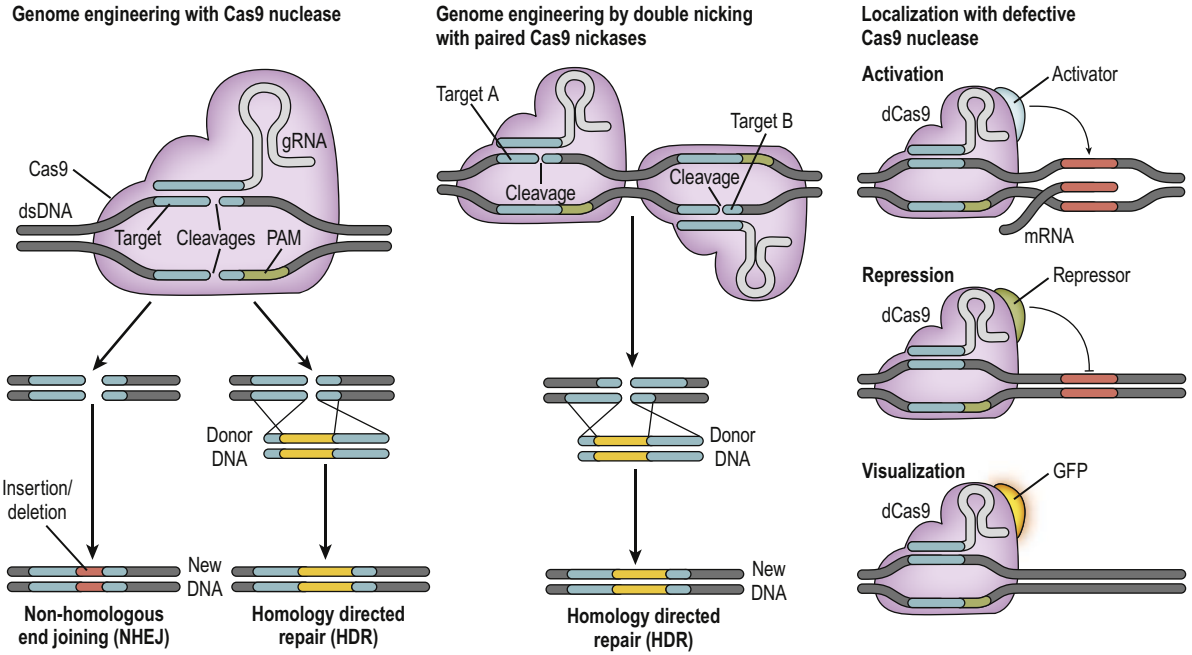


Fig. 3.12 (A) Wild-type Cas9 nuclease site specifically cleaves double-stranded DNA activating double-strand break repair machinery. In the absence of a homologous repair template, non-homologous end joining can result in indels disrupting the target sequence. Alternatively, precise mutations and knock-ins can be made by providing a homologous repair template and exploiting the homology directed repair pathway. (B) Mutated Cas9 makes a site-specific single-strand nick. Two sgRNA can be used to introduce a staggered double-stranded break which can then undergo homology directed repair. (C) Nuclease-deficient Cas9 can be fused with various effector domains allowing specific localization. For example, transcriptional activators, repressors and fluorescent proteins.

uk-genetics-testing-network-ukgtn-data). Examples of gene causing mutations are given in Table 3.2.

The opportunity with our current platforms has enabled on large populations to appreciate target genetic association in complex disease and develop opportunities with mendelian randomization to ascertain causality nor not. For example, GWAS has highlighted the high-risk alleles in macular degeneration on both chromosome 1 and 10 in addition to many genes in cholesterol pathway, for example, opening avenues of potential therapeutic targets. Data from glaucoma, myopia and others have identified novel at-risk alleles. This may help in identifying and stratifying patient risk of progression and thus directing appropriate therapy for the right patient at the right time, but also where to dive deeper into understanding mechanisms. Ultimately causality is the holy grail. If risk is present at the gene level in complex disorders, mendelian randomization may help. Using myopia as an example, it has been shown that more time in education increases the prevalence of myopia. In AMD, mendelian randomization has implicated HDL-cholesterol mechanisms. Mendelian randomization can tell us the relevance or not of gene association.

The X chromosome

Historically, many molecular genetic studies have been based on an investigation of the X chromosome and of X-linked disorders. Mapping disorders such as red–green colour blindness (Xq22–28), blue cone monochromacy (Xq28) and congenital stationary night blindness (Xp11) is possible with the use of RFLP and recombinant DNA technology.

Retinitis pigmentosa

Retinitis pigmentosa is the term used to describe a heterogeneous group of rod–cone dystrophies that have a variety of clinical appearances by virtue of varying inheritance patterns. Studies have documented the frequencies of the various modes of inheritance. Approximately 43% are inherited by autosomal transmission, 20% by autosomal recessive transmission and between 8% and 25% by X-linked recessive transmission. Figures again vary, but approximately 20%–25% of cases of retinitis pigmentosa appear to be isolated, or at least have unidentifiable patterns of inheritance. As the disease may

Table 3.2 **Ocular disease-causing genes**

Diagnoses eligible for inclusion	Genes offered for testing
Achromatopsia	CNGA3, CNGB3
Albinism	*Recessive*: TYR, OCA2, TYRP1, SLC45A2 *X-linked*: GPR143 (OA1)
Aniridia and other developmental eye anomalies	PAX6, WT1#, DCDC1#, ELP4# (# del/dup testing only)
Axenfeld–Rieger syndrome	PITX2, FOXC1
Best disease	BEST1
Bietti crystalline corneoretinal dystrophy	CYP4V2
Choroideraemia	CHM
Chronic progressive external ophthalmoplegia (CPEO)	POLG
Cone rod dystrophy	ABCA4, RPGR, CRX, GUCY2D *(codon R838)*
Congenital cranial dysinnervation diseases (CCDD)	KIF21A, CHN1, SALL4, TUBB3, HOXA1, PHO2A, ROBO3, HOXB1
Familial/congenital nystagmus (X-linked cases only)	FRMD7
Congenital stationary night blindness/Oguchi disease	GPR179, RHO, NYX, TRPM1, SAG
Corneal dystrophy	TGFBI, KRT3, KRT12
Doyne honeycomb dystrophy	EFEMP1
Familial exudative vitreal retinopathy	FZD4, LRP5, NDP, TSPAN12
Fundus albipunctatus/Bothnia retinal dystrophy	RDH5, RLBP1
Glaucoma (juvenile open angle and congenital only)	CYP1B1, OPTN, MYOC
Hermansky–Pudlak syndrome	HPS1 and HPS3
Juvenile X-linked retinoschisis	RS1
Kearns–Sayre syndrome (KSS), mitochondrial encephalopathy, lactic acidosis and stroke-like episodes (MELAS), myoclonis epilepsy associated with ragged-red fibres (MERRF), neuropathy, ataxia and retinitis pigmentosa (NARP)	Mitochondrial gene panel
Leber's hereditary optic neuropathy (LHON)	LHON panel (MT-ND4, MT-ND1, MT-ND6/mutations 11778G>A, 3460G>A, 14484T>C and 14459G>A)
Lowe syndrome	OCRL
Microphthalmia and anophthalmia	RAX, SOX2, OTX2, VSX2, STRA6 and SIX6del/dup analysis
Neurodegeneration with brain iron accumulation (NBIA)	FA2H, MMIN, PANK2, PLA2G6
Occult macular dystrophy	RP1L1 (R45W)
Optic atrophy, dominant	OPA1, OPA3
Papillorenal syndrome	PAX2
Pattern dystrophy	PRPH2
Retinitis pigmentosa (RP) and retinal degenerations	*Dominant*: (panel including RHO, PRPH2, RP1, IMPDH1, PRPF8, NR2E3, PRPF3, TOPORS, PRPF31, RP1, KLHL7, SNRPN200), CA4, CRB1, CTRP5 *X-linked*: RPGR, RP2
Retinoblastoma (12/50 enrolled for 2013)	RB1
Sorsby fundus dystrophy	TIMP3
Stargardt disease	ABCA4, ELOVL4, RDS
Stickler syndrome^ (6/50 enrolled for 2013)	COL2A1
Usher syndrome^ (limit exceeded for 2013)	Usher panel (CDH23, CLRN1, DFNB31 (WHRN), GPR98, MYO7A, PCDH15, USH1C, USH1G, USH2A)

The genes highlighted in this table emphasize the increasing number of genes that may be tested to detect or to diagnose disorders. While not a complete set and, moreover, the genes included for testing vary worldwide, the advent of exome and whole genome deep sequencing illuminates continually other disease-causing genes.
(Courtesy National Eye Institute, National Institutes of Health (NEI/NIH). Adapted from http://www.nei.nih.gov/eyegene/genes_eyegene.asp).

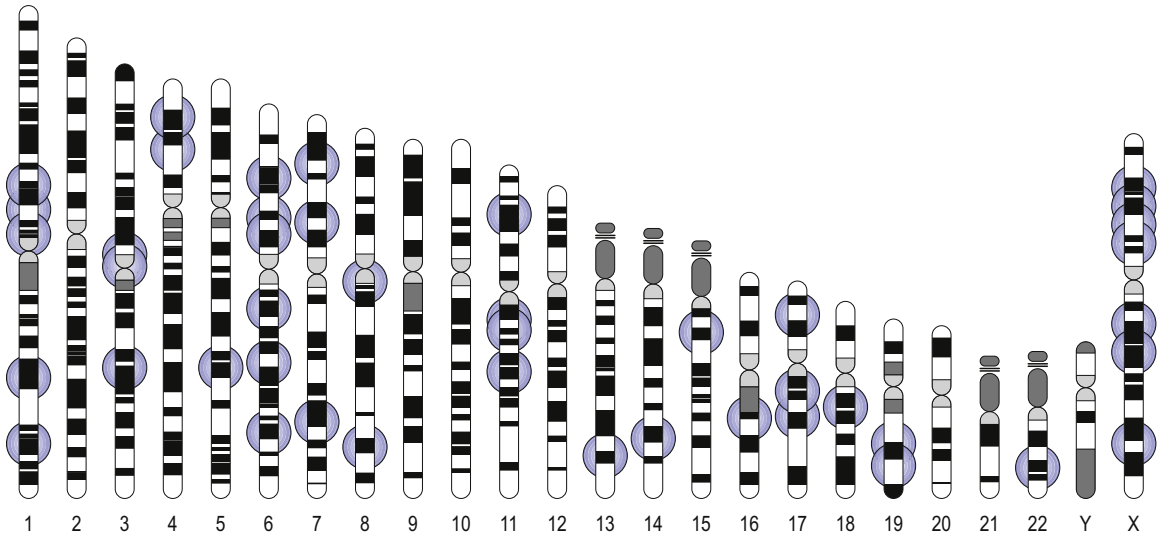


Fig. 3.13 A schematic representation of the human chromosomes, showing the various loci implicated in retinal disorders. (Figure courtesy Dr Z. Mohamed, PhD thesis, University of Aberdeen, UK.)

be classified according to Mendelian inheritance patterns, the possibility of single-gene defects is high and this has prompted an energetic search for the isolation of such genes (Fig. 3.13).

X-linked retinitis pigmentosa

Patients with this form of retinitis pigmentosa present with symptoms of night blindness from childhood; they have progressive constriction of visual fields and loss of vision in midlife, although the severity of the disease does vary. The gene for this condition has been mapped to Xp11.3 (short arm of the X chromosome). Further evidence also maps the X-linked gene to Xp21, particularly in families where the female carrier demonstrates the golden tapetoretinal reflex. These gene loci have recently been designated *RP2* (Xp11.3) and *RP3* (Xp21.1). Currently, probes are available for identifying both loci and may be used for prenatal diagnosis and genetic counselling.

Autosomal dominant retinitis pigmentosa

Several mutations have been found in candidate genes in up to 30% of patients with autosomal dominant retinitis pigmentosa. Mutations in two genes have been studied in particular. These are the rhodopsin gene on chromosome 3q (accounting for 20% of all cases) and the peripherin gene on chromosome 6p. The rhodopsin molecule, composed of 348 amino acids, exists as a seven-loop transmembrane protein in the rod outer segment. The C-terminus of the protein is in the cytoplasm and the N-terminus of rhodopsin is in the intradiscal space. Throughout the protein, several regions are affected by mutations, which fall into three main groups: (1) mutations affecting amino acids in the intradiscal space; (2) mutations affecting amino acids in the transmembrane domain; and (3) mutations affecting amino acids in the cytoplasm. Most of these mutations probably destroy the three-dimensional (tertiary) conformation of the protein and in some way affect protein function. To date, over 150 different mutations have been reported, the majority of which are point mutations, although deletions have also been discovered.

The peripherin gene codes for the retinal degeneration show (RDS) protein found in rodents (see Chapter 9, p. 519), which is a component of the rod outer segment disk membranes. More than 20 mutations have been discovered in this gene in association with autosomal dominant retinitis pigmentosa and other retinopathies, for example retinitis pigmentosa albescens and hereditary maculopathies. Recently, other genes have been identified in autosomal dominant retinitis pigmentosa and these include one at the centromere of chromosome 8 and both the long and short arms of chromosome 7.

Autosomal recessive forms of retinitis pigmentosa have been less well studied; to date, only one gene mutation has been identified in a rhodopsin gene. Recently there has been a report that certain patients with autosomal dominant retinitis pigmentosa have defects in the gene coding for cyclic guanosine monophosphate phosphodiesterase (see Chapter 4) (Box 3.14).

Box 3.14 **Photoreceptor degenerations and ciliopathies**

Examples of photoreceptor degeneration are via genetic mutations that lead to cell functional deficits. It was thought that the rate of cell death or dysfunction increases over time. However, mathematical modelling has determined that it is not a cumulative damage that occurs but that the risk of cell death remains constant and the genetic mutation results in random cell death over time—'one-hit' biochemical model. Such examples of neuronal degeneration are photoreceptor degenerations.

- Ciliopathies define a dysfunction of cilia function in retinal photoreceptors.
- Mutations, for example, that may occur in genes such as ***CEP290*** and ***RPGR*** give rise to photoreceptor dysfunction and degeneration and are ubiquitous in many cells for centrosome function.
- Retinitis pigmentosa related to ciliary dysfunction can be an isolated feature or a part of a syndrome such as Bardet–Biedl syndrome (BBS).

Choroideraemia

Choroideraemia presents in childhood with night blindness, leading to eventual loss of vision in later life. Fundal changes show granular pigmentary changes early in the course of the disease and choroidal atrophy in the late stages. Like X-linked recessive retinitis pigmentosa with the *RP3* gene defect, carrier females also demonstrate patchy non-progressive equatorial pigmentary changes, as a consequence of lyonization. The gene locus for choroideraemia has been identified at the locus Xq21. Several point mutations, deletions and RFLPs resulting from aberrant splicing at the exon–intron junction have been associated with the locus and the disease is caused by deletion to *Rab1* gene.

Norrie disease

Norrie's disease affects males from birth. Features include bilateral congenital blindness secondary to retinal dysplasia and retinal vascular anomalies, impaired hearing, mental retardation, retinal detachment and cataract. This condition is similar to another congenital blinding condition, namely familial exudative vitreoretinopathy, which also shows abnormal vascular development, particularly of the retinal vessels. The gene locus for Norrie disease has been identified as Xp11.1, Xp11.3. Similar gene loci have been suggested for familial exudative vitreoretinopathy, suggesting that these two disorders are related by expressing different mutations at the same gene loci, producing the two phenotypes. Mutations and deletions have been demonstrated at the gene loci for Norrie's disease. The gene's official symbol is *NDP*, a gene for a protein called norrin integral to specialization of retinal cells (Table 3.3).

Table 3.3 **Subretinal delivery of gene therapy from identification of gene mutations**

Disease	Vector-gene
Leber's congenital amaurosis caused by RPE65 mutation	AAV2-hRPE65v2
Choroideraemia	rAAV2-Rep1
Usher type 1B	Ushstat®

Retinoblastoma

Retinoblastoma is a tumour of primitive photoreceptor cells. It is the commonest ocular malignancy of childhood with a prevalence of approximately 1 in 20,000, showing an equal sex distribution. Approximately 40% of cases are inherited. Unilateral tumour, however, is nearly always sporadic, with no family history. Bilateral cases usually have a strong family history of retinoblastoma and are inherited by autosomal transmission. These patients are also at greater risk in later life of developing osteosarcoma. It has been shown that retinoblastoma may be secondary to a deletion of 13q14 in about 4% of patients. The gene locus that encodes the enzyme d-esterase is also situated on chromosome 13 and patients with retinoblastoma show a much reduced level of this enzyme, indicating that the d-esterase and retinoblastoma loci are closely linked. Knudson proposed a hypothesis to explain the fact that inherited cases are usually bilateral, multifocal and of early onset, whereas sporadic cases are unilateral and solitary. The 'two-hit hypothesis' proposes two mutational events in inherited retinoblastoma. The first mutation is present in the germinal cell and would therefore be present in every cell. The second 'somatic' mutation must occur to induce tumour growth in cells with the initial mutation by releasing suppression or regulation of the retinoblast. Two somatic mutations to one retinoblast must occur in sporadic cases of retinoblastoma and are therefore likely to be solitary and unilateral. This retinoblastoma gene has therefore been proposed as a *tumour suppressor gene*, whose presence in normal retinoblasts prevents uncontrolled mitosis. The DNA sequence of the retinoblastoma gene has now been identified and the surrounding complex gene locus identified as a *retinoblastoma predisposition gene*, which is structurally altered in patients with retinoblastoma, although its true function is unknown. As mentioned previously, patients who survive retinoblastoma have an increased risk of developing osteosarcoma and, interestingly, in isolated patients with osteosarcoma the retinoblastoma gene has been found to be deleted.

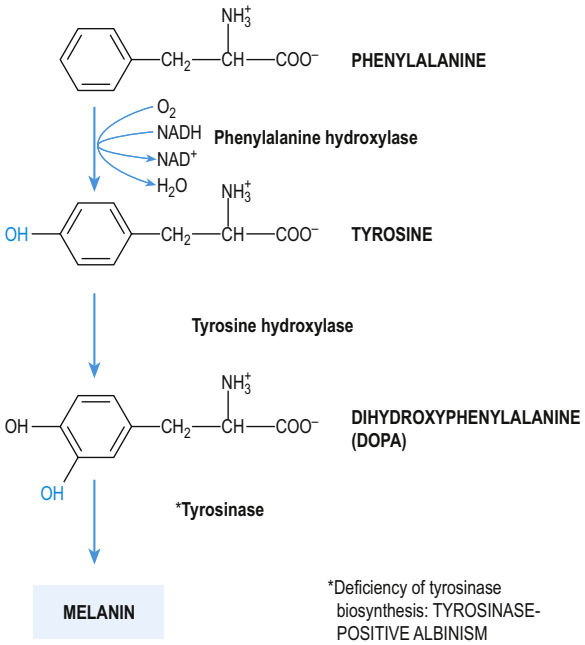


Fig. 3.14 Enzyme defects in albinism.

Albinism

Albinism is a cause of poor visual acuity and nystagmus in children and is divided broadly into two groups: oculocutaneous and ocular albinism. Patients with oculocutaneous albinism may be further differentiated into tyrosinase-producing (tyrosinase-positive) and tyrosinase-non-producing (tyrosinase-negative) groups, shown in children over the age of 4 years by hair bulb incubation in tyrosinase solution (Fig. 3.14).

Oculocutaneous albinism

This form of albinism is inherited in an autosomal recessive Mendelian fashion and within the tyrosinase-negative group gives rise to severe disease with profound visual loss, photophobia and nystagmus, as well as the classic features of iris transillumination, absent fundal pigmentation and absent foveal reflex. Most of the optic nerve fibres cross at the chiasma (90%) and further neuronal disorganization occurs within the lateral geniculate body.

Ocular albinism

Ocular albinism is present when most of the hypopigmentation (hypomelanosis) is confined to the ocular structures. Ocular albinism may be inherited in an autosomal recessive (Nettleship–Falls syndrome) or an X-linked recessive pattern. X-linked recessive ocular albinism gives rise to ocular disease of moderate severity, with a prevalence of approximately 1 in 50,000. Affected males have reduced visual acuity, nystagmus, strabismus and iris translucency. Fundal examination shows classic hypopigmentation and foveal hypoplasia. In this condition giant melanosomes are present in the retinal pigment epithelium and are also found in skin biopsies; they are similar to the melanosome aggregates found in one of the albinoid syndromes, Chédiak-Higashi syndrome (associated with phagocytic dysfunction). Carrier females, whose visual acuity is normal, also demonstrate iris transillumination, retinal pigment epithelial granularity and a preponderance of giant melanosomes in the skin. At-risk females, however, can pose a diagnostic problem and accurate genetic counselling will be available only when future genetic diagnostic tests and techniques for identification of the candidate gene have been developed.

Myotonic dystrophy

Myotonic dystrophy presents with progressive muscle weakness early in adult life. The condition is characterized by expressionless face, frontal balding, gonadal atrophy and myotonia when shaking hands. Patients often develop cataracts and may also develop a pigmentary retinopathy. Myotonic dystrophy is transmitted as an autosomal dominant trait with an incidence of 1 in 20,000. Gene loci for this disease have been localized with the use of RFLP and DNA probes to chromosome 19. However, because of the recent discovery of an unstable DNA mutation consisting of an increased number (more than 50) of a nucleotide triplet (CTG repeats), whose protein product is a member of the protein kinase family, family studies can confirm or exclude those at risk.

Mitochondrial inheritance

As has been stated above, mitochondria contain specific circular DNA that replicates separately from the nuclear DNA and is inherited solely from maternal mitochondria. Recently, some inherited disorders have been identified as having a mitochondrial mode of transmission, because they do not follow classic Mendelian patterns of inheritance.

Leber's hereditary optic neuropathy

This condition is characterized by rapid onset of visual failure particularly in boys, but it may affect either sex. The result of the initial hyperaemic disk swelling and peripapillary telangiectasia is optic atrophy and visual failure. Mothers characteristically pass the disease to their sons, but sons never transmit it (i.e. there is no male-to-male transmission). A characteristic point mutation causing histidine to be inserted instead of arginine at the 340th amino acid of

NADH in complex I of the respiratory chain has been demonstrated in patients with this type of optic neuropathy. Other point mutations in mitochondrial DNA have also been documented. However, there has been no explanation as to why males are predominantly affected in this disorder, which cannot be explained purely on the basis of a single mitochondrial gene defect. With the advent of mitochondrial DNA analysis, investigation of patients with optic neuropathy of uncertain aetiology can be carried out to determine whether Leber's neuropathy is the cause. Also, recent studies have shown that the genotype of the condition is associated with a variable phenotype, in that some families with specific gene mutations demonstrate recovery of vision in up to 50% of patients. In addition, other gene mutations are linked to Leber's hereditary optic neuropathy and are associated with generalized neurological abnormalities.

Other mitochondrial disorders

Other mitochondrial inherited disorders also affect the eye. These include the mitochondrial myopathies, of which the most documented is Kearns–Sayre syndrome. This syndrome occurs secondary to multiple point mutations within the mitochondrial genome, which in turn lead to multiple deletions of varying size. The heterogeneity of the mutations accounts for the variance of the clinical signs encountered, which include a pigmentary retinopathy and progressive myopathy, involving cardiac and proximal limb muscles and a progressive external ophthalmoplegia.

Further reading and references are available online at https://expertconsult.inkling.com

Further reading

Burnight, E.R., Giacalone, J.C., Cooke, J.A., Thompson, J.R., Bohrer, L.R., Chirco, K.R., et al., 2018. CRISPR-Cas9 genome engineering: treating inherited retinal degeneration [review]. Prog. Retin. Eye Res. 65, 28–49. https://doi.org/10.1016/j.preteyeres.2018.03.003. Epub 2018 Mar 22.

Cepko, C.L., Austin, C.P., Yang, X., Alexiades, M., Ezzeddine, D., 1996. Cell fate determination in the vertebrate retina. Proc. Natl. Acad. Sci. U.S.A. 93 (2), 589–595.

Chu, C.J., Barker, S.E., Dick, A.D., Ali, R.R., 2012. Gene therapy for noninfectious uveitis. Ocul. Immunol. Inflamm. 20, 394–405.

Cideciyan, A.V., et al., 2018. Dec 17. Nat Med. https://doi.org/10.1038/s41591-018-0295-0. [Epub ahead of print].

Clarke, G., Collins, R.A., Leavitt, B.R., Andrews, D.F., Hayden, M.R., Lumsden, C.J., et al., 2000. A one-hit model of cell death in inherited neuronal degenerations. Nature 406, 195–199.

da Cruz L., et al. (2018). Phase 1 clinical study of an embryonic stem cell-derived retinal pigment epithelium patch in age-related macular degeneration. Nat Biotechnol. Apr;36(4):328–337. doi: 10.1038/nbt.4114. Epub 2018 Mar 19

Davey Smith, G., 2010. Mendelian randomization for strengthening causal inference in observational studies: application to gene × environment interactions. Perspect. Psychol. Sci. 5, 527–545.

Jacobson, S.G., Cideciyan, A.V., 2010. Treatment possibilities for retinitis pigmentosa. N. Engl. J. Med. 363, 1669–1671.

Klug, W., Cummings, M., Spencer, C., Palladino, M. Concepts of genetics plus mastering genetics with eText – Access Card Package. International Edition, 10th ed. 2011

Lipinski, D.M., Thake, M., MacLaren, R.E., 2013. Clinical applications of retinal gene therapy. Prog. Retin. Eye Res. 32, 22–47.

Luo, S., et al., 2018. Biparental inheritance of mitochondrial DNA in humans . PNAS 115 (2018), 13039–13044.

MacLaren, R.E., Pearson, R.A., MacNeil, A., Douglas, R.H., Salt, T.E., Akimoto, M., et al., 2006. Retinal repair by transplantation of photoreceptor precursors. Nature 444, 203–207.

Petersen-Jones, S.M., Annear, M.J., Bartoe, J.T., Mowat, F.M., Barker, S.E., Smith, A.J., et al., 2012. Gene augmentation trials using the Rpe65-deficient dog: contributions towards development and refinement of human clinical trials. Adv. Exp. Med. Biol. 723, 177–182.

Power, C., Rasko, J.E., 2011. Promises and challenges of stem cell research for regenerative medicine. Ann. Intern. Med. 155, 706–713.

Snead, M.P., Yates, J.R., 1999. Clinical and molecular genetics of Stickler syndrome. J. Med. Genet. 36, 353–359.

Takahashi, K., Yamanaka, S., 2006. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell 126, 663–676.

Wang, S.W., Mu, X., Bowers, W.J., Klein, W.H., 2002. Retinal ganglion cell differentiation in cultured mouse retinal explants. Methods 28 (4), 448–456.

West, E.L., Pearson, R.A., MacLaren, R.E., Sowden, J.C., Ali, R.R., 2009. Cell transplantation strategies for retinal repair. Prog. Brain Res. 175, 3–21.

Chapter 4

Biochemistry and cell biology

CONTENTS

Introduction
The 'omics' creed
Cells and tissues
- The cell
- The extracellular matrix

Biochemical pathways that affect ocular function
- Glucose metabolism and tissue glycation
- Oxidation/reduction and free radical production
- Lipids and lipid peroxidation

The ocular surface
- The tear film
- The conjunctiva
- The lids

Cornea and sclera
- The cornea
- The sclera

The uveal tract
- The iris
- The ciliary body
- The choroid

Aqueous humour dynamics
- Aqueous humour is secreted by the ciliary body epithelium
- Neural/autonomic control of aqueous secretion
- Aqueous humour outflow from the eye

The lens
- Lens transparency
- Lens metabolism
- Ageing in the lens and cataract formation

The vitreous body
- The vitreous body is a true connective tissue containing collagen, gags and cells
- Physicochemical properties of the vitreous gel

The retina
- The neural retina is highly organized in layers
- Blood flow in the retina
- Photoreceptors
- The retinal pigment epithelium

The chemistry of the visual response
- Photochemical reactions in the retina
- Synaptic events between photoreceptors and cells of the inner nuclear layer

Conclusion

Introduction

The eye is a miracle of self-organization. While many tissues such as muscle and kidney comprise predominantly a restricted set of more or less specialized cell types, almost all cell types are represented in the eye including secretory cells, neuronal cells, vascular cells, specialized fibroblasts, tissue myeloid cells and supporting cells, and the matrices contain all of the molecular components found in other tissues. Uniquely, the cells and tissues have been customized for the eye and the forces leading to this event appear to be intrinsic to the cells, since optic cup and eye morphogenesis can be induced in embryonic stem cells *in vitro* given the right conditions. The great leaps forward in genomics (see Chapter 3, p. 148) now being conducted at the single cell level have allowed a profound understanding of how morphogenesis is shaped through interactions and competition between cells responding to these external matrix forces (Guillaumet-Adkins, 2017). Remarkably, only four transcription factors seem to be necessary to encourage developing eye tissues *in vitro* to take shape, for example the differentiation of 'retinas in a dish' from progenitor cells under control of the Onecut (Oc 1-3) genes (a transcription factor family which make a single 'cut' at their DNA-binding domain). Interestingly, these factors

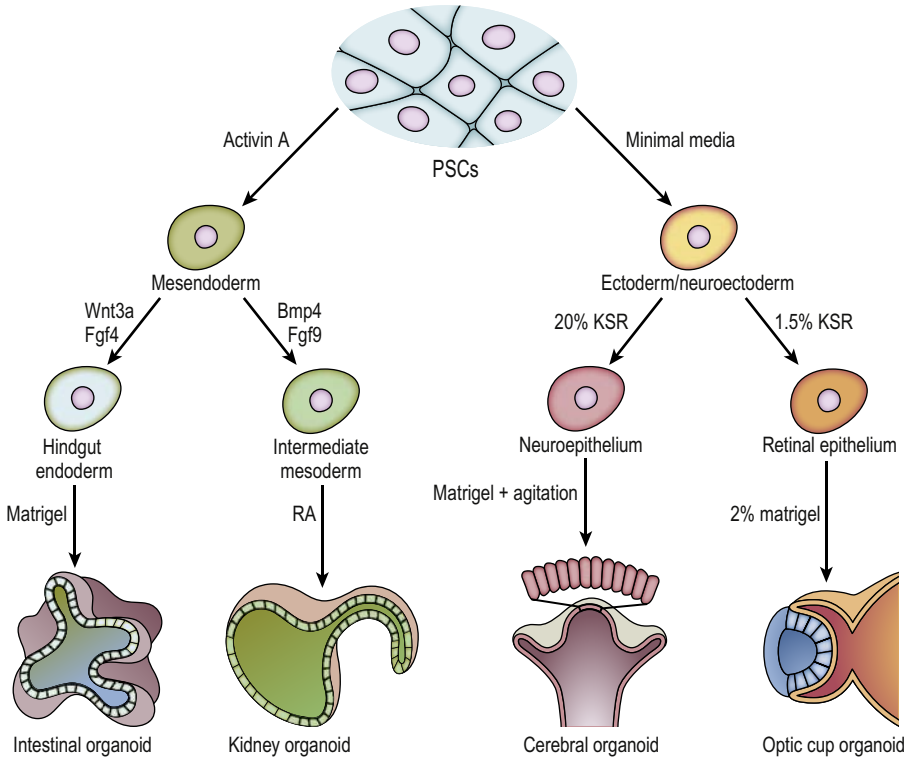


Fig. 4.1 Organoids are tissue culture methodologies for generating tissues *in vitro* from pluripotent stem cells (*PSCs*) dependent on the set of growth factors used. Different combinations of customized serum replacement media and/or collagen matrices promote neuroectoderm, retinal, brain or renal organoids. (From Lancaster and Knoblich, 2014.)

are not eye specific but also support development of other tissues such as pancreas via common downstream targets such as the PAX6 and the Prox1 gene (Fig. 4.1).

The unique feature of cells and tissues in the eye is that they are organized for the transmission, reception and conversion of light energy into cellular signals. Cells respond to stimuli in a remarkably similar manner. What differentiates one cell from another are the stimuli each cell responds to and the mechanisms it uses to respond. Cells have the genetic potential to express any type of receptor but their unique specialization is down to the limited set of membrane receptors they express. Through these receptors, cells respond to a specific stimulus by activating an intracellular second messenger system that has a limited generic range, i.e. the same set of signalling molecules is frequently activated by a wide range of ligand–receptor interactions: the specificity lies in the ligand binding. This produces a programmed response in the cell, resulting in an effect, e.g. aqueous secretion from acinar lacrimal gland cells after stimulation by adrenaline (epinephrine), an ocular muscle action potential after neurotransmitter release from nerve endings, or rhodopsin activation by a photon of light. Remarkably, the process of signal induction and transmission is based on a very limited set of biochemical reactions often involving a glucose-dependent, ATP-energy-driven mechanism, such as the addition or subtraction of a phosphate group to the signalling molecule, respectively via enzyme activity provided by a kinase or a phosphatase.

Cells are organized, singly or in groups (tissues), to receive information from the environment (via membrane receptors), to signal this information to the intracellular compartment (via signalling networks), to convert the message into cellular responses (gene activation and protein transcription) and to relay this information to the outside world (e.g. changes in cell behaviour, tissue function, secretion, etc.). Innumerable molecules and genes are involved in a single response by the cell through these networks, in which thousands of molecular interactions are connected through molecular 'hubs' (similar to hubs regulating aircraft traffic) (see next section). Cells may also respond simultaneously to several stimuli through several receptors and controlling the flow of this information can be difficult. In recent years, this function seems to be attributed to that mysterious cellular constituent chromatin, whose histone tails may act as signal storage and

converter devices, similar to digital–analogue converter devices in the electronic industry. In this way, accumulated repeated rapid ON–OFF signals over time can induce epigenetic and indeed phenotypic change in cells and tissues, which not only influence cellular functions as they age but can also be transmitted to the next generation.

Further information on the mechanism of how induced epigenetic changes can lead to heritable functional effects is available online at https://expertconsult.inkling.com

These concepts have been derived from the vast amount of information made available through full genetic analysis of various organisms (genomics) and the use of novel methods of investigation including microarray technology and informatics.

The 'omics' creed

No longer a philosophy, omics has become a creed based on in depth analysis of quantitative data. The science of 'omics', a term applied to a body of work, knowledge or data, including genomics, transcriptomics, metabolomics, proteomics, signalomics and the microbiome, has generated a vast amount of information leading to some further 'omic' subdivisions such as cancer genomics or toxicogenomics. The concept has reached the clinic in the form of 'morphomics' whereby analytic CT scanning rather than body mass index is used to quantify adipose tissue in patients receiving haematopoietic stem cell transplants.

Genomics examines the many genes that might be involved by increased or decreased expression, while transcriptomics studies the many transcription factors that might be activated or deactivated in any one cellular behaviour, such as cell division, and metabolomics investigates the many biochemical pathways that might be used, or not, in conversion of one molecular species to another in the process of energy generation and consumption. For the epigenetics researcher, there is even an epigenome, which provides information on DNA methylation, histone modifications and chromatin remodelling. Meanwhile, the microbiome is a term applied to the databank relating to gut commensals and has considerable influence on the immune system (see Chapter 7, p. 384). In several of these '-omes', molecular networks are entrained, which in themselves reveal the extensive interdependence that one system has on another. In addition, the notion of central ('hub'-based) molecular species without which the entire network would collapse allows a hierarchy of importance to be applied to molecules. This is demonstrated in the genetic mutagenesis studies in which certain molecules, such as transforming growth factor-β, are lethal to the embryo when deleted, while others, such as plasminogen activator inhibitor 1, barely alter the murine phenotype.

Signalling networks are a prime example of how cellular information is transmitted. There are hundreds to thousands of signalling receptors in the cell membrane interacting with around 10 second messenger 'hubs' in large interacting intracellular networks of several thousand cellular proteins. Cellular proteins for particular signalling pathways form complexes or platforms that might be bound to intracellular organelles such as endoplasmic reticulum (ER) or might be located in the cytoplasm as mobile units, allowing several pathways to 'cross-talk'. The density of the signalling events is very high and difficult to model, but algorithms are being developed for what is known as 'combinatorial complexity' and are based on assuming the simplest possible interactions.

Further information on the modelling of 'combinatorial complexity' in signalling networks is available online at https://expertconsult.inkling.com

Examples of typical general second messenger systems include:

- receptor tyrosine kinase-linked receptor systems (RTK)
- ion channels and pumps
- G protein-driven messengers
- seven-transmembrane loops.

There are also several more generic types of receptor.

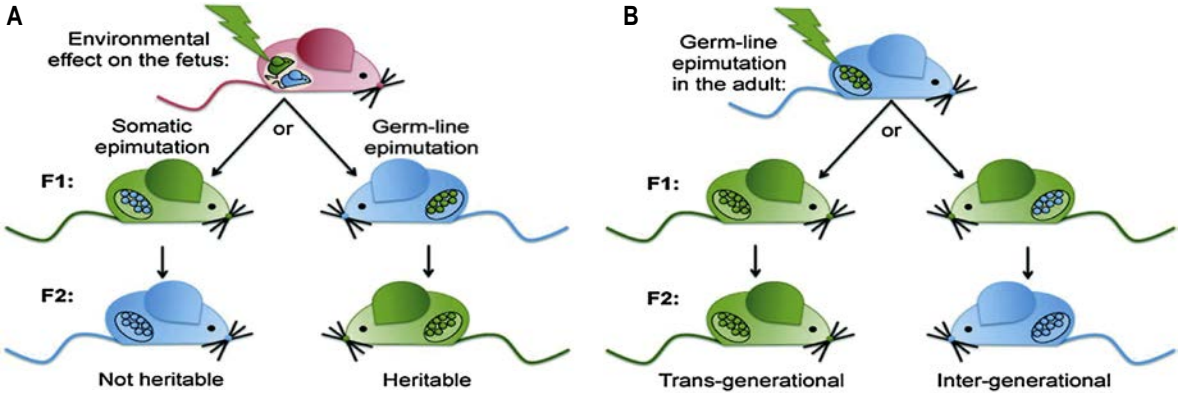
Each of these may interact with other intracellular signalling systems and the signalling systems themselves may be 'customized' to respond selectively depending on the conditions (Fig. 4.2). In addition, each receptor may be represented many times in a single cell (it is estimated that there are between 5000 and 10,000 major histocompatibility (MHC) class II molecules on a dendritic cell), while several different ligand–receptor pairs may act in clusters at the cell surface, as for instance in the T-cell receptor synapse (see Chapter 7, p. 427).

Signalling networks behave similarly to other biological networks, such as metabolic and gene transcriptional networks, and probably represent a basic biological organizing system. Three basic concepts underpin a signalling network: signalling pathways, signalling modules and signalling nodes. In addition, receptors exist not only on the cell surface but intracellularly, for instance on endosomes and in the nucleus, and ligands may have to be transported intracellularly to interact with their receptor. Thus, the cellular location of signalling pathways impinges directly on the ultimate response.

Information transmitted via ligand/receptor interactions can also be bidirectional, as for instance in the regulation of stem cells in stem cell niches through integrin molecule binding to extracellular matrix proteins (Fig. 4.3).

Epigenetics and inheritance

Studies in mice have shown that epigenetic changes to the gene pool of the fetus can lead to damage to either the somatic cells (genes not transmitted) or the germ-line cells (genes passed on). If the damage is not sufficient to kill the fetus the first-generation progeny (F1) will only pass on their mutated genes to the next generation (F2) if the germ-line genes have been affected (eFig. 4.1). If, however, there is epigenetic germ-line damage and it is stably expressed in the adult, it may be passed on as part of the normal germ-line genes (i.e. is passed down several generations—transgenerational) unless reprogramming occurs in the embryo before implantation. Reprogramming can be viewed as a type of fail-safe repair mechanism. Epigenetic mechanisms of disease arise from environmental damage and are increasingly recognized as determinants of disease risk in children and grandchildren of affected individuals for instance in increased risk for diabetes or obesity.

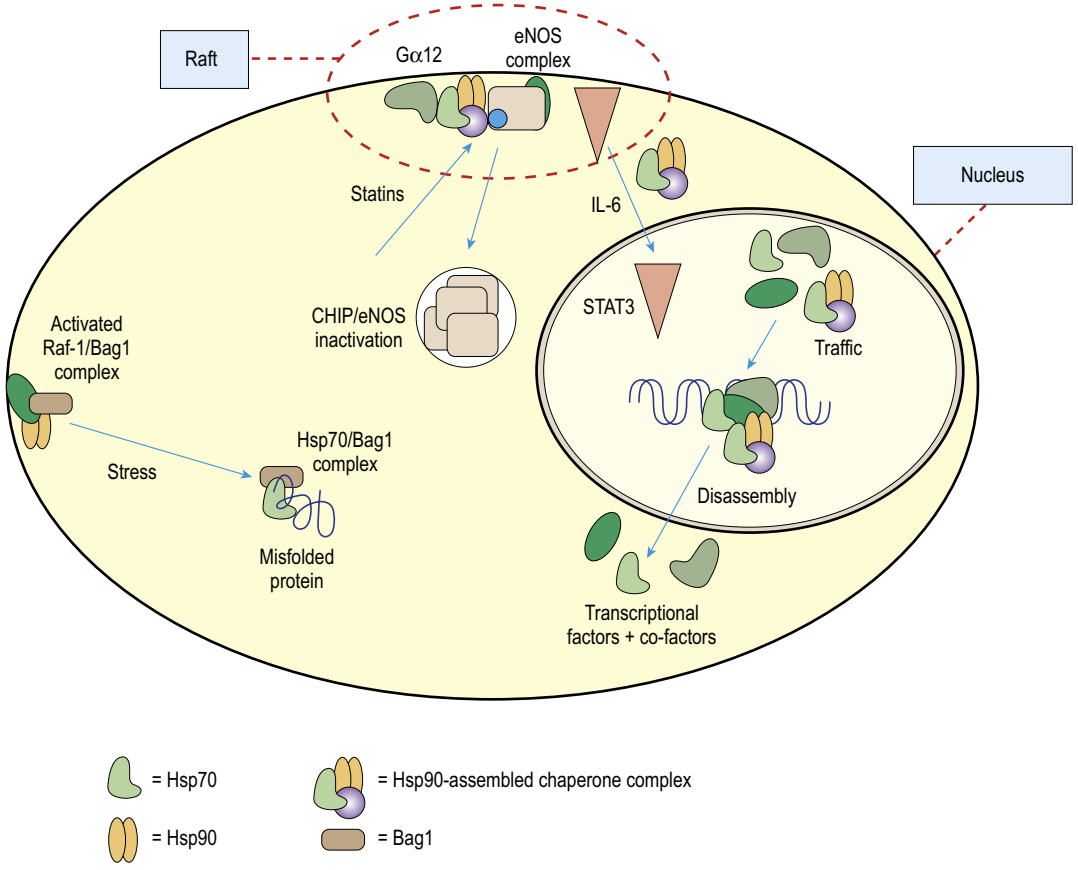


eFig. 4.1 Epigenetic inheritance could occur through several mechanisms: (**A**) the fetus (F1) might be affected by environmental factors such as chemicals or drugs which would lead to either somatic (*blue* cells) or germ-line (*green* cells) mutations. In this case only the somatic mutations are heritable. (**B**) If a germ-line mutation is inherited, it can either be inherited over many generations or the germ-line cells can be reprogrammed back to normal cells and the genetic defect only lasts for one or two generations (inter-generational). In the figure, *pink* and *blue* mice are normal, *green* mice have the genetic defect. (From Stringer et al., 2013.)

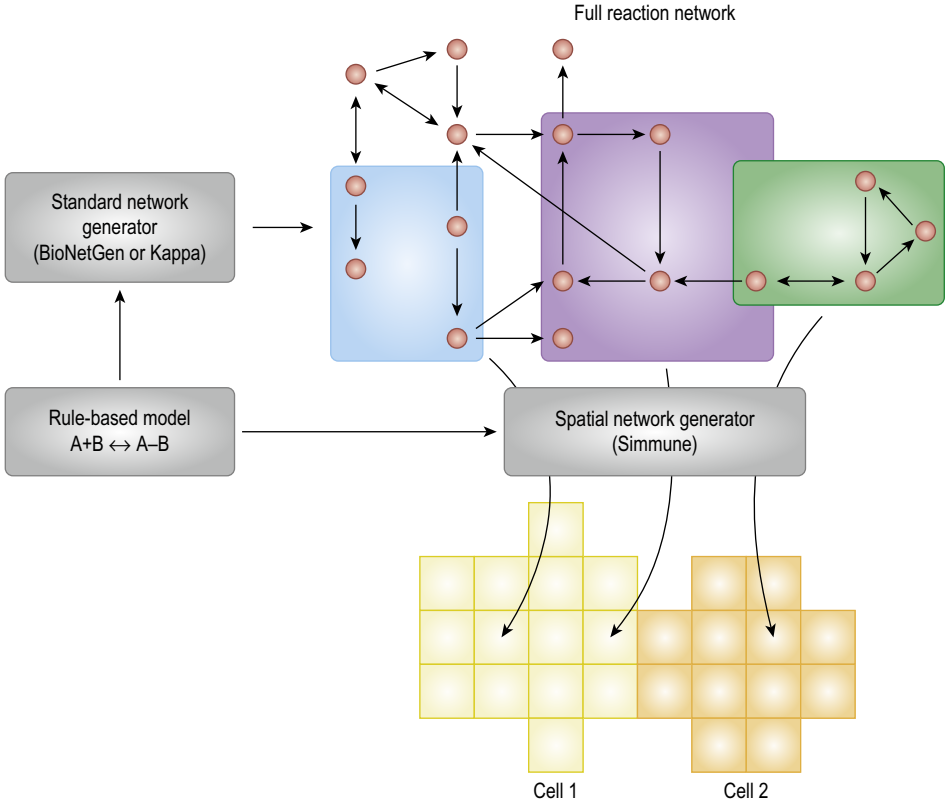
Signalling networks and combinatorial complexity

Many thousands of molecules and genes are involved in a single response by a cell through intracellular networks; in this process immense numbers of molecular interactions are connected through a system in which some molecules form many interactions with others in 'hubs' while other molecules make very few interactions and are on the periphery of the network. These concepts are derived from the vast amount of information made available through full genomic analysis of various organisms and the use of novel methods of investigation including microarray technology and bioinformatics. Furthermore, hub molecules are frequently transient in activity, one molecule acting as a hub during activation of one signalling pathway while the same molecule acts merely as a relay station during activity of another pathway. Information usually proceeds from 'outside in' to the cell, but on occasion information initiated outside the cell can be relayed back to extracellular targets. Many of the cellular proteins may not be directly involved in signalling but may act as adaptors or may amplify/diminish the overall response. In addition, other proteins act as 'chaperones' to protect proteins and signalling molecules for optimal function (eFig. 4.2).

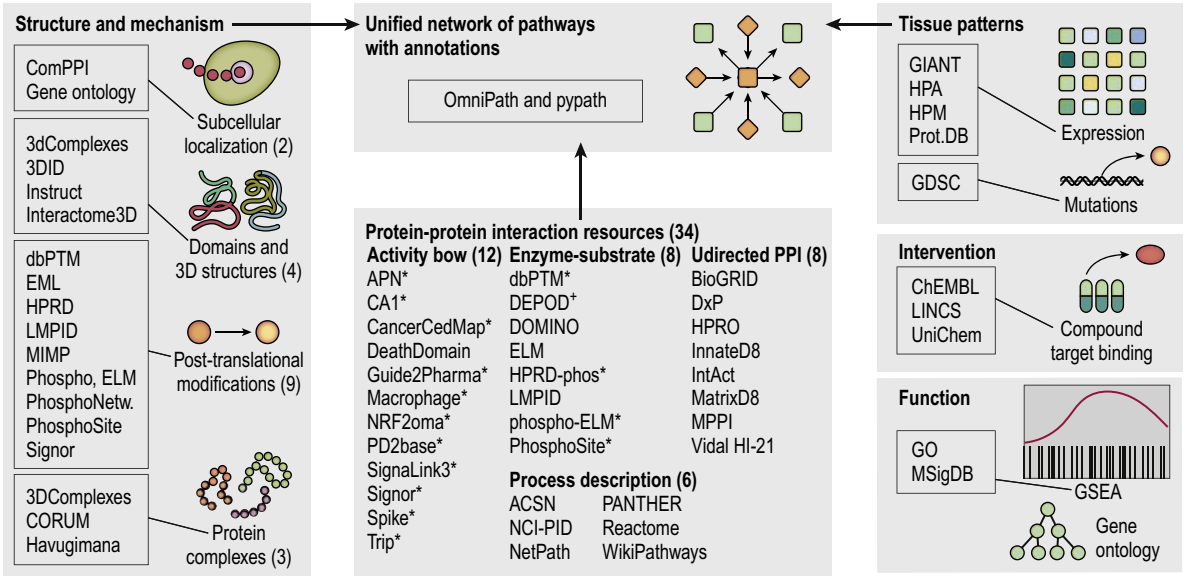
The combinatorial complexity of signalling events has been addressed using simplified rule-based modelling systems, which assist greatly in predicting the spatial location of signalling events within the cell. This has important consequences for the effectiveness and duration of the signal (eFig 4.3). Several resources for use in investigating signalling pathways with the hubs and nodes are now available such as Omnipath (eFig 4.4).



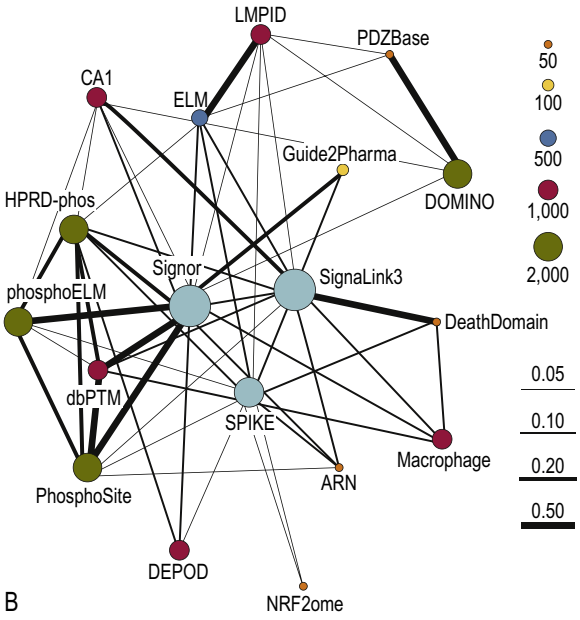
eFig. 4.2 Molecular chaperones in the regulation of signalling: a few recent advances. Chaperone proteins are everywhere in the cell, shepherding essential functional proteins such as enzymes and signalling molecules to ensure their proper functioning.



eFig. 4.3 A systems biology and big data approach to cell signalling. A combination of rule-based modelling with dynamic cellular geometries is one approach to determining the location of sites within the cells where networks are generated with the possibility to revise them as new information is obtained. (With permission from Sneddon and Emonet, 2012.)

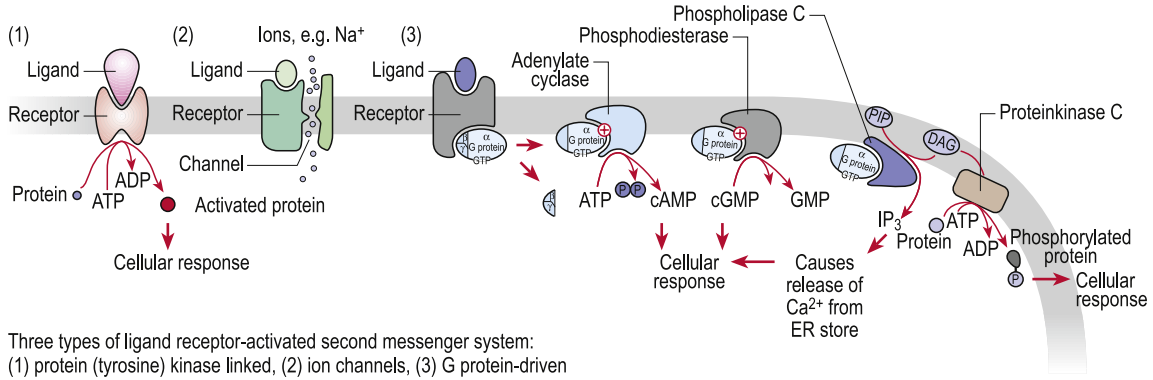


A



B

eFig. 4.4 Resources featured in OmniPath and pypath. (**A**) Overview of OmniPath and pypath. These resources can provide direction (*) or sign and direction (+) of interactions. Data types provided by the listed resources can be integrated with pathways in different ways using databases such as*GO*, Gene Ontology; *HPA*, Human Protein Atlas; *HPM*, Human Proteome Map; *PhosphoNetw*, PhosphoNetworks; *Prot.DB*, ProteomicsDB. (**B**) Overlap of interactions across causal resources as in (**A**). Circle size denotes the number of interactions per resource, and line widths show the overlap of interactions between them. (From Turei et al., 2016.)



Three types of ligand receptor-activated second messenger system:
(1) protein (tyrosine) kinase linked, (2) ion channels, (3) G protein-driven

Fig. 4.2 Ligand–receptor-activated second messenger systems.

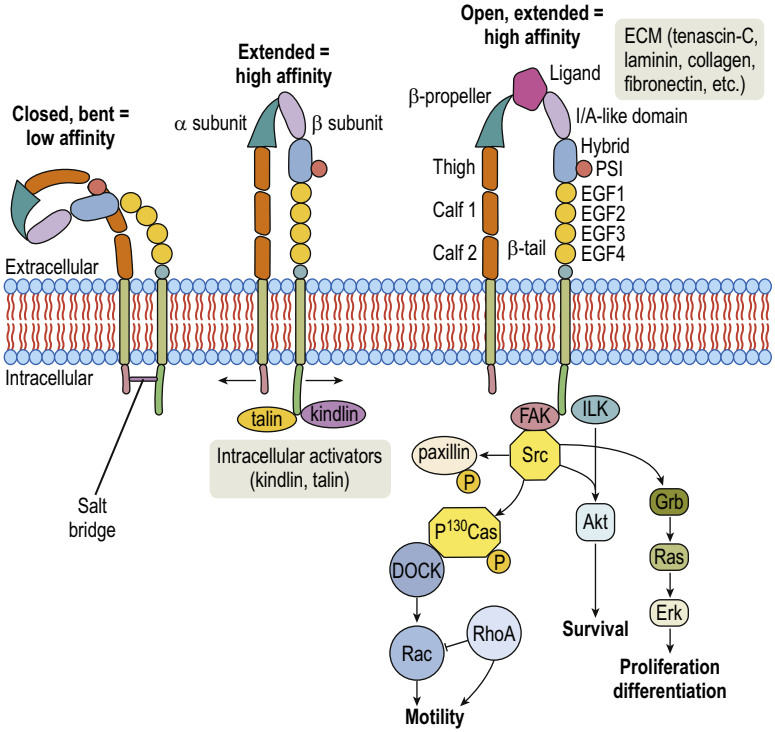


Fig. 4.3 Integrin structure and activation. Activation of integrin heterodimers leads to intracellular signalling cascades and resulting processes such as cell motility, cell survival, cell differentiation and neurite outgrowth. Conformational changes in the integrin molecule from a quiescent inactive state to full on induction of intracellular signalling represented as both 'Inside–Out signalling' and 'Outside–In signalling'. (From Cheah and Andrews, 2018.)

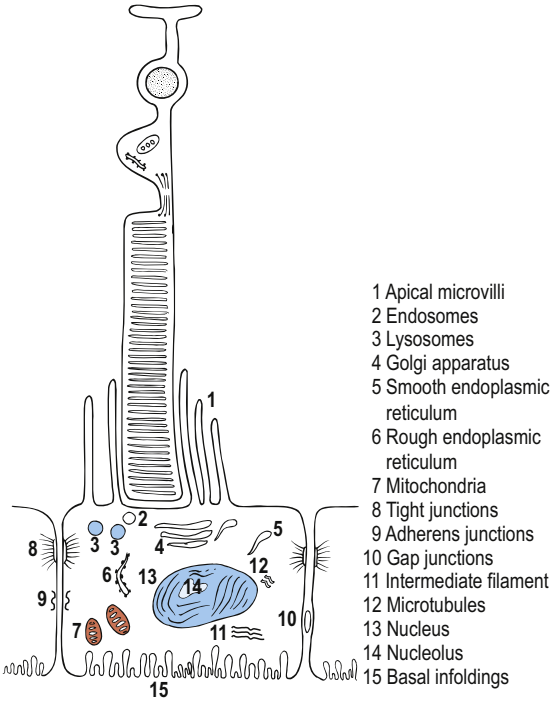


Fig. 4.4 Diagram of RPE and photoreceptor cells. The photoreceptor outer segment lies in close apposition to the RPE cell, enclosed in a sheath of apical microvilli. The RPE cell is a terminally differentiated epithelial cell with several functions, one of which is to transport fluid across the cell towards the basal infoldings and into the choroid sink.

Cells and tissues

The cell

General structure

Technology drives science and there is no better example of this than the discovery of the cell as the basic unit of living organisms by Anton van Leeuwenhoek using a compound optical microscope. The basic structure of the mammalian cell can be illustrated by the retinal pigmented epithelial (RPE) cell (Fig. 4.4) as it contains most of the recognized cellular structures and intracellular organelles. While all cells have the potential and machinery for mitosis and motility, many adult tissue cells such as the RPE cell are considered terminally differentiated, non-motile cells, except under pathological conditions. The RPE cell is an example of a bidirectional transporting epithelial cell with polarity, i.e. an apical surface with microvillous processes and a basal surface with numerous infoldings. The RPE is also an example of how the basic structure of the cell has been modified extensively, as in several types of specialized cells in the eye. However, cells define tissues and for decades it has been assumed that all cells of one type in a tissue are similar if not identical, e.g. a kidney cell or a corneal endothelial cell. However, the technological advances in single cell genomics and transcriptomics have revealed that there is great heterogeneity in cells within a tissue and that the cell-to-cell variability of a particular cell type varies considerably depending on the age, state of activation, plasticity, shape and size. This applies par excellence to the RPE cell, which, although terminally differentiated, has the potential to divide, does so at a very slow rate and undergoes marked morphological changes with age. Dysfunction of this critically important cell underlies the pathology of age-related macular degeneration (AMD), one of the commonest causes of blindness in developed nations (see Chapter 9, p. 518).

The plasma membrane

The plasma membrane, which surrounds all cells, is a selective two-way barrier to passive diffusion, which also has active transport mechanisms subserved by specialized proteins (for instance ion channels, pumps and transporters) floating in a lipid bilayer, composed of phosphoglycerides, sphingolipids and sterols, which forms spontaneously as a result of its content of phospholipids. The barrier function of the plasma membrane applies even to ions and is the basis of the electrical potential that occurs across cell membranes and is energy-driven. For example, a group of proteins known as the 'endosomal sorting complex required for transport' (ESCRT) is central to many plasma membrane processes such as endosomal sorting, trafficking of vesicles, biogenesis of organelles and maintenance of plasma membrane integrity, and requires a specific ATPase (the AAA-ATPase vacuolar protein sorting 4, Vps4p) (Fig. 4.5). Many other proteins are also suspended in the plasma membrane, such as receptors for hormones, neurotransmitters, viruses and other cells. Many of these receptors have a three-part structure with an extracellular, variably sized, component, a transmembrane component, and a short intracellular section coupled to the second messenger system, which may activate or inhibit (regulate) cellular function. The plasma membrane is a variably thick lipid bilayer crowded with many membrane proteins whose functions not only serve transmembrane traffic via receptors and transporters but also have a direct role to play in the physiology and indeed the survival of the cell: this is in part determined by the types of motion and tension which different membrane proteins exert (Box 4.1).

Many of these proteins are held in patchy distributions within the lipid bilayer, termed lipid (micro) domains or 'lipid rafts', which themselves have variable constitution. The term lipid raft, however, has fallen out of favour mainly because of the heterogeneity of the lipid composition, the predominance of proteins and the dynamic changing nature

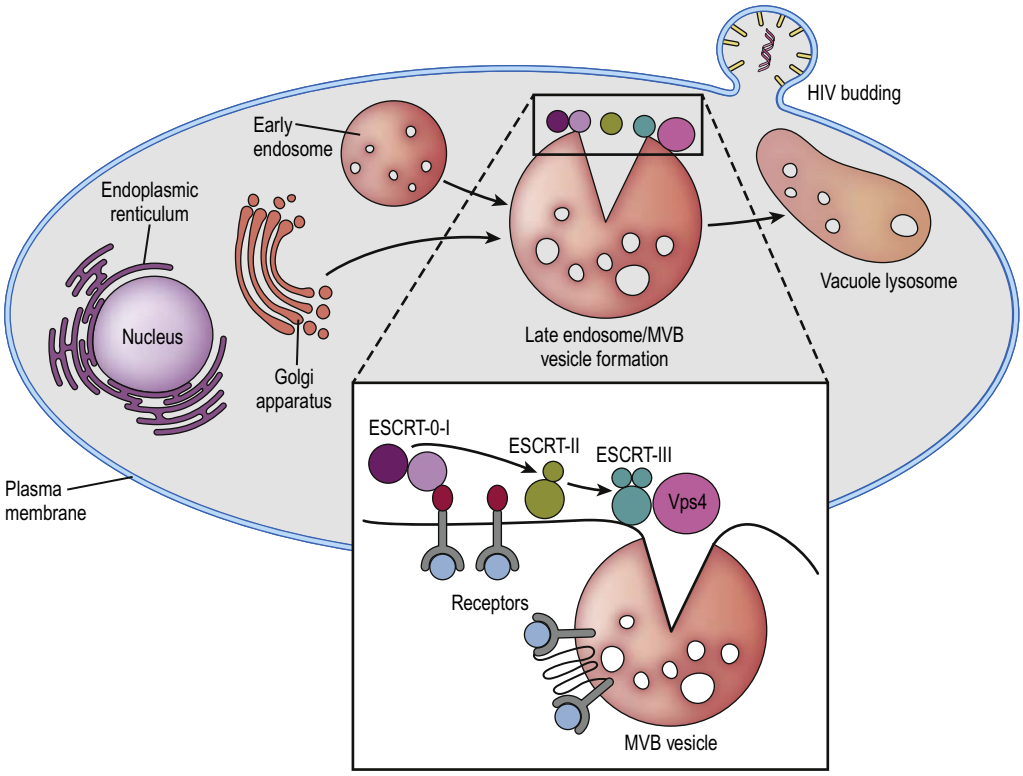


Fig. 4.5 Cellular transport is a complex process but viruses can hijack this machinery with ease. Endosomal sorting complexes required for transport *(ESCRTs)* and the AAA-ATPase vacuolar sorting protein *(Vps4p)* are two examples of intracellular structures required for normal homeostatic protein handling. However, they inadvertently facilitate formation of intraluminal vesicles (multivesicular bodies, *MVBs*) and viral budding, for instance in Ebola and HIV infections. MVBs budding off from the initially infected cell disseminate to infect other cells. (From Ahmed et al., 2019.)

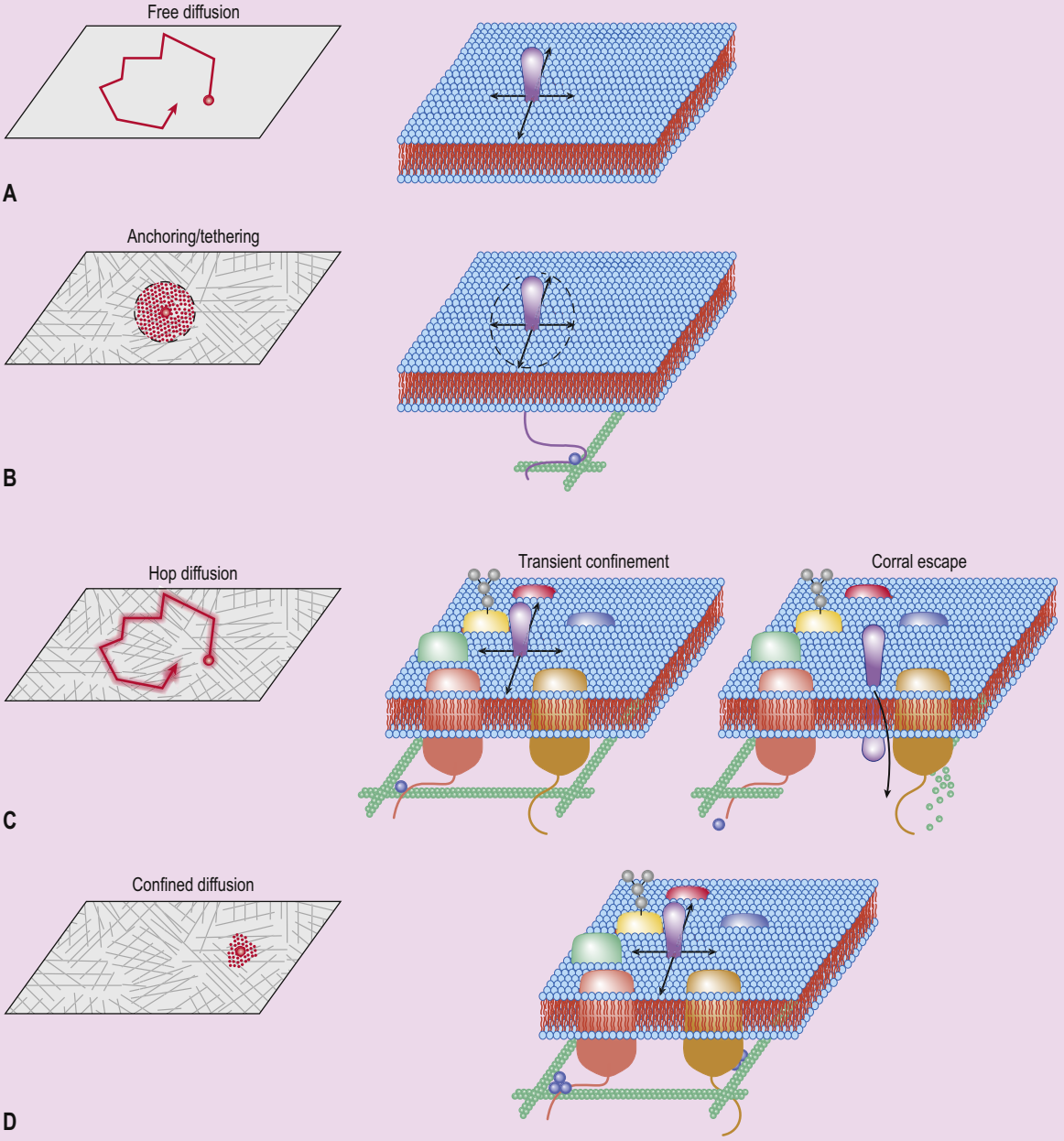
of the membrane patches, and the term membrane nanodomain or transient nanodomain is more in use (Lu). Other microdomains decorate the cell surface in the shape of irregularly pitted invaginations and evaginations such as specialized structures for endocytosis (clathrin-coated pits) and caveolae, embedded in a glycoprotein-rich matrix (glycocalyx). These are also involved in a plethora of functions including cell signalling, protein trafficking, cell movement, waste disposal (exocytosis) and even cell survival. In virally infected cells, these microdomains provide the envelope for budding virus particles (Fig. 4.5). Microdomains are detergent-resistant and usually contain a specific protein such as caveolin in caveolae, or several proteins, as occur in lipid rafts, areas specialized for specific functions such as the immunological synapse in antigen-presenting cells (see Chapter 7). Other microdomains include tiny domains (nanodomains), which contain GTP-binding protein (inhibitory) (GPI)-anchored proteins (important in some types of signalling) and glycosphingolipids, transient confined zones of varying size, and small regions composed of more fluid lipids.

Cells of the nervous system are especially rich in lipid rafts, which contain a high content of cholesterol, and sphingolipid rafts appear to have an organizational function either as discrete functional elements in which both the ligand and receptor are present, or where the activation of the receptor is dependent on recruitment to the raft of the effector ligand. It is likely that the enormous complexity associated with microdomain function is being revealed using a systems biology approach such as a combination of RNA sequencing, proteomics and transcriptomics.

Many other specializations occur in the plasma membrane, depending on the cell type, such as junctional complexes, gap junctions, desmosomes, hemidesmosomes and contact sites with the basement membranes (see later). In the eye these membrane specializations are developed to a high level. For instance, the photoreceptor cell (Fig. 4.4) (see Chapter 1, p. 40) is a highly polarized structure comprising a receptor component, a nucleus and a synapse. The rod photoreceptor (specialized for scotopic vision; see Chapter 5, p. 279) develops as an evagination of the plasma membrane,

Box 4.1 **Membrane proteins have many ways of moving**

Free diffusion (**A**) of proteins in the lipid bilayer allows random or directional movement depending on external triggers. Many proteins are, however, anchored or tethered either to cytoskeletal proteins or to extracellular matrix proteins (**B**). Examples include GPI-tethered proteins. Some proteins are only transiently tethered during a particular function, for instance in the immunological synapse of immune cell activation (**C**). Some of the proteins can escape from the synapse or corral and become internalized. In contrast some proteins have limited movement and are confined to a region of the plasma membrane, as in confined diffusion (**D**).



which folds upon itself many times to form stacks of membranous discs by fusion of the peripheral disc membrane. The plasma membrane is typical of any cell, i.e. it comprises a lipid bilayer containing a high concentration of membrane proteins. The lipid bilayer is a self-assembling sheet of phospholipid that adopts the bilayer format because of the physicochemical properties of the polar phospholipids, ensuring that the polar groups are external and the hydrophobic groups form the inner layer of the leaflet. The photoreceptor can adopt this special arrangement because it has more cholesterol in its bilayer, not only making it less fluid but also preventing crystallization of the membrane by inhibiting possible phase transition of the hydrocarbons. The synaptic terminal of the photoreceptor interacts with mobile Ca^{2+} channels in the bipolar cell which appear to function within lipid rafts confined in their mobility by attachments to the cytoskeleton (see Part B of the figure in Box 4.1).

Endoplasmic reticulum and Golgi apparatus

A wide variety of cell organelles are embedded in a cytoplasmic gel, which is traversed by a system of membranes, the ER. The ER, a series of thin bilayered membranes, is a flowing dynamic system constantly forming and reforming. Cisternal, tubular and vesicular elements exist. The rough ER (RER) is distinguished from the smooth ER as a ribosome-studded structure that is highly developed in secretory cells such as the lacrimal gland acinar cell, and is specialized in other cells, e.g. the sarcoplasmic reticulum of striated (including extraocular) muscle. The RER is arranged *en face* in rows or rosettes of ribosomes (polysomes). Newly synthesized proteins come off the ribosomes and are threaded through the lipid bilayer into the interior of the ER where they are post-translationally folded, ready for secretion via the Golgi apparatus by vacuolar budding and fusion with the plasma membrane for exocytosis (Box 4.2). Endocytosed surface proteins, particularly receptors, may traffic through tubulovesicular extensions and become recycled to the plasma membrane either of the cell itself or as part of the secreted vesicle (exosome, microvesicular body) (Fig 4.6). Recycling occurs through two related mechanisms, fast recycling involving the protein endophilin (Hauke, 2015) and slow recycling involving the better known clathrin (Fig 4.6). The smooth ER is also the site of synthesis of molecules such as lipids, triglycerides and steroids, and is prominent in cells such as the RPE and meibomian gland cells; the Golgi apparatus is a membranous stack of flattened cisternae that receives proteins, now equipped with a leader sequence required for secretion, for sorting and exocytosis from the smooth ER; the smooth ER also plays a role in lipid trafficking and modification.

The ER and the Golgi apparatus have other functions: they are involved in signalling, for instance during mitogen activation of cells for proliferation, via the small GTPases, which are activated at the plasma membrane and also signal in the ER and Golgi body (Box 4.3). In fact, small GTPases and other molecules, such as phosphoinositides, provide a signature for each organelle and are involved in the specific lipid membrane folding that characterizes each organelle (Box 4.3). The ER also provides a regulatory role in ensuring quality control of good protein secretion (GPS). This is regulated by a process of ubiquitinylation, a post translational modification by which the level of ubiquitin molecules added to the protein determines whether the protein is targeted for secretion or degradation in the proteasome.

Further information on Protein modification in the context of ubiquitinylation and sumoylation is available online at https://expertconsult.inkling.com

If proteins are not folded correctly, they are unable to function and the ER generates a 'stress response', also known as the unfolded protein response (UPR), which is one form of ER stress. Together with a second fail-safe mechanism, mediated by a kinase known as the mammalian target of rapamycin (mTOR), the UPR and the mTOR pathways control many cellular processes including programmed cell death (apoptosis), protein translation, waste disposal (autophagy), glucose metabolism and energy supply (ATP), and responses to inflammatory stimuli.

Further information on Proteins involved in ER stress is available online at https://expertconsult.inkling.com

The ER also forms the nuclear envelope during telophase, when a series of flat vesicles surround the chromosomes and fuse at their edges. The envelope contains many *nuclear pores*, which are composed of eight cylindrical filamentous structures in a highly organized arrangement. Pores act as molecular sieves, permitting rapid passage of small (4.5 kDa) proteins and slower passage of larger molecules (12–70 kDa). The outer aspect of the nuclear envelope in secretory cells is lined with ribosomes and polysomes, while the inner surface is in contact with a nuclear filamentous matrix.

Mitochondria

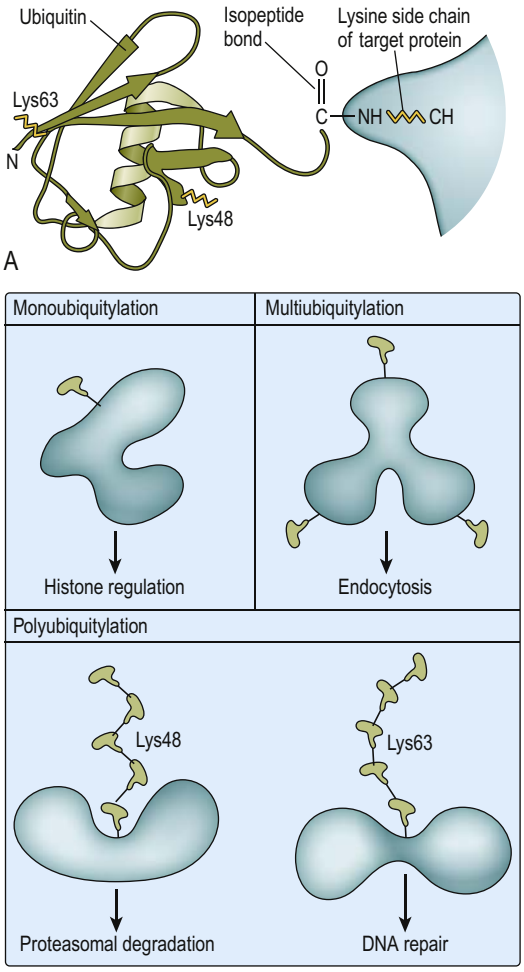
Mitochondria are small (2 μm long) oval-shaped organelles comprising a two-membrane system of compartments, the inner one of which is composed of two domains, the inner boundary membrane (IBM) and the invaginations, folded into structures termed cristae (Fig. 4.7). At the cristae junction with the IBM, the mitochondrion inner membrane contact site and organizing centre (MICOS) is located and separates the two populations of proteins: those in the IBM assist in protein movement (translocases) and those in the cristae contain the proteins involved in the respiratory chain, the F1FO-ATP synthase and carrier proteins for ADP/ATP (see later).

The intermembrane space contains carrier proteins that are responsible for the transport of metabolites between the two compartments and also between the cytosol and

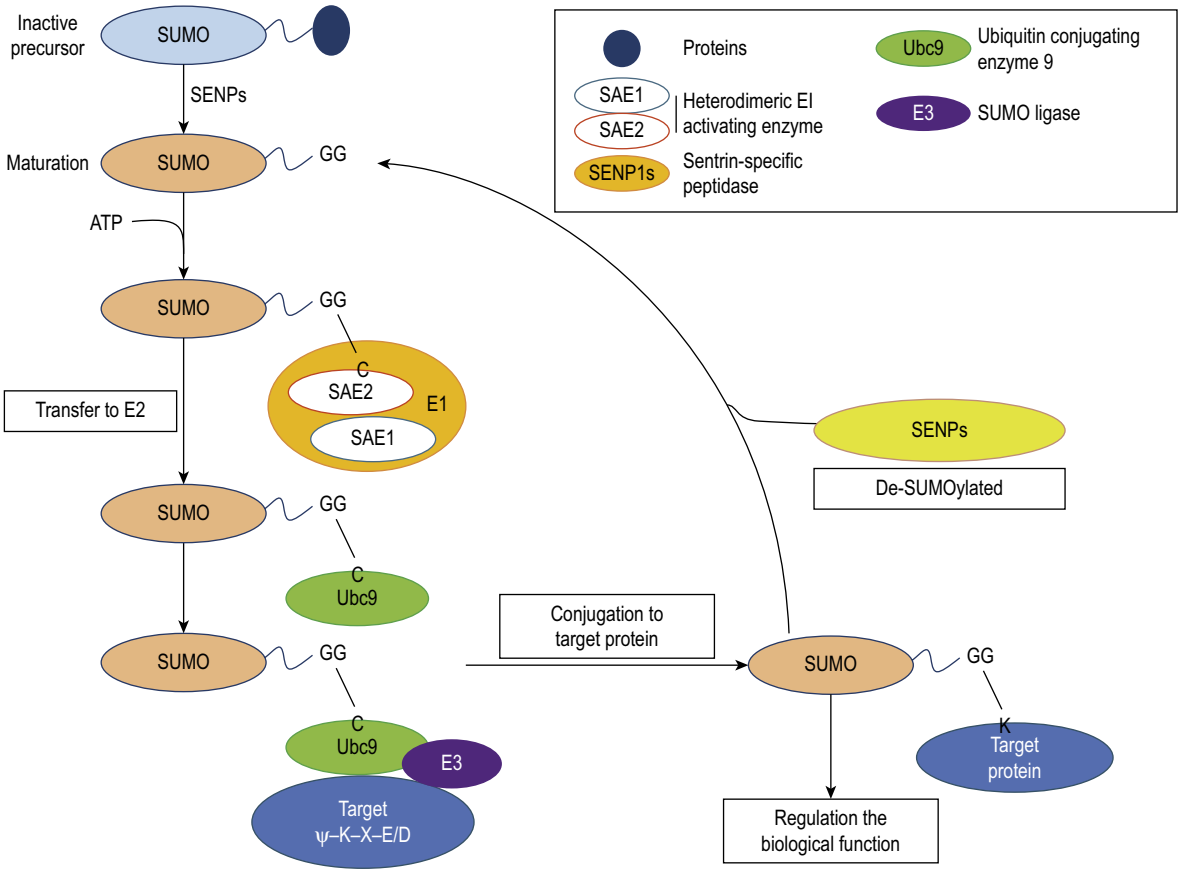
Ubiquitinylation and sumoylation

There are many post-translational modifications to proteins that alter cellular activity including phosphorylation, glycosylation, acetylation, ubiquitinylation, and sumoylation. Ubiquitinylation is an essential process for regulation of core signalling network proteins and uses E3 ligases and deubiquitinylases to add or to remove ubiquitin molecules on proteins, thus determining the protein's folded state and its fate. The way a protein is folded critically determines its function. This process is monitored by the ubiquitous cellular protein, ubiquitin (Ub), which is bound to the protein as it emerges from the endoplasmic reticulum. Ubiquitinylation at single sites prepares the protein for specific functions such as histone regulation or endocytosis (eFig. 4.5). As more and more Ub molecules are added to the protein, it may be targeted for degradation in the proteasome, a filtering complex of enzymes or for DNA repair functions. As it enters the proteasome, the Ub molecules are removed from the protein and it becomes susceptible to degradation and/or recycling/removal. This process is tightly regulated (eFig. 4.5).

Sumoylation is a similar post-translational modification process (SUMO = small ubiquitin-like modifiers) intrinsic to many cellular processes such as cell growth, migration and response to stress. Sumoylation involves ligases similar to ubiquitinylation and prepares proteins for processes such as subcellular localization, DNA binding, transcriptional regulation and many other functions (eFig. 4.6).



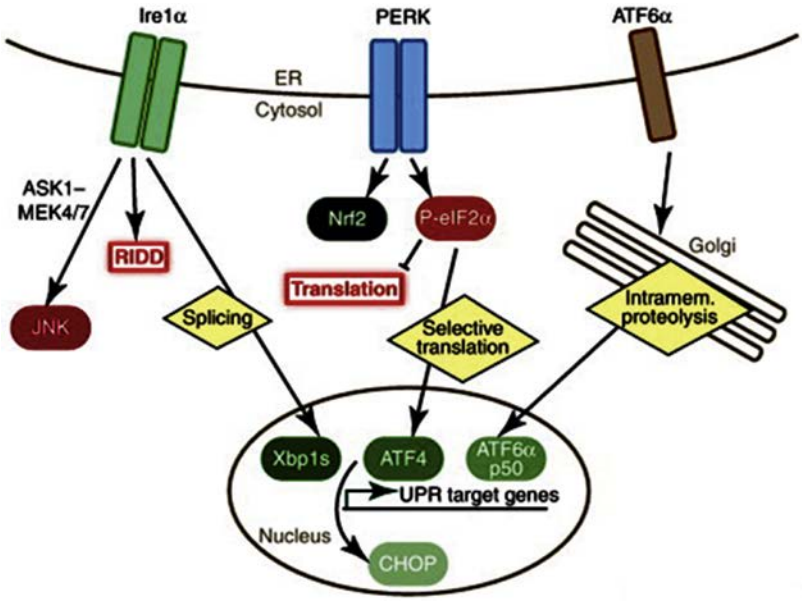
eFig. 4.5 (**A**) The three-dimensional structure of ubiquitin, a small protein of 76 amino acids. A family of special enzymes couples its carboxyl end to the amino group of a lysine side chain in a target protein molecule, forming an isopeptide bond. (**B**) Some modification patterns have specific meaning to the cell. Note the two types of polyubiquitination differ in the way the ubiquitin molecules are linked together. Linkage through Lys48 signifies degradation by the proteasome, whereas linkage through Lys63 has other meanings. Ubiquitin markings are 'read' by proteins that specifically recognize each type of modification. (From Alberts et al., 2015.)



eFig. 4.6 Biochemical process of SUMO modifications in mammal cells. All small ubiquitin-like modifier (*SUMO*) paralogues are synthesized as pre-proteins that are first cleaved by a SENP to expose a carboxy-terminal diglycine (*GG*) motif (maturation). An ATP-requiring activation step by the heterodimeric E1activating enzyme (including SAE1 and SAE2) then generates a SUMO–SAE2 thioester. SUMO is then transferred to the E2 conjugating enzyme Ubc9, again forming a thioester. This last step usually requires a SUMO E3 ligase to bring about an isopeptide bond between the SUMO C-terminus and a lysine within the target protein. (From Yang, 2018.)

Proteins involved in ER stress

Cells respond to unfavourable conditions by demonstrating a stress response. For instance, this occurs where there is an excess of free radical generation by the cell but can also occur when normal physiological mechanisms are tested, as in a condition known as ER stress. The cell deals with stress through a variety of mechanisms, including the ER stress responses and a cell signalling system known as the mammalian target of rapamycin (mTOR) response, which is implicated in many conditions involving inflammatory responses in the eye (eFig. 4.7).The specific protein machinery of ER stress is described in the legend.



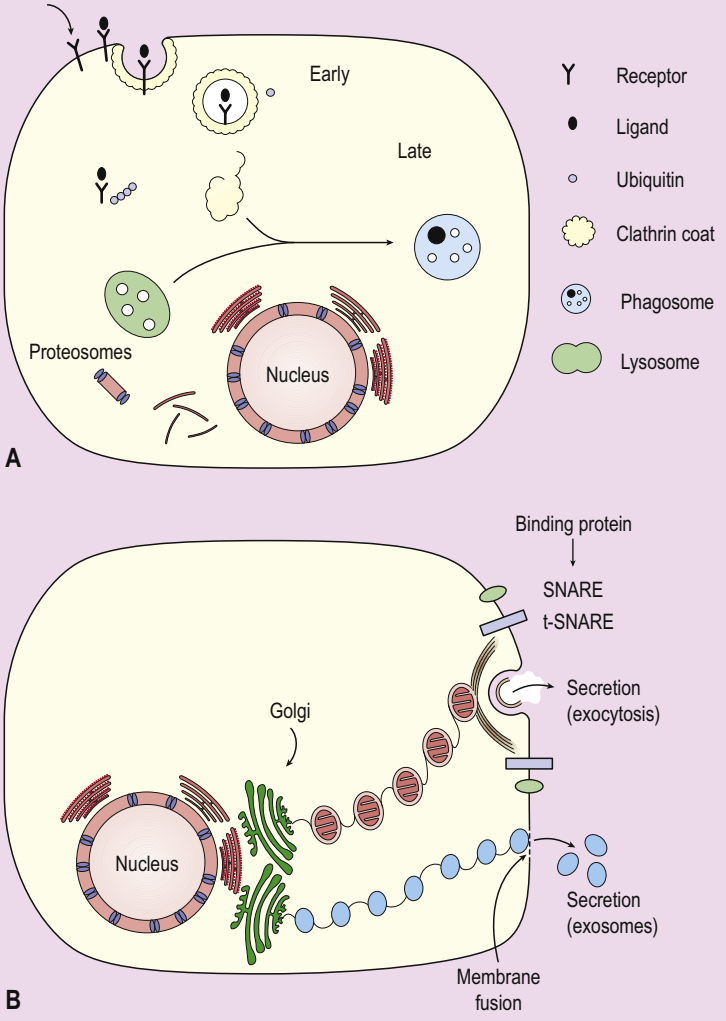
eFig. 4.7 *IRE-1α* is an ER stress sensor that signals to the nucleus to activate the unfolded protein stress response *(UPR)* and maintain cellular physiology. However, if the cell fails to respond, convergent signals through *PERK* activate a nuclear transcription factor *(CHOP)*, which initiates apoptosis. A similar signal can also be initiated via *ATF6α* to activate CHOP. (From Appenzeller-Herzhog and Hall, 2012.)

Box 4.2 **Endocytosis, exocytosis and exosomes**

Endocytosis is generally achieved via incorporation of ligand–receptor complexes in clathrin-coated vesicles. This applies to soluble proteins and to small and large particles such as viruses, which frequently use constitutive cell surface receptors to enter cells.

Clathrin-coated vesicles start as small pits on the cell surface. When the vesicle is fully intracellular it loses its clathrin coat and becomes an endosome, which fuses with primary lysosomes that have a high content of acid hydrolases and other proteases. These lead to degradation of the ingested material and further processing depending on the cell type. Certain cell surface receptors are recycled to the cell membrane during this process to engage further extracellular ligand (**A**). Clathrin-coated pits are normally restricted to the region of the plasma membrane by the cortical cytoplasm actin organization. Relaxation of this actin assembly by proteins such as latrunculin B allows movement of the coated pits.

Exocytosis occurs by a similar process but in reverse. Cytoplasmic granules such as those in exocrine glands or in granulocytes are lipid vesicles containing material for extrusion. Secretory vesicles bud from the Golgi apparatus and are transported towards the plasma membrane by the cytoskeleton. There, they fuse in a lipid microdomain assisted by proteins such as SNARE and the SNARE-binding proteins in a 'targeting patch' (**B**). Fused secretory vesicle membranes act as targets for further secretory granule fusion in cells such as mast cells. In addition, vesicles termed exosomes may be 'pinched' off cells such as macrophages and immune dendritic cells in the same way as platelets are pinched off megakaryocytes. Exosomes convey material (information) from one cell to another and the nature of the information varies from cell to cell.



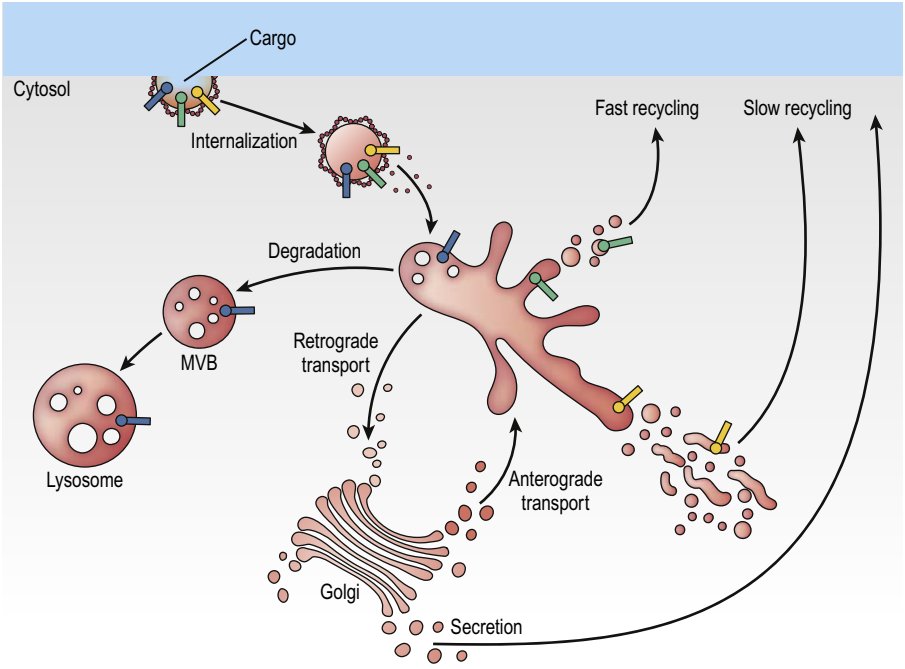


Fig. 4.6 Overview of endocytic pathways. Once internalized from the plasma membrane, membrane-bound vesicles that carry receptors from the cell surface fuse with the early endosomes (EEs). The EE serves as a sorting station from which either tubulovesicular carriers deliver cargo to the endo-lysosomal system for degradation or cargoes are recycled directly or indirectly to the plasma membrane via the endocytic recycling compartment. (From Naslavsky and Caplan, 2018.)

the outer compartment. Their transport systems include antiport, aspartate/glutamate, ornithine/citrulline, maleate/citrate, symport, pyruvate/H^+, and urea and porphyrin synthesis. Mitochondria are the powerhouses of the cell and have several essential metabolic functions as they contain all the elements for the respiratory assembly, for the citric acid cycle and for fatty acid metabolism. Their main functions therefore are to act as the site of energy-rich adenosine triphosphate (ATP)/guanosine triphosphate (GTP) formation, to function as a calcium store mainly in the form of calcium phosphate, to engage in the uptake of energy-rich substances, and to facilitate the oxidative breakdown of ATP. Mitochondria and the ER form an integrated system known as the ER-mitochondria organizing network (ERMIONE) with extensive areas of contact (Fig. 4.7) enabling the transport of lipids and calcium between the two compartments. This structural organization is central to the production of reactive oxygen species (ROS) and calcium signalling.

Mitochondria are powerhouses of nutrient handling and energy storage, essential for cell growth and proliferation. Uncontrolled, this can lead to tumours, and therefore a balance between homeostatic metabolism and cell death is essential. Mitochondria are therefore regulators of cell death through several routes (Fig. 4.8).

The central inducer of cell death is mitochondrial cytochrome c, which is released into the cytoplasm and complexes with Apaf-1 and caspase 9 (the apoptosome) and ultimately leads to caspase 3 and 7 as the final executioners. More recently, mitochondria have been recognized as sensors and regulators of cytosolic Ca^{2+} levels, through Ca^{2+} transporters, which have central physiological roles in signalling, proliferation, metabolism and ultimately cell survival.

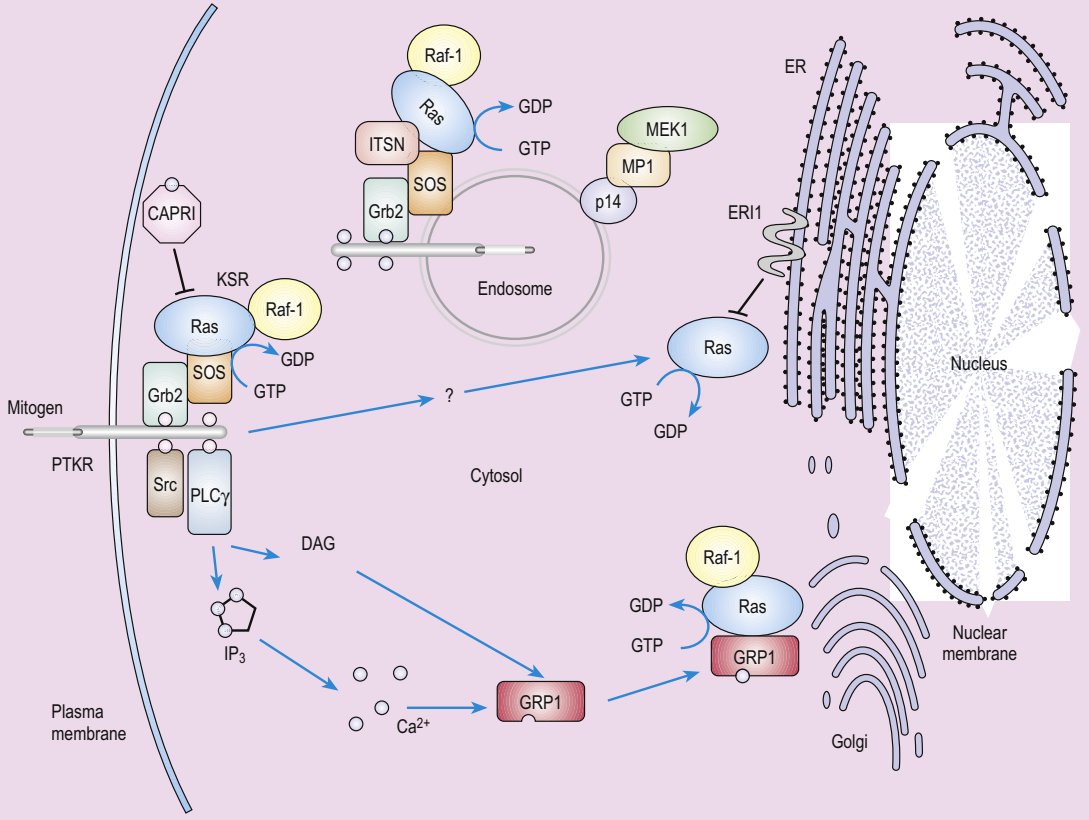
There are many 'death receptors' (receptors which initiate the cell programme leading to death of the cell, i.e. apoptosis), including FasL, TNFR and granzyme B, in addition to environmental conditions that promote cell death, such as hypoxia, genotoxic damage, cytokine or nutrient deprivation (linked to the tricarboxylic acid cycle and cytochrome c) (Fig. 4.9), excess steroid exposure, UV exposure and toxic drug exposure, and the balance between life and death is maintained by pro- and anti-apoptotic Bcl-2 proteins (see p. 174).

Mitochondria have their own complement of DNA (but no histones, and thus are not susceptible to epigenetic changes; see

Further information on a mechanism of how induced epigenetic changes can lead to heritable functional effects is available at https://expertconsult.inkling.com

Box 4.3 **Schematic representation of how mitogens activate cells**

Ligand binding to the tyrosine kinase receptor (PTKR) activates the small GTPases, Ras, through a complex of signalling molecules (Grb2/SOS). Activated Ras on endosomes from the plasma membrane induces signalling in this organelle as well as in the Golgi apparatus (indirectly via phospholipase CgCa and another protein known as GRPI) while inhibition of the small Ras occurs in the endoplasmic reticulum via an inhibitor protein ERII (see figure). In this example of a signalling network, interstin and kinase suppressor of Ras act as scaffold proteins and p14 acts as an adaptor protein in Ras-independent activation of the kinase, MEK-1 (mitogen-activated protein (extracellular signal-regulated [ER kinase] kinase), by the endosome. The integration of organelles and signalling networks with cytosolic proteins is thus central to proper functioning of the cell in response to an external stimulus such as a mitogen.



and eFig. 4.1) as well as ribosomal RNA/transfer RNA, and generate a series of mitochondrion-specific proteins associated with mutations and a number of discrete syndromes, some with ophthalmological consequences (see Chapter 3, p. 155). Since mitochondria originate only from ova, transmission of these genetic defects is purely maternal.

Mitochondria can also become damaged and are removed by a process termed mitophagy, similar to autophagy (see next section), which involves mitochondrial receptors binding to phagophore LCR proteins leading to engulfment. Alternatively, mitophagy can occur via a process of ubiquitinylation involving adaptor proteins such as optineurin (Villa).

Other cytoplasmic organelles

The cell contains many other cytoplasmic organelles including various types of endosome, lysosomes, phagosomes, autophagosomes, tubular-vesicular structures and numerous multivesicular bodies (MVB). Autophagosomes are

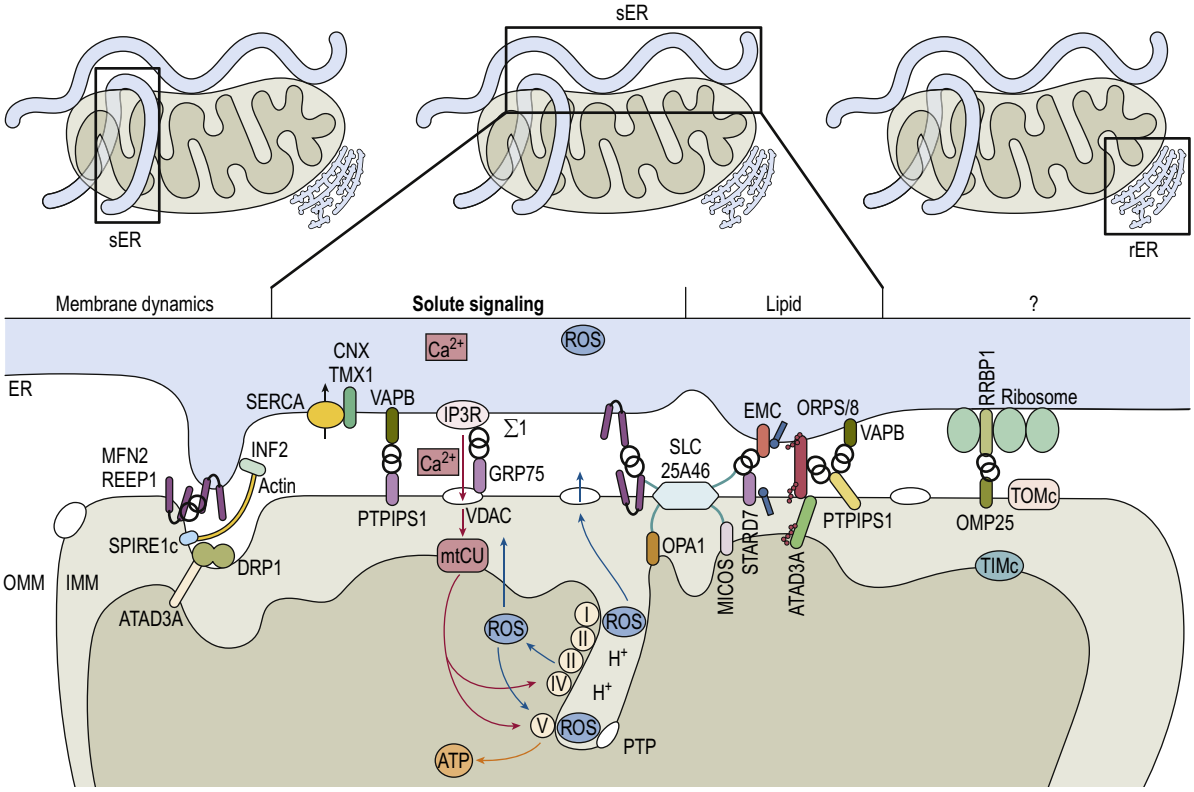


Fig. 4.7 Endoplasmic reticulum (*ER*) mitochondrial contacts assist through increased membrane dynamics in signalling via lipid transfer. *Top*, schematized mitochondrion with a fission groove engaged in various contacts with smooth ER tubules and a rough ER sheet. *Below*, the molecular contributors of the main known functional interactions at ER–mitochondrial contacts are summarized. For bi/multipartite linkers the coupling domains are shown as *grey clasps*. Local Ca^{2+} fluxes are indicated with *red arrows* while main mitochondrion-derived reactive oxygen species (O_2^-, H_2O_2) fluxes with *blue arrows*. *ATAD3A*, ATPase family AAA domain-containing protein 3A; *CNX*, calnexin; *DRP1*, dynamin-1-like protein; *EMC*, ER membrane protein complex; *ER*, endoplasmic reticulum; *GRP75*, glucose-regulated protein 75; *IMM*, inner mitochondrial membrane; *INF2*, inverted formin 2; *IP3R*, inositol trisphosphate receptor; *MFN2*, mitofusin 2; *MICOS*, mitochondrial contact site and cristae organizing system; *mtCU*, mitochondrial Ca^{2+}uniporter; *OMM*, outer mitochondrial membrane; *OPA1*, optic atrophy protein 1; *ORP*, oxysterol-binding protein-related protein; *PTP*, permeability transition pore; *PTPIP51*, protein tyrosine phosphatase interacting protein 51; *REEP1*, receptor expression-enhancing protein 1; *rER*, rough ER; *ROS*, reactive oxygen species; *RRBP1*, ribosome-binding protein 1; *sER*, smooth ER; *SERCA*, sarcoplasmic reticulum calcium transport ATPase; *SLC25A46*, solute carrier family 25 member 46; *SPIRE1c*, spire type actin nucleation factor 1; *STARD7*, steroidogenic acute regulatory protein-related lipid transfer domain containing protein 7; *TIMc*, translocase of the inner membrane complex; *TMX1*, thioredoxin-related transmembrane protein; *TOMc*, translocase of the outer membrane complex; *VAPB*, vesicle-associated membrane protein-associated protein B; *VDAC*, voltage-dependent anion channel. (From Csordas, 2018.)

housekeeping structures that remove and recycle exhausted or damaged components by fusing with lysosomes to form a phagosome, similar to phagosomes, which form after engulfment of infectious bacteria in macrophages. Autophagosome formation (autophagy) is thus a normal physiological process central to the health of the cell and regulated by specific genes (Atg genes) and proteins including optineurin (OPTN), mutations in which are associated with glaucoma (see Chapter 9, p. 516). One set of proteins (LC3 /Atg8) is involved in the docking of the cellular debris and closure of the autophagosome after fusion and is used in studies as a biomarker. Autophagy is a prominent feature of programmed cell death (apoptosis).

Recently, another type of organelle has been described, namely membraneless organelles (MLOs) (Fig. 4.10). MLOs measure 0.1–3.0 μm and are considered regions of liquid/liquid phase separation where proteins segregate into complexes in the cell. The nucleolus (see later)

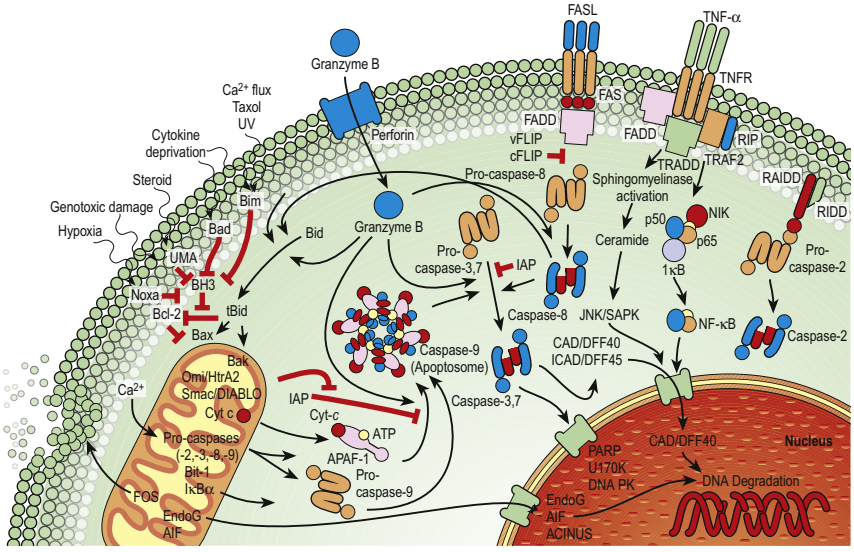


Fig. 4.8 Mitochondria and apoptosis: the signalling pathways. Cytochrome c *(red spots* in Fig. 4.9*)* in mitochondria controls many apoptosis pathways involving caspases, granzymes and perforin molecules as well as several pro- and anti-apoptotic genes such as Bax, Bcl2, BH3, Bid and Bim. Many of these, plus numerous other proteins, form a complex known as the apoptosome.

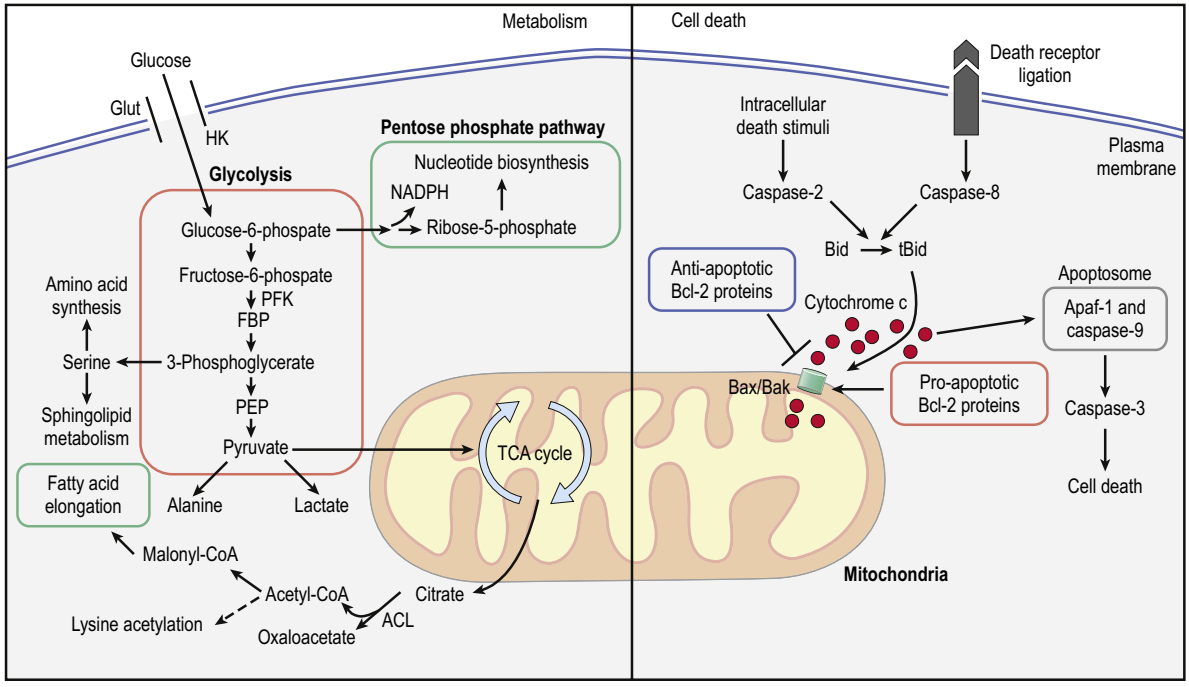


Fig. 4.9 Cell metabolism and cell death (apoptosis) are intricately linked through the tricarboxylic acid *(TCA)* cycle. Cross-talk in the *mitochondria* between the signalling pathways for these two systems provides cross-regulation and determines whether a cell will live or die. *Cytochrome* c is central to both pathways in generation of ATP for energy and caspase-1 for apoptosis. (From Andersen and Kornbluth, 2013.)

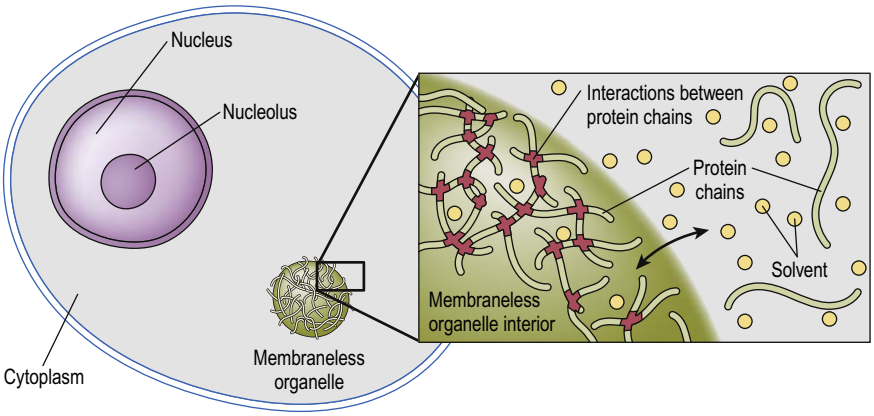


Fig. 4.10 *Membraneless organelles* are present in the cell, creating foci where interactions can occur between cognate molecules in the relative absence of interference from other cellular constituents. They develop through phase separation similar to that of oil and water but less definitive. About 20 such organelles have been described. (From Crabtree and Knott, 2018.)

represents one such organelle and there are thought to be around 20 similar structures in the cytoplasm. They are sites of chemical reaction where interacting proteins can be brought together mainly to control reaction rates and avoid interference from other cellular constituents or toxic reactions. Small molecules are allowed to enter MLOs depending on how they interact with MLO proteins. Examples of the sort of activity that might be undertaken by MLOs include unwinding and separation of DNA double helices, which has been demonstrated *in vitro*. Protein aggregates in MLOs may also participate in disease pathology as for instance in Alzheimer's disease (see Chapter 9, page 512).

The nucleus

The nucleus is contained within a perforated sac, formed by pores (nuclear pore complexes). These allow transport of proteins and other signalling molecules passage between the transcriptional machinery and the appropriate instructions from outside. Both import and export of materials are receptor-mediated in an 'address-label'-like system based on small GTPases (see Box 4.3).

The main nuclear component is chromatin, a complex structure of highly extended DNA, RNA and protein in the interphase (non-dividing) cell, which becomes greatly condensed (400-fold) to form chromosomes during cell division. Packing of chromatin is achieved by interaction between negatively charged DNA and certain basic proteins (histones), which carry a positive charge at the pH of the cell. The basic packing unit of chromatin is the nucleosome in which a segment of DNA is coiled round eight histone proteins and their position in the genome determines access of regulatory proteins to DNA. Where DNA coiling occurs, chromatin 'beads' form, known as chromomeres. Chromomeres contain 500–1000 nucleosomes (Olins and Olins, 2018).

Euchromatin is less packed than heterochromatin, the proportion of which varies from nucleus to nucleus and may be characteristic of certain cell types, e.g. the 'clock-face' heterochromatin characteristic of plasma cells. Genome stability is dependent on minimizing DNA damage, e.g. during cytokinesis and chromatin condensation, and chromatin regulators exist to ensure stability. Histone-modifying enzymes include acetyl and methyl transferases and deacetylases. which modify genes via reversible nucleotide methylation and acetylation.

The nuclear membrane also contains receptors for ligands, which may be synthesized in the cytoplasm or may have been endocytosed through plasma membrane receptors. Typical nuclear membrane receptors include steroids, growth factors such as fibroblast growth factor and novel groups of proteins known as the peroxisome proliferator-activated receptors (PPARs), which are involved in many cell processes such as lipid and glucose homeostasis, wound healing, and inflammation generally. PPARs are unique receptors that allow integration of signals mediated by lipophilic ligands with plasma membrane-derived signals. Nuclear membrane receptors may be organized for induction or suppression of genes in a coordinated fashion (Fig. 4.11). In addition, a major function of the nucleus is to manage the packaging of mRNA into ribonucleoprotein particles called mRNPs, a process mediated by specific protein complexes known as internal ribosome entry elements. These complexes contain very large numbers of proteins involved in the transcription of mRNA in the nucleus to translation and degradation of mRNA in the cytoplasm (Fig. 4.12). The USA National Institutes of Health has set up the 4D Nucleome Program to understand the principles

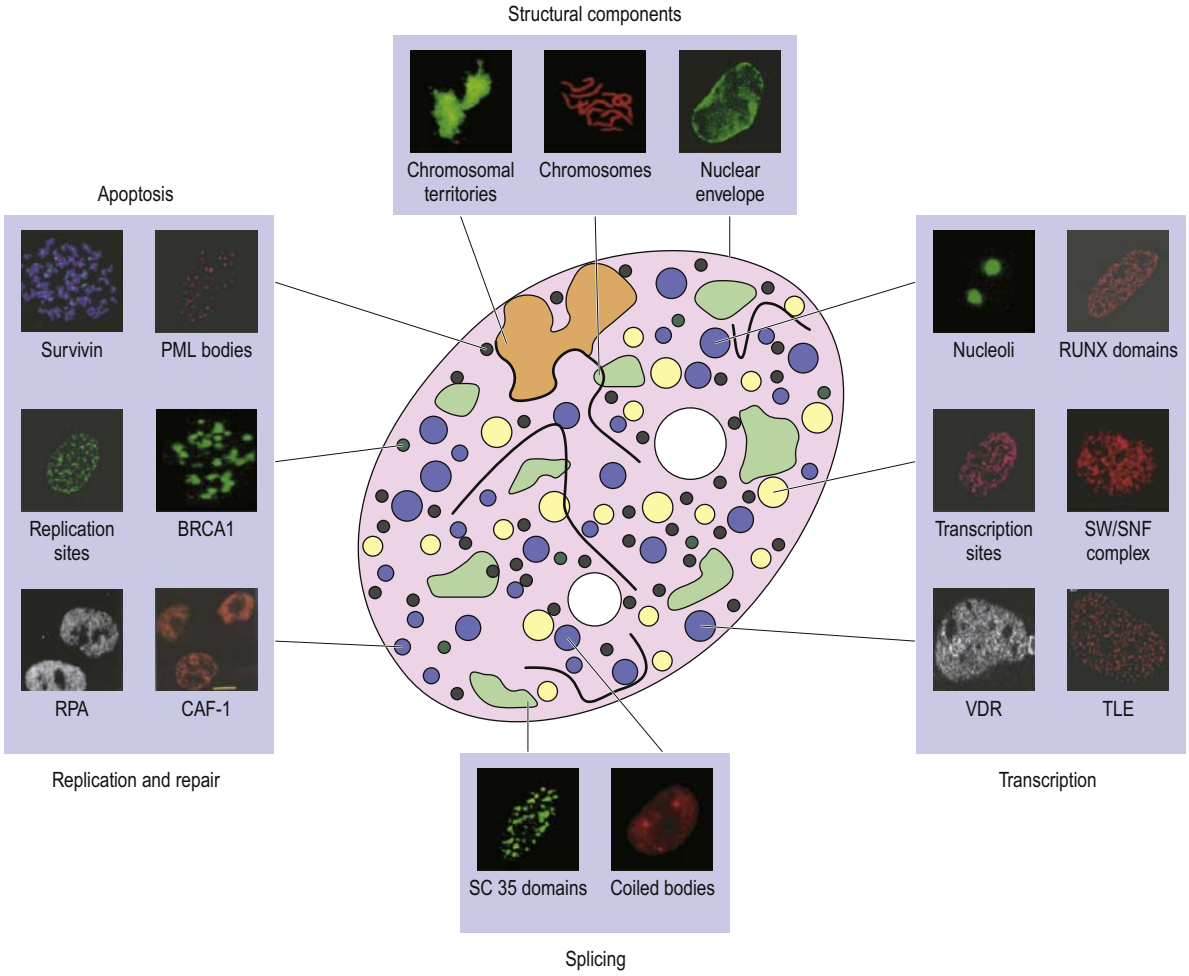


Fig. 4.11 Subnuclear organization of different nuclear structures. The nuclear architecture is functionally linked to the organization and sorting of regulatory information. Immunofluorescence microscopy of the nucleus *in situ* has revealed the distinct non-overlapping subnuclear distribution of vital nuclear processes. (From Stein et al., 2003, with permission from Elsevier.)

behind the three-dimensional organization of the nucleus in space and time (the 4th dimension) and how gene expression reflects the organization of its components.

The nucleolus is an MLO (see page 175) composed of RNA and a fibrillar matrix and is the site of ribosomal RNA synthesis and intense transcriptional activity. In addition, it may contain non-coding regions of RNA involved in the stress response. The nucleolus develops during the late stages of mitosis in association with specific regions on the chromosomes, known as the nucleolar organizer centre.

The intracellular matrix

The cytoplasm is a highly viscous aqueous medium that has deformability (elasticity). Physically, it exists at different times as a gel or as a sol. The cortical cytoplasm (ectoplasm) is more akin to a gel structure, while the endoplasm is usually more fluid. Thus, the cortical cytoplasm restricts movement of organelles such as coated vesicles (see Box 4.2). The gel-like properties of the cytoplasm are the result of the binding of 'structured' water molecules and Ca^{2+} ions to polymeric filaments; in fact, water, as the main constituent of cells, has a significant role even within the molecular interstices of cell membrane receptors where in its 'ordered' state it contributes to the tertiary structure of the molecule as shown by rhodopsin molecules in the photoreceptor discs (Fig. 4.13).

There are four main types of intracellular filaments: microfilaments, such as actin; intermediate filaments; group 3 cytoskeletal fibres, such as tubulin and myosin; (see Fig. 4.14) and septins.

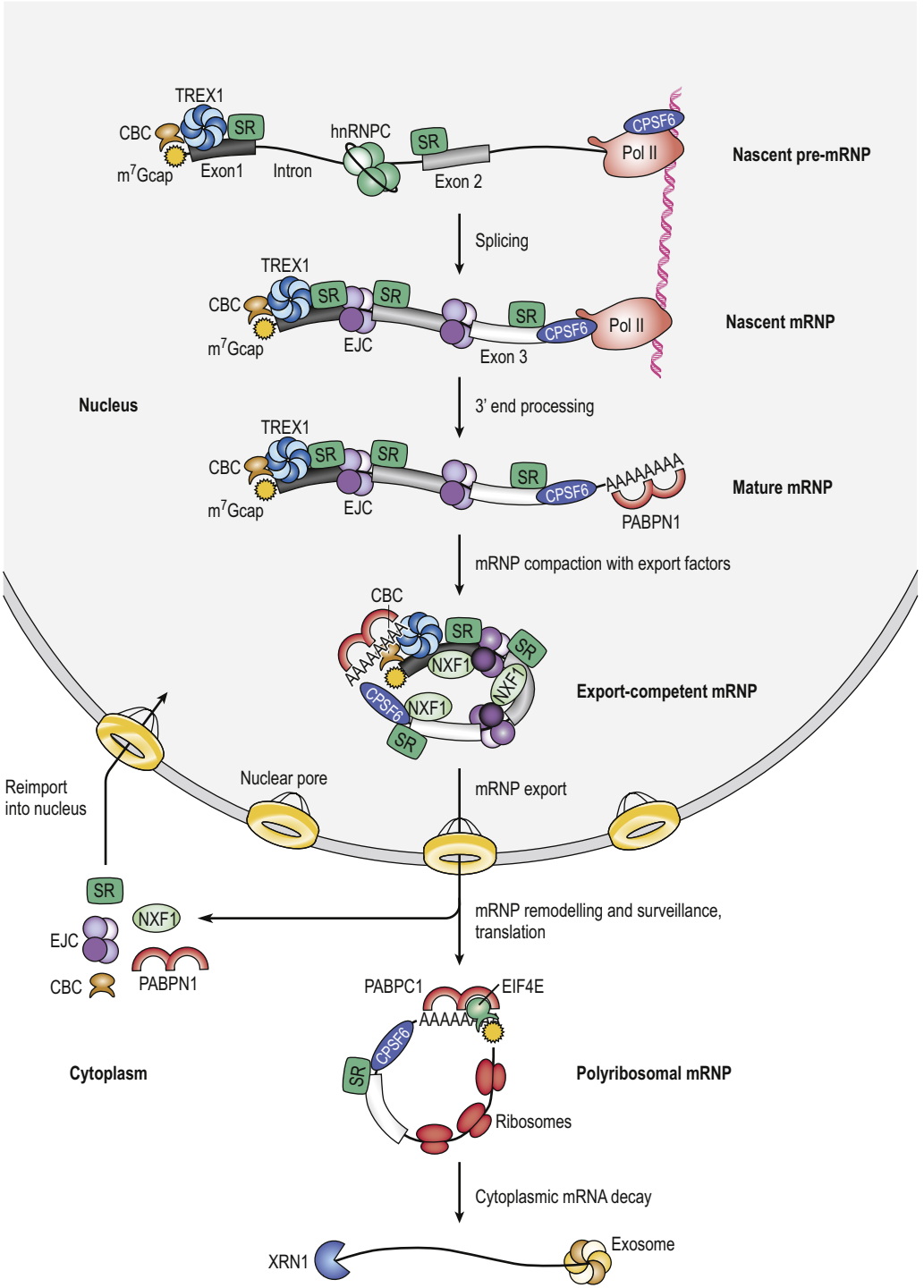


Fig. 4.12 Summary diagram of how messenger ribonucleoprotein particles are assembled and remodelled. It is now realized that there is considerable diversity in how these molecules are generated. In the diagram one example is shown of an RNP cap binding complex *(CBC)* facilitating splicing of the 3′-end formation and export together with bound mRNAs through the nuclear pores into the cytoplasm. Finally, it is then bound to cytoplasmic proteins in ribosomes and becomes actively translated mRNA. (From Muller-McNicoll and Neugebauer, 2013.)

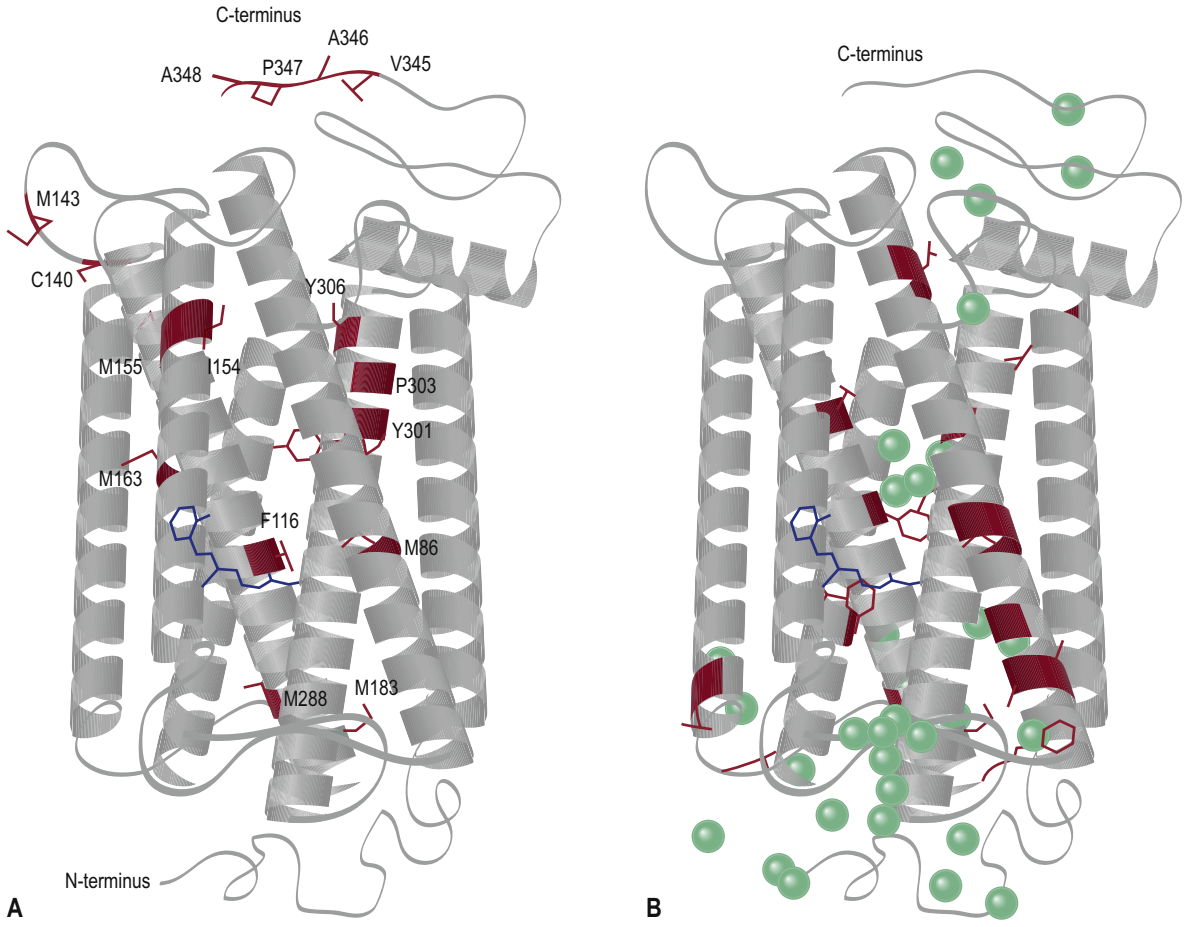


Fig. 4.13 Ribbon diagram of the rhodopsin molecule as it flows in the photoreceptor disc membrane. Vitamin A bound to rhodopsin is shown in *blue* in (**A**). In (**B**), water molecules represented as *green spheres*, interact intimately with the binding of vitamin A to rhodopsin and are bound and released in concert with vitamin A as it is bound and released from the rhodopsin scaffold. It is believed that the water molecules modify, if not determine, how vitamin A interacts with rhodopsin. (From Orban et al., 2010.)

Microfilaments. Microfilaments (5–7 nm), such as actin, tropomyosin and troponin, are universal constituents of cells and are involved in almost every cellular activity, including cell motility, contractility, endo- and exocytosis (see Box 4.2) and maintaining structural integrity, where they insert into cell adhesion junctional complexes (cadherins).

Further information on cadherins, filaments and cell junctions is available online at https://expertconsult.inkling.com

Actin bundles are brought to a highly developed level in muscle cells. Actin occurs in several forms within the cell, depending on its associated protein, e.g. as a fine lattice meshwork or as a sheaf of fibres (stress fibres). Stress fibres are often the most prominent cytoskeletal features in a cell: they not only generate force but they respond to mechanical tension by reinforcement and in some respects are similar to sarcomeres of muscle cells.

The cellular distribution and, particularly, the degree of polymerization of microfilaments are determined by the nature of proteins that bind to them, i.e. the microfilament-associated proteins (Table 4.1). Monomeric soluble actin (G-actin) is converted to gel-phase polymerized fibrils (F-actin) by association with certain proteins. In smooth muscle cells and fibroblasts, filamin assists polymerization; in contrast, in lymphocytes, profilin and thymosin maintain G-actin in the depolymerized state, presumably to facilitate flexibility in cell shape during rapid migration

Filaments and junctions

Intercellular connections or junctions are structures that make cells into tissues and tissues into organs. There are several types, each of which is composed of junction-specific proteins. These are described later and illustrated in eFig 4.8 (A–C).

(A) Tight junctions

Tight junctions have no detectable 'space' between the cell membranes. They are also known as zonulae occludentes; they form a barrier to paracellular diffusion of all molecules, including water and ions. They are also involved in regulation of epithelial cell proliferation and differentiation. They occur at such sites as the blood–aqueous barrier of the ciliary body and the blood–retinal barrier at the apex of the RPE cell. They consist of a system of ridges and grooves, as seen by freeze–fracture studies. There are four major classes of tight junction proteins:

- transmembrane proteins: occludin, claudin
- adaptors: ZO-1, cingulin, MUPPI
- transcriptional and post-transcriptional regulators: AP-1
- signalling proteins: αPKC, CDK4.

Claudins have unique abilities to selectively permit transport of charged ions and vary from tissue to tissue.

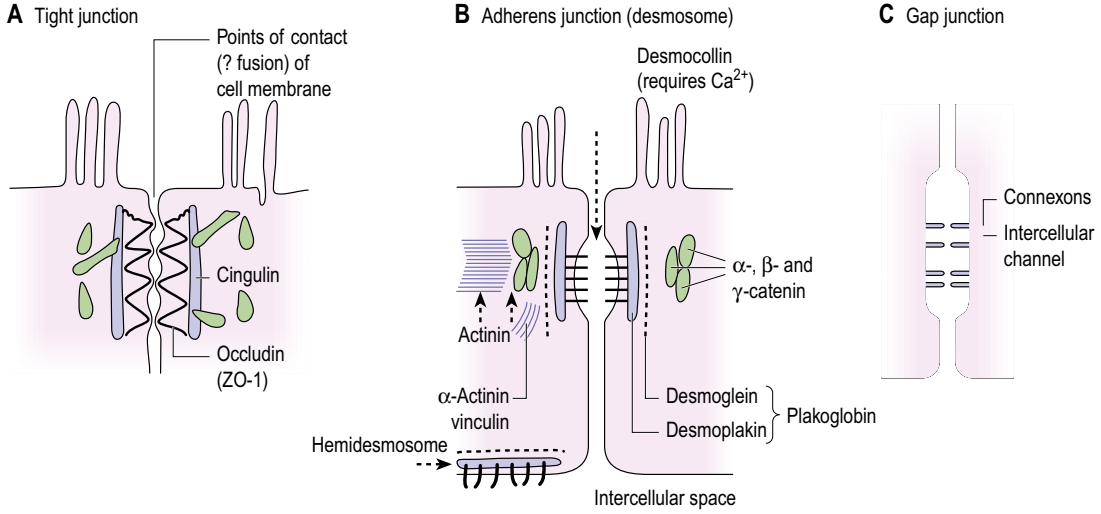
(B) Adherens junctions

Desmosomes form specialized adherens junctions of 20 nm width between cells. Two types are described: spot desmosomes (at single site) and belt desmosomes (as a ring round the apex of the cell). The latter are also known as zonulae adherens. Their probable role is mechanical adhesion: cytoplasmic filaments insert into spot (cytokeratin) and belt (actin) forms. Several proteins are involved in different regions of the desmosome, including desmocollin, desmoglein and desmoplakin, forming a subfamily of cadherins. The cadherins bind to β-catenin and then to actin-based cytoskeletal proteins such as vinculin and α-actinin, ZO-1 and actin itself. The function of tight junctions (see previous section) is dependent on the integrity of the adherens junctions. Hemidesmosomes have a similar plaque formed in this case by bullous pemphigoid antigen (BPAG1).

(C) Gap junctions

Gap junctions are so-called because there is a 2-nm gap at this adhesion site between cells. Gap junctions occur in the basal regions of epithelial and other cells. They are a highly organized structure composed of 'connexons', plasma membrane domains containing connexins, which have a role in permitting the passage of larger ions such as Ca^{2+}, and other signalling molecules, permitting a coordinated response by a group of cells, as in ocular or cardiac muscle (**C**).

Synapse: a specialized form of junction between nerves, or between nerves and muscles, and characterized by synaptic vesicles in the axon terminal and pre- and postsynaptic thickening of the plasma membrane (see Chapter 5, p. 315, Fig. 5.29). In addition, the immunological synapse (see Chapter 7, p. 459) between two immune cells has many similarities to a neurological synapse.



eFig. 4.8 The three classes of cell-to-cell adhesions.

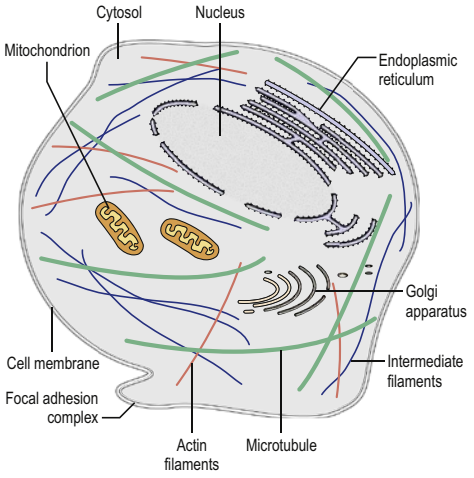


Fig. 4.14 Line diagram showing the major cellular organelles and structural proteins. Organelles such as *endoplasmic reticulum* and *mitochondria* are now known to make physical contact, to become tethered to each other and to transfer information including proteins and lipids across the tethers acting like 'gangways between ships'.

Table 4.1 **Function of actin-binding proteins**

Function	Protein	Cell/structure
Gelation	Filamin	Smooth muscle/ fibroblasts
Bundling	α-actin	Muscle
	Fimbrin	Microvilli
	Talin	All cells
Severing	Gelsolin	Macrophages
	Villin	Microvilli
	β-actinin	Skeletal muscle
Depolymerizing	Profilin	Lymphocytes
	Thymosin	All cells
	Actobinin	All cells
Membrane-binding	Vinculin	Adhesion sites
	Spectrin	Red cells, photoreceptors
Receptor transport	Capping proteins	Leucocytes
Junctional complex	Radixin	Liver cells

within tissues. The specificity of these actin-binding proteins is remarkable; examples include the ankyrin–spectrin–actin combination, which determines the red cell biconcave shape, and the spectrin–peripherin–actin combination between rod outer segment discs and plasma membrane. Actin polymerization and depolymerization is a highly regulated process requiring addition of actin monomers to one end of the microfilament and removal at the opposite end, each of which has separate K_{on} and K_{off} constants. This process is under the control of actin-depolymerization factor (cofilins), which alters these rate constants as necessary, leading to changes in the twist of the molecule and promoting severance of the actin filament.

Regulation of coordinated changes to the actin cytoskeleton is under the control of intracellular enzymes, particularly the Rho family of GTPases (this includes enzymes such as Rho, Rac and Cdc42, the last of which is a cell cycle-related protein). Certain kinases, known as the p-21 activated kinases (PAKs), are involved in regulating some of the diverse changes induced by Rac and Cdc42. PAKs may determine such cellular responses as polarity in epithelial cells and motility in fibroblasts. For example, coordinated functional regulation of the interaction between dynamin, cortactin, actin-binding protein 1, neuronal Wiskott–Aldrich syndrome protein 1 (N-Wasp-1), profilin and actin occurs during the pseudopodial extension and vesicle formation of migrating fibroblasts (Fig. 4.15). Central to this process is the ARP 2/3 complex, which is required for actin nucleation. For instance, the Arp 2/3 complex is also involved in a relatively recently described family of actin-binding proteins, the coronins, which are involved in leucocyte migration.

Several isoforms and at least three families of actin exist (α, β, γ) and some of these are located in specific parts of the cell, e.g. β-actin in the cell cortex.

Intermediate filaments. Intermediate filaments (10–12 nm) are coiled α-helices, which act as stretchable components of the cytoskeleton scaffold. They occur in cell-, tissue- and differentiation-specific distribution in both cytoplasmic and nuclear compartments of the cell and are classified into five groups depending on domain and sequence homology. Their major function is to protect cells from mechanical and non-mechanical damage (Fig. 4.16), and gene mutations account for around 30 different diseases in humans, mostly related to skin, muscle and nerve dysfunction. In general, intermediate filaments are important for the correct positioning and function of cell organelles such as mitochondria and ER.

There are five classes of cytoplasmic intermediate filaments, often used to characterize cells in tissue culture or tumours, and one class of nuclear intermediate filaments:

- *keratins:* found in epithelial cells; over 50 individual members grouped into two types, I and II, exist as heterodimers
- *vimentin:* found in mesenchymal cells
- *desmin:* interconnects myofibrils of muscle cells, at site of Z disc and M line, thus maintaining their register
- *glial fibrillary acidic protein* (glia)
- *neurofilaments:* e.g. S100 protein in neuroectoderm; connect with microtubules via small projections

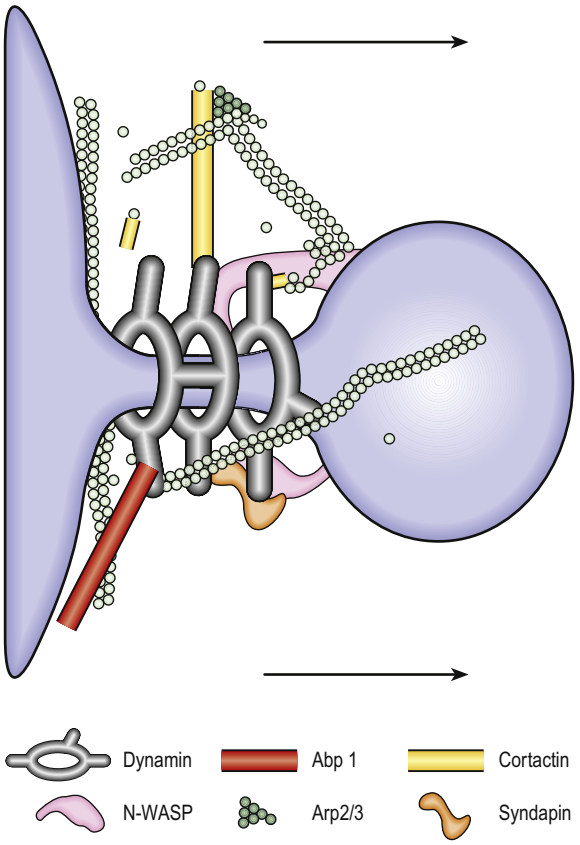


Fig. 4.15 Dynamin-mediated tubulation and vesiculation of secretory and endocytic compartments. Model showing dynamin-mediated tubulation/vesiculation of a membrane as it might occur at the plasma membrane. Dynamin, its binding protein and associated proteins are likely to function together during vesicle formation. The molecular 'pinchase' activity of dynamin, together with the enhanced actin filament nucleation at the membrane interface, results in the tubulation and severing of the membranous vesicle necks (see text for details). Large *black arrows* indicate the generation of force and the movement of nascent vesicles. (From Orth and McNiven, 2003, with permission from Elsevier.)

- *lamins:* line the inner surface of the nuclear envelope as a fibrous lattice (karyoskeleton); more than five types.

One function of intermediate filaments is to ensure normal cytoplasmic positioning and function of different organelles. Lamins undergo considerable molecular disruption during mitosis and may communicate with cytoplasmic intermediate filaments, but how this occurs is not clear. Others, such as keratins, provide mechanical strength to junctional structures such as desmosomes.

Further information on cadherins, filaments and cell junctions is available online at https://expertconsult.inkling.com

Cellular junctions are highly specialized: desmosomes and hemidesmosomes (junctions between the cell and basement membrane in epithelia) have some ultrastructural similarity, but at the molecular level there are clear differences between the constituent proteins. Desmosomes are formed by a series of proteins spanning the cell membrane and the intercellular space (desmoglobin, desmoplakin), while hemidesmosomes contain other proteins such as the $\alpha_6\beta_4$ integrin receptor, which binds to laminin in the basement membrane, plus other proteins such as the bullous pemphigoid and the pemphigus antigens.

They may also have a function in the positioning of the nucleus in the cell in a cage-like bundle of fibrils. Intermediate filaments maintain cell homeostasis by dealing with stress in its various guises (Fig. 4.17) and the cell responds in a number of ways such as generating 'inclusion bodies', producing new and regenerating intermediate filaments, or by reorganizing the filaments into variously sized bundles.

Interactions between intermediate filaments and microfilaments are mediated by plectin, a >500 kDa dumb-bell-shaped protein that can self-associate and in addition can bind at both ends of the hemidesmosomal protein $\alpha_6\beta_4$ integrin and probably to other junctional proteins.

Group 3 cytoskeletal fibres. A third group of cytoskeletal fibrillar elements also exists, but the fibres are less easy to categorize. Several thick filaments occur as part of the cytoskeleton, such as myosin (myosin comprises 25% of cytoplasmic protein in striated muscle) and microtubules. Microtubules are cylindrical structures about 24 nm wide, comprising 13 globular elements composed of the heterodimer ($\alpha\beta$) tubulin. Stable microtubules occur in flagellae and cilia, while labile microtubules are found in structures such as the muscle spindle. Microtubules are involved in movement and cell motility, including intracellular transport as in axoplasmic flow. They are often required to switch very rapidly from an extending to a shrinking fibril, as in processes such as cell division where a multiprotein structure termed the kinetochore engineers chromosome segregation by facilitating chromosome attachment to spindle microtubules. Rapid growth to shrinkage in the microtubule is achieved through GTP to GDP hydrolysis, mediated through the Rho family of small GTPases; this is known as 'catastrophe', while the reverse change is termed 'rescue' (Fig. 4.18).

These changes occur in any microtubule-related function, including cell movement, and drugs such as colchicine (used in Behçet's disease) and taxol (proposed as prophylaxis for proliferative vitreoretinopathy, see Chapter 9, p. 498) disrupt microtubule organization and inhibit cell motility and cytokinesis.

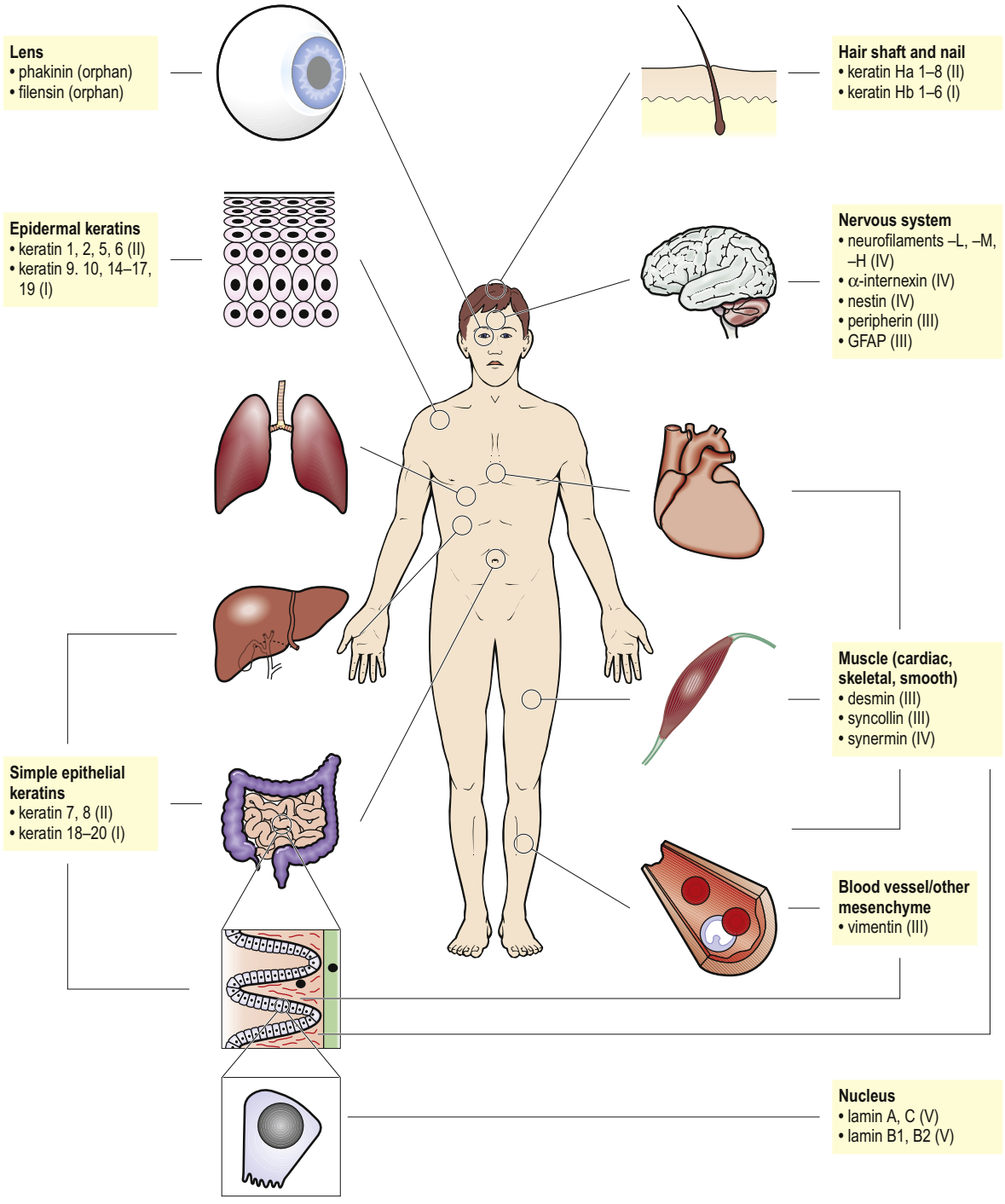


Fig. 4.16 Distribution of intermediate filament (IF) proteins in the human body. IF proteins include five major types (I–V) and a separate 'orphan' category (IF type and category are listed in parentheses). Lamins (type V) are found in the nucleus of most mammalian cells, whereas the remaining IFs (types I–IV) are cytoplasmic and expressed in a cell-tissue-selective manner. For each tissue, representative major IFs are listed for the principal cell type. (From Toivola et al., 2005, with permission from Elsevier.)

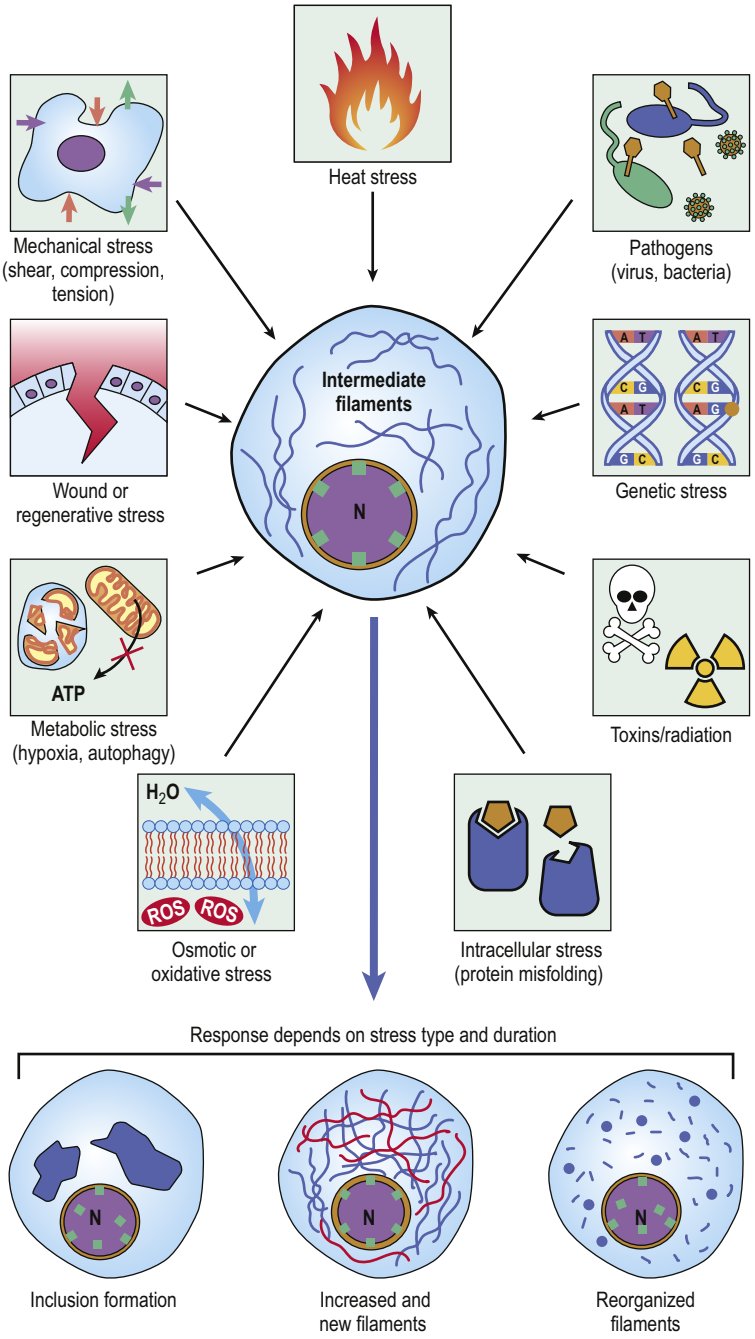


Fig. 4.17 Cells respond to stress through their intermediate filaments by (a) forming inclusion bodies, (b) making new filaments or (c) reorganizing their filaments. (From Toivola et al., 2010.)

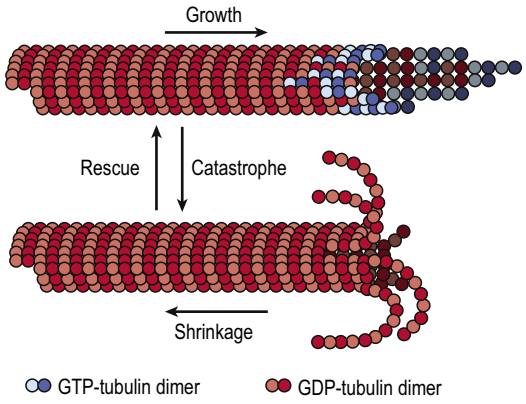


Fig. 4.18 Microtubules switch between phases of growth and shrinkage generated by GDP-GTP-mediated incorporation and removal of tubulin dimers in the region of the tubulin cap. (From Bowne-Anderson et al., 2013.)

Group 3 fibres are thus a class of superfine filaments which combine to form an intracellular meshwork in which proteins do not exist in solution as previously surmised, but are attached to and transported along the filaments with other structures such as 'free' ribosomes and small vesicles (polysomes). This arrangement has special relevance to highly organized cells such as lens fibres. Transport of molecules along microtubules involves two families of molecular motors, kinesins and dyneins.

Certain discrete cellular structures are composed of microtubules, such as the centriole (the microtubular organizing centre, MTOC, a cylindrical structure comprising nine groups of triplet microtubules) and the mitotic spindle. Cilia are remarkably constant structures in eukaryotic cells, composed of nine peripheral and two central bundles of three fused microtubules. Movement occurs by sliding of the outer arm of dynein along the core of tubulin.

Septins. Septins are a family of proteins expressed in all cell types that are integrators of functions of other microfilaments such as the ATPase role of actin and the dynamic functions of microtubules. There are 13 septins in humans, mutations of which are known to cause specific diseases. Septins are important in cells with high migratory functions, such as lymphocytes. In addition, septin 8, which lies on chromosome 5q31, is associated with retinal degeneration. Their main function in the cell appears to be supportive, e.g. as a scaffold for protein recruitment and as an intracellular compartment organizer, and they are called septins for this reason. For example, several septins in one cell combine to form a hexameric filament of around 25 nm long, which acts as an intracellular 'septa' (barrier to diffusion) (Fig. 4.19).

In conclusion, therefore, there are extensive interconnections between the four filamentous systems orchestrating many fundamental types of cell behaviour. For instance, in cardiac muscle, the close communication between cytoplasmic filaments and nuclear lamins through linker proteins such as plectin have been revealed in detail (Fig. 4.20). The cell cytoplasm also contains several storage products such as glycogen granules, lipid droplets and melanin in melanosomes, often inside endosomes, which, as indicated above, are transported around the cell by MT-associated proteins and made to function by active fibre contraction as in actin-mediated exocytosis (see Box 4.2). With age, some cells, such as the RPE cell, accumulate unwanted intracellular bodies such as lipofuscin granules.

Intracellular signalling mechanisms

As stated above, cells respond to external stimuli by means of cell surface receptors, which convert the external stimulus to a series of intracellular signals (second messengers) directed towards specific cellular functions such as protein transcription for growth control or ion-channel gating in neural responses (see Chapter 5, p. 294, Box. 5.6).

Second messenger systems are based on a network of reactions involving an agonist, a receptor and an interacting set of coupled proteins. Cyclic adenosine monophosphate (cAMP) is the archetypal second messenger, and the result of such a response is the phosphorylation of a regulatory intracellular protein via a kinase, such as serine-threonine kinase or tyrosine protein kinases, which demonstrate great substrate specificity (Fig. 4.21). In contrast, the phosphatases, such as protein phosphatase-1 (PP1), which usually bring the kinase reaction to a close, are much more widespread in their range of activity and, consequently, have many fewer genes expressed.

During the last 30 years, a family of ligand receptors known as WNTs has been shown to have widespread roles in embryonic development, cancer and wound healing. These molecules define the interaction between matrix and cells via intra- and inter-cell signalling and are essential for developmental processes such as patterning. The name WNT derives from the discovery that the int1 (integration gene) and the (Wg) Wingless gene in *Drosophila* were identical and there are now known to be several Wnt genes with multiple effects on extracellular matrix proteins as well as on many cell functions, including immune cells where this signalling pathway promotes immunological tolerance (see Chapter 7).

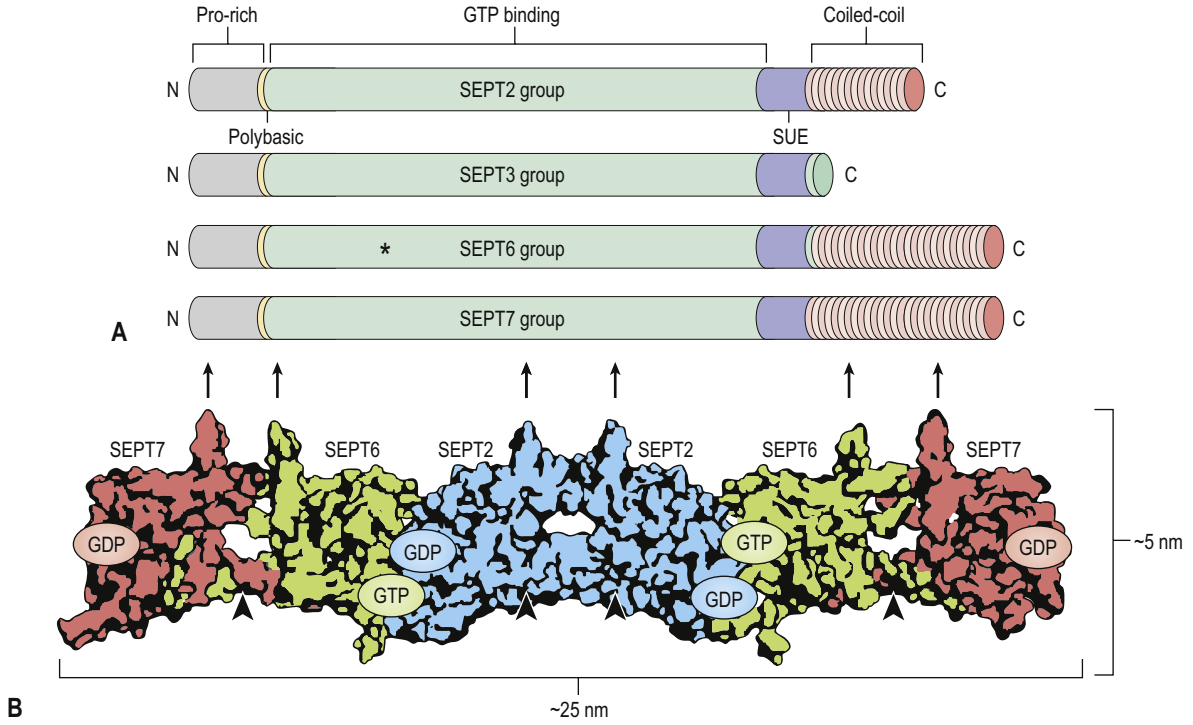


Fig. 4.19 There are 13 human septins classified into four groups *(SEPTS 2, 3, 6 and 7)*. They have three conserved domains: a phosphoinositide-binding region; a *GTP-binding region*; and the septin unique element *(SUE)*. (From Mostowy and Cossart, 2012.)

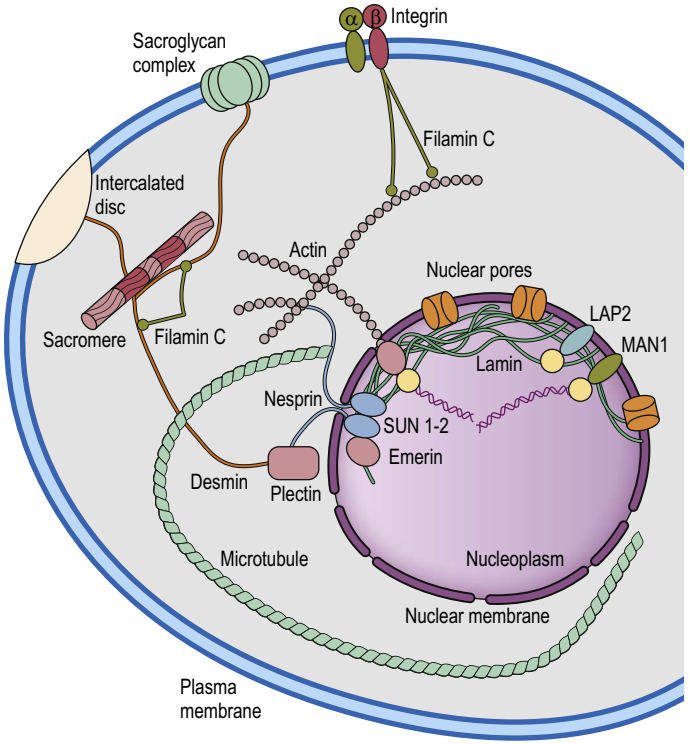


Fig. 4.20 Schematic representation of intermediate filament (IFs) proteins interacting in the cell . IFs are precisely termed: the form intermediates between major cellular elements, e.g. nesprin connects microtubules and actin with lamin inside the nucleus tethered to DNA-containing chromatin and is available to ensure proper intracellular molecular movement during such processes as cytokinesis/cell division. (From Goldman, 2018, originally modified from *Biology of the Cell*.)

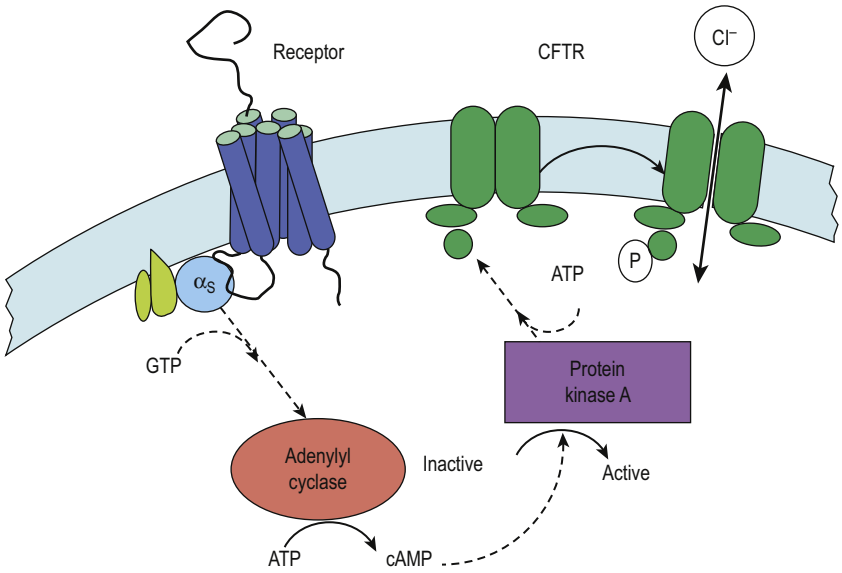


Fig. 4.21 An example of how cAMP regulates the G protein signalling response. The G_{as} protein signalling pathway is dependent on the cyclic AMP-dependent response in zebrafish oocytes and CFTR is activated by the binding of a ligand to a membrane receptor, causing the receptor to interact with a G protein (G_{as}). This ultimately leads to a cAMP signalling cascade, ending in the activation of CFTR. (From Cohen, 2008.)

Further information on Wnt signalling and cell function is available online at https://expertconsult.inkling.com

The extracellular matrix

Cells exist within a structural framework, the extracellular matrix (Box 4.4), which is secreted by cells such as the myofibroblast during wound healing and consists of several classes of macromolecules, the most abundant of which is collagen (accounting for 30% of total protein in the organism) (Fig. 4.22 and Video 4.1).

Genetic and protein sequence analysis of extracellular proteins has shown that, despite the great variety of matrix molecules, extensive sequence homology exists between them with recurring structural motifs (Box 4.4). For instance, the 'epidermal growth factor' motif appears to be present in several apparently different extracellular matrix proteins.

Extracellular matrix proteins determine the structural nature of the tissue

Tissues are defined by their extracellular matrix constituents. For instance, the cornea contains type I collagen filaments, which do not form fibrils of diameter greater than 5 nm; this is important in corneal transparency. The vitreous contains an isoform of type IX collagen, which is different in certain characteristics from cartilage type IX collagen.

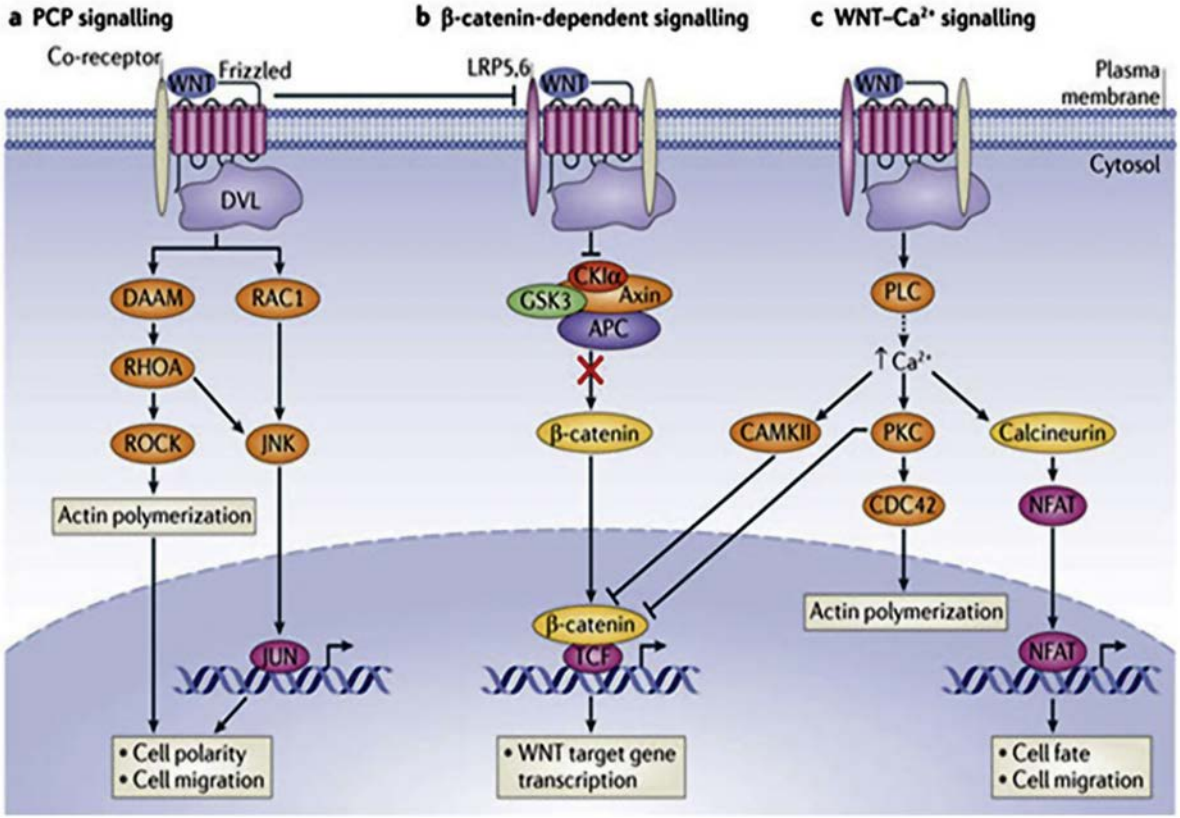
Collagen fibres are composed of a triple helix of three polypeptide chains, of which there are 46 different types and which assemble into 29 distinct collagen molecules (Table 4.2). Types I–IV are the most common collagens of which Type I accounts for 90% collagen in the body.

At least 22 of these have been detected in the developing and adult eye. The eye and cartilaginous tissues share six types of collagen not found very frequently in other tissues; indeed, the collagen triple helix (Fig. 4.23 and Box 4.4) has been found as a domain of many proteins and it has been suggested that all such proteins should be included in the collagen family. Collagen types are determined by the combination of the three types of chain forming the α helix core. For instance, type I collagen triple helix is made up of two unique α_1 chains and a unique α_2 chain, coded as $[\alpha_1(I)]_2\ \alpha_2(I)$; type II collagen is composed of three identical unique α_1 chains, $[\alpha_1(II)]_3$. Each collagen therefore has a set of unique chains that make up the triple helix.

In certain collagens, particularly the non-fibrillar collagens such as types IV and IX, the protein is composed of short segments of triple helix (COL1, -2, -3, etc.) interspersed with sections of non-collagenous (NC1, NC2, etc.) protein. These are also sometimes referred to as FACIT collagens (fibril-associated collagens with interrupted triple helices) (Table 4.2). These proteins usually act as bridges or networks for binding other proteins and forming complex protein aggregates such as basement membranes. Collagens in essence are the structural components that hold organs and tissues together and the mechanism by which they form is interesting: the three polypeptide chains form a monomer by self-assembly from a small nucleus in a

Wnt signalling and cell functions

Wnt signalling pathways are central to many developmental and cell biological functions (eFig. 4.9), including planar cell polarity (PCP) mediated by Wnt/Frizzled interactions; β-catenin-dependent signalling, which is extensively involved in cell–cell interaction and cell adhesion and includes interaction of Wnt with the low-density lipoprotein receptor (LRP)5,6; and Wnt-calcium (Wnt-Ca^{2+}) signalling through activation of calcineurin/NFAT signalling is involved in cell fate and cell migration. This latter pathway is activated in dendritic cells but not macrophages and induces apoptosis in DC (see Chapter 7, p. 427).



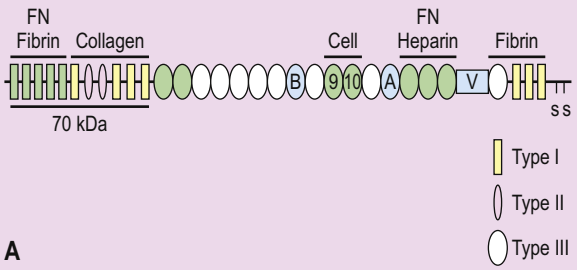
eFig. 4.9 Wnt signalling involves the co-receptor Frizzled to promote actin polymerization and cell polarity (planar cell polarity, PCP), the low-density lipoprotein receptor (LRP) 5,6 to mediate β-catenin signalling to the nucleus and Wnt-Ca^{2+} signalling, which activates the calcineurin-NFAT pathway determining cell fate. (From Niehrs, 2012.)

Box 4.4 **Extracellular matrix proteins**

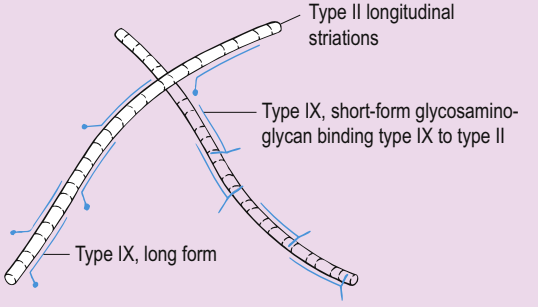
Extracellular matrix proteins appear to be made up of building blocks (structural motifs), which are protein domains with extensive homology to existing protein structures. For instance, three fibronectin domains are reproduced in many other proteins; the epidermal growth factor-like domain also appears in modular form in proteins such as plasminogen; and the arginine–glycine–aspartate (RGD) cell adhesion site is present in many molecules (**A**).

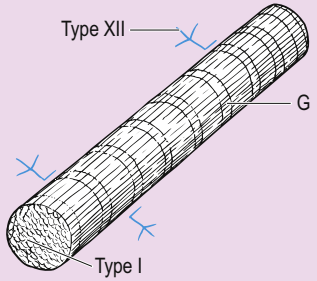
Fibronectin is a dimer containing discrete domains for the attachment of other molecules (**A**). The modular structure of fibronectin shows that it consists of 12 type I modules (rectangles), two type II modules (violet ovals) and 15–17 type III modules (ovals). The alternatively spliced domains IIIB, IIIA and the V region are shown in yellow. Binding domains for fibrin, collagen, cells and heparin are indicated; dimer forms via cysteine pair at the C-terminus (SS).

Laminin is a cross-shaped trimer in which the B1, B2 and A chains form a triple helix in the stem of the cross. Laminin also has several discrete domains for attachment of various molecules. In addition, it contains a cryptic cell adhesion-binding site, which becomes exposed after partial proteolysis (**B**). Fibrillar collagen is organized in tissues by association with smaller non-fibrillar collagens, e.g. type I with type XII collagen, and type II with type IX collagen (**C**).

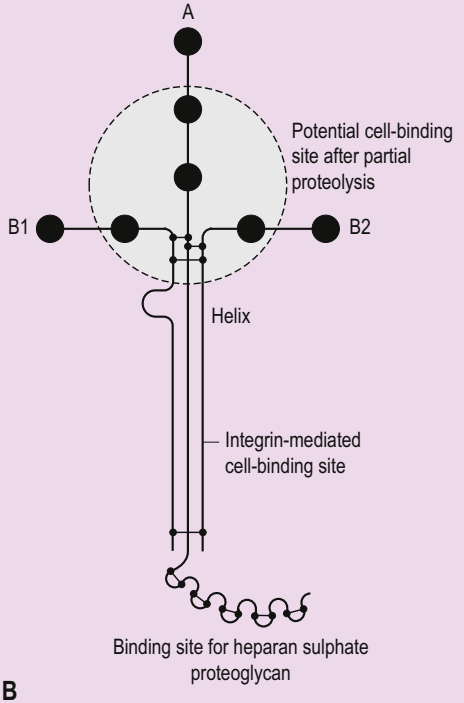


A





C



B

(Part A reproduced from Mao and Schwarzbauer, 2005, with permission from Elsevier; Parts B and C courtesy Yamada and Miyamoto, 1995.)

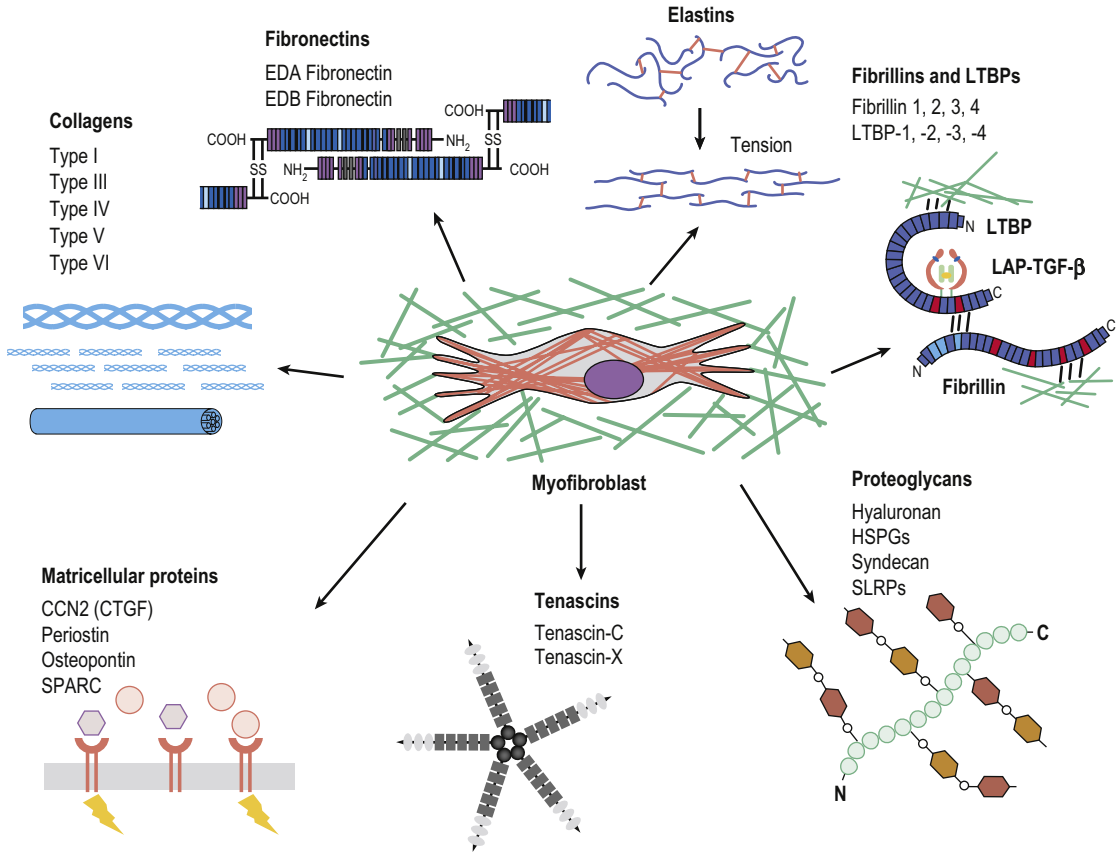


Fig. 4.22 The myofibroblast matrix. Schematic of some of the ECM molecules relevant to tissue fibrosis. The myofibroblast (*centre*, with *red stress fibres* containing α -smooth muscle actin) lies enmeshed in its ECM *(green)*. Components of the ECM are depicted (clockwise, from the 12 o'clock position): *elastins, fibrillins and LTBPs, proteoglycans, tenascins, matricellular proteins, collagens and fibronectins*. The myofibroblast encounters, signals, and modulates the expression of these various components as outlined in the text. (From Klingberg et al., 2013.)

Table 4.2 Some collagen types

Fibrillar	Anchoring	FACIT[a]	Transmembrane	Beaded filament	Network-forming	Multiplexin
I, II, III, V XI, XXIV XXVII	VII	IX, XII XIV, XVI XIX, XX XXI, XXII	XIII XVII XXIII XXV	VI, XXVI XXVIII	IV, VIII, X	XV, XXVI XXVIII

[a]*FACIT*, Fibril-associated collagen with interrupted triple helices.

zipper-like fashion, much in the way that crystallization occurs, and the formation of fibrils from monomers is also entropy driven. Indeed, the semi-crystalline packing of collagen fibres, achieved through lysine hydroxylysine cross-links, confers tensile strength to tissues.

A subcategory of collagenous proteins is included as components of transmembrane proteins, including types XIII, XVII, XXIII and XXV (Table 4.2), and remains following cleavage of the soluble fragment of the cytokine or adhesion molecule by enzymes known as sheddases. They

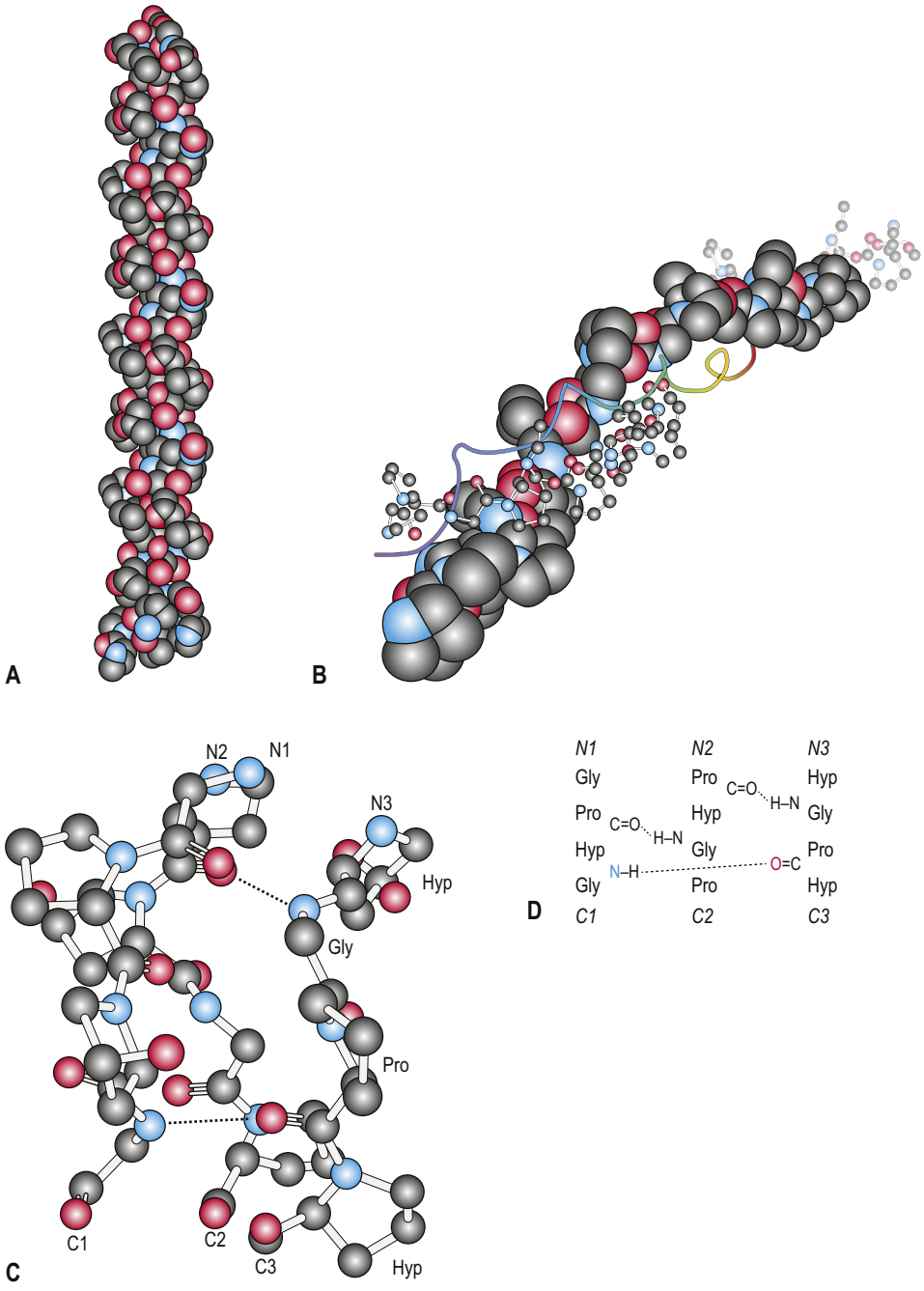


Fig. 4.23 Overview of the collagen triple helix. (**A**) Crystal structure of a collagen triple helix, formed from (ProHypGly)4– (ProHypAla)–(ProHypGly)5. (**B**) View down the axis of a (ProProGly)10 triple helix with the three strands depicted in space-filling, ball-and-stick, and ribbon representation. (**C**) Ball-and-stick image of a segment of collagen triple helix highlighting the ladder of interstrand hydrogen bonds. (**D**) Stagger of the three strands in the segment in panel **C**. (From Shoulders and Raines, 2009.)

exist as homotrimers of a collagen-specific α-chain. Some transmembrane collagen-like molecules occur as specific receptors such as the macrophage scavenger receptors. Sheddases are of several types and include the ADAMs (proteinases of the A Disintegrin and Metalloproteinase family) and ADAMTs (ADAMS with a thrombospondin, T motifs), which are involved for instance in the shedding of molecules such as soluble tumour necrosis factor-α-converting enzyme (TACE) from leucocytes (see Chapter 7, p. 411). This general process is important in establishing the soluble regulatory constituents of particular extracellular matrices.

The transmembrane collagen XVII (originally described as bullous pemphigoid antigen 180, BP180) is important in cellular adhesion because it binds α_6 integrin and laminin 5 extracellularly and β_4 integrin, plectin and BP230 intracellularly. Degradation products of collagens, such as the non-fibrillar collagens XVIII and XV, which yield the anti-angiogenic products endostatin and restin after cleavage, have major roles in regulating cell function through inhibition of matrix metalloproteinases.

The eye contains a wide variety of different collagens. For instance, the cornea contains types I, V and VI in the stroma, and types IV and VII are present in the subepithelial layer. In addition, type XVI transmembrane linker collagen has been found in the basal epithelial cell matrix. The iris contains collagen types I, III and IV, and the zonule contains type IV. The lens contains only type IV and the vitreous contains collagens II, IX and XI, complexed to the extracellular matrix protein fibrillin. This complex has an important role in vitreous matrix organization.

Certain types of collagen are unique to ocular tissue in that their structure has been modified. Examples include type VIII in Descemet's membrane, types II and IX in vitreous (both similar but not identical to cartilage collagens) and type III in the distensible tissue of the choroid. Several of the newer collagens contain domains similar to type IX, and it has been suggested that these may represent a subfamily of type IX collagens (Table 4.2). The more recently described types XV, XVIII and XIX are located in basement membranes and may have a role in the formation of blood vessels.

Collagen does not simply fill the space between cells but engages four types of specific receptors on cells: integrins, discoidin domain receptors (DDRs 1 and 2, a subfamily of receptor tyrosine kinases involved in signalling to the cell), glycoprotein VI (GPVI) receptor (on platelets) and leucocyte-associated immunoglobulin-like receptor-1 (LAIR-1). DDRs and integrins are exquisitely specific in the collagen domain to which they bind and their main function is to maintain tissue stability during intra-tissue cell migration, while GPVI and LAIR-1 are involved in haemostasis and immunoregulation after tissue injury.

Elastin is the second major insoluble protein of the extracellular matrix. While collagen occurs in all tissues, elastin is present only in deformable tissues such as blood vessel walls, lung parenchyma and the zonule of the lens. Unlike collagen, elastin does not contain any methionine residues and is therefore separable from collagen after digestion with cyanogen bromide. Elastin is formed from soluble tropoelastin monomers, secreted from fibroblasts and smooth muscle cells, which initially self-associate through hydrophobic domains in a process termed coacervation *in vitro,* which *in vivo* corresponds to coalescence into spherical globules of 1–6 μm size. This process of self-association aligns lysine residues to form cross-links, via allysine intermediates, desmosine and isodesmosine, a process facilitated by the elastin co-acervated droplets depositing onto fibrillin microfibrils together with fibulin (see later) (Fig. 4.24). The final fibre is 90% elastin and 10% fibrillin and cross-links are mediated via a set of lysyl oxidase enzymes, mutations in which have been associated with the pseudo-exfoliation syndrome, the commonest cause of secondary glaucoma and a risk factor for successful cataract surgery because of the role of elastin in zonule formation. Elastin has little order to its tertiary structure; instead, it forms a random coil, which tends to become more ordered during deformation.

As for collagen, elastin degradation products have a role in regulating cell behaviour in the matrix and under certain circumstances may be pathogenic, e.g. in promoting tumour invasion and angiogenesis. Even smaller molecules

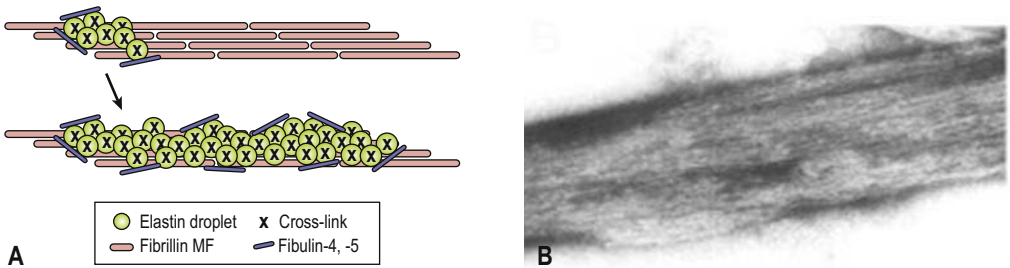


Fig. 4.24 Elastin 'droplets' are deposited on fibrillin microfibrils and cross-linked to the fibrillin molecule via lysyl oxidase in association with fibulin molecules. (From Muiznieks and Keeley, 2012.)

derived from a range of matrix proteins (matrikines) have regulatory activity for connective tissue cells; for instance, the tripeptide GHK, which has stimulatory activity for several cell types and promotes collagen biosynthesis and wound healing overall.

Fibrillins are a family of proteins serving as components of extracellular elastin microfibrils that are responsible for the biomechanical properties of tissues. Cysteine-rich glycoproteins, composed of multiple repeats of a Ca^{2+}-binding epidermal growth factor-like domain, are secreted in a proform and polymerize extracellularly. They are present in vitreous (see earlier section) and are an important component of the zonule. Fibrillins have a structural role in long-range elasticity of tissues and are a component of elastic fibres (see earlier section). They also have a role in the fine-tuning of growth factor signals such as those involving transforming growth factor-β (TGF-β) and particularly latent TGF-β-binding proteins (LTBP) involved in morphogenesis. Mutations in *FN1* cause a dominant form of Marfan syndrome.

Tenascins are extracellular matrix proteins best described as adhesion modulating, matricellular proteins which do not form major fibrillary structures such as collagen or elastin. They are module-built proteins comprising EGF-like repeats, fibronectin-type III repeats and fibrinogen domains (Fig. 4.25). Several forms are described (tenascin-C, -X, -W, etc.) and phenotypes in knockout mice are broadly normal although there may be some subtle defects in wound healing. Tenascin-Y appears to be restricted to neural tissue, while mutations in tenascin-X have been linked to Ehlers–Danlos syndrome, in conjunction with known defects in pro-collagen and elastin biosynthesis. Tenascins are widely distributed and appear to have an anti-adhesive role, particularly antagonizing the effects of fibronectin. Interestingly, fibronectin is found in the anterior lens capsule, while tenascin is found in the posterior capsule.

Laminins are an integral part of basement membranes and have a characteristic hetero-trimeric cross-shaped structure composed of an α chain, a β and a γ chain constructed to shape and an α coiled rod (see Box 4.4). There are 16 isoforms variously made up for one α (α1–α5), 3β (β1–β3) and 3γ (γ1–γ3) chains that are specific for different cell types (e.g. kalinin in epithelial cells) and are named according to their chain components. For instance, human laminin (laminin-332 composed of α3, β3 and γ2 chains) is specific for epithelial cells and binds to the $\alpha_3\beta_1$ and $\alpha_6\beta_4$ integrins.

Further information on the detailed structure of human laminin332 is available online at https://expertconsult.inkling.com

The infrastructure of basement membranes is composed of type IV collagen in a highly organized lattice network

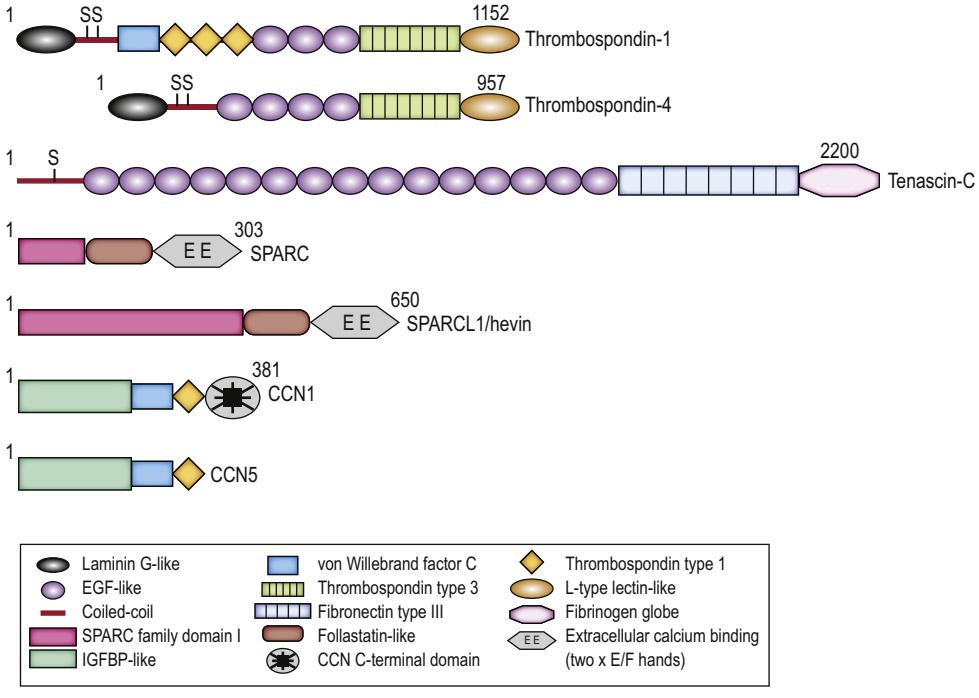
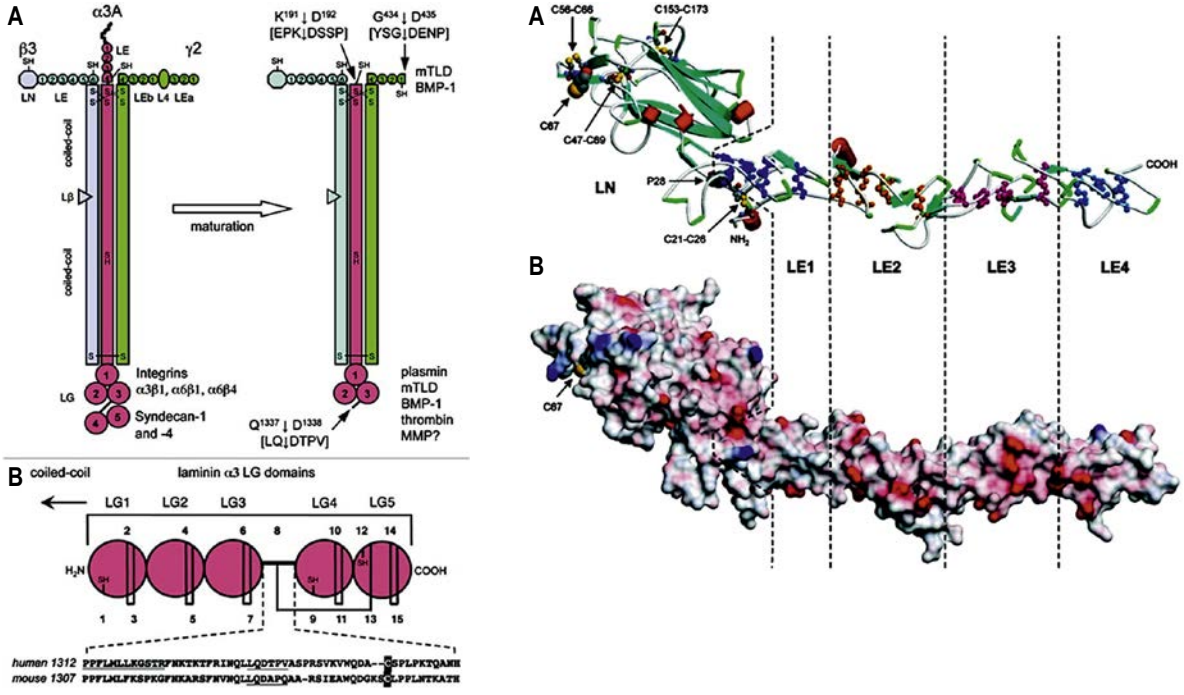


Fig. 4.25 Schematic diagrams of the domain organization of representative members of the protein families discussed in this chapter. (Not to scale.) (From Mosher and Adams, 2012.)

Human laminin-332

Laminin-332 (eFig. 4.10) is specific to the epithelial basement membrane and is central to epithelial sheets bound to mesenchyme, which have to withstand external forces such as in the skin and other mucosal layers. Matrix metalloproteinases are involved in post-translational processing of laminins as they incorporate themselves into the basement membrane to form anchoring adhesions and reciprocally, MMPs (MMP-3, -12, -19 and -20) are activated by laminin degradation products to process the γ2 chain and induce epithelial migration during wounding. Laminin in the basement membrane is tightly bound to nidogen (entactin), a protein that mediates the binding of laminin to heparin-containing proteoglycans (heparin sulphate proteoglycan) such as perlecan and syndecan in the basement membrane matrix.



eFig. 4.10 Structure of human laminin. Human laminin is known as laminin 332 on the basis of its three subunit chains (α3A, β3 and γ2). Each chain is composed of different domains that are indicated in (**A**). An odd number of cysteine groups allows SH bonds to form in the L domains after maturation of the molecule. The first three LG domains (LG1–3) interact with α3β1, α6β1 and α6β4 integrins while the last two (LG45) contain binding sites for syndecan-1 and -4. Newly synthesized laminin 332 undergoes maturation by proteolytic processing at the α3A chain N- and C-terminus as well as at the γ2 chain N-terminal extremity. A schematic structure of laminin domains is shown in (**B**), with numbered cysteine groups and the disulphide bridges.

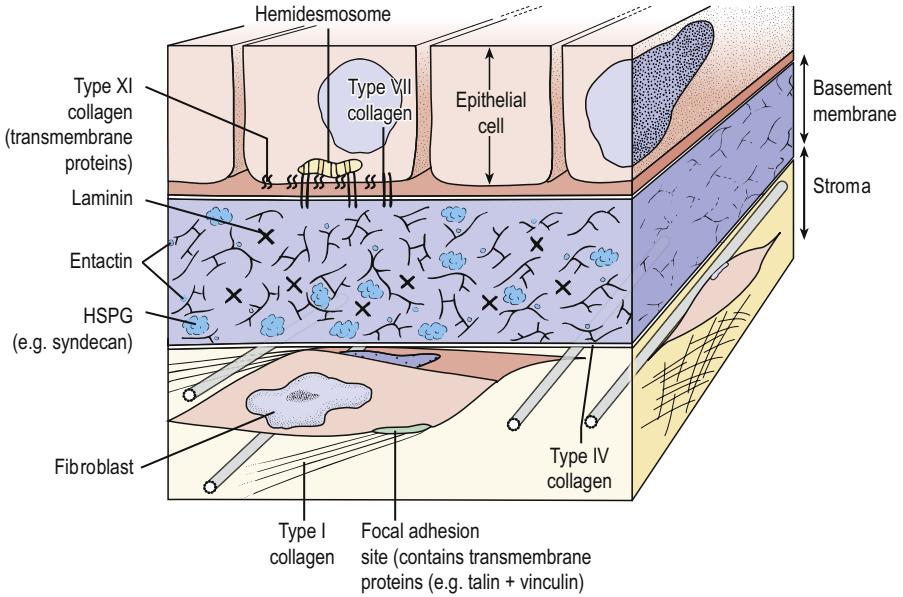


Fig. 4.26 Structure, composition and relationships of the basement membrane.

to which the complex of laminin–nidogen is also bound. The proteoglycan then acts as a space-filling molecule in the basement membrane (Fig. 4.26). Cells are bound to the basement membrane via anchoring fibrils containing type VII collagen and transmembrane proteins such as types XI and XVIII collagen.

Most laminins are produced by the cells adhering to the basement membrane in which they are found; interestingly, the laminins of the internal limiting membrane (ILM) of the retina are synthesized by ciliary body and lens cells. ILM laminin and fibronectin are substrates for ocriplasmin, a truncated form of the fibrinolytic enzyme now recommended for treatment of certain types of vitreo-retinal disease involving traction on the macular retina.

Non-structural proteins

Many other proteins are distributed throughout the extracellular matrix, which, although not having a direct structural role, have important functions in cell–matrix and cell–cell interactions. Many of these proteins are engaged in self-organization with other ECM proteins (see p. 185) to form fibrillar networks. Probably the best known of these is fibronectin, a 250 kDa heterodimer that has multiple cell and molecule binding domains.

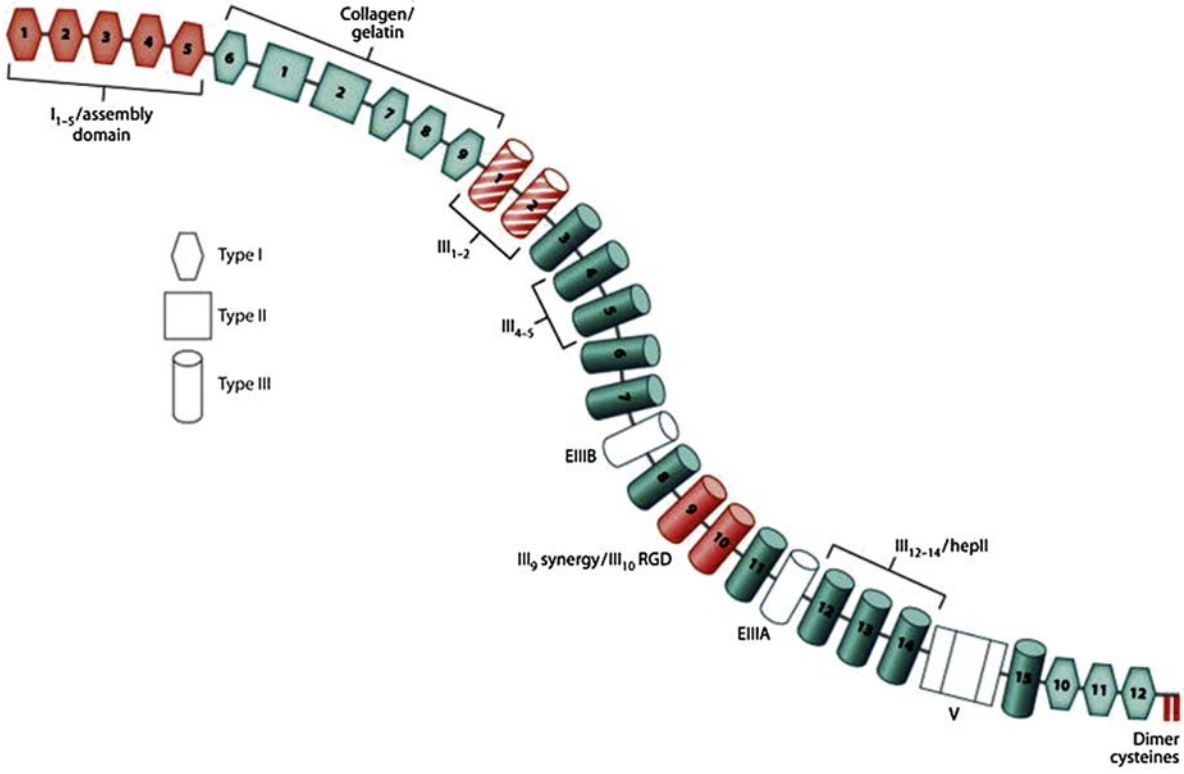
Further information on the great 'tissue glue' protein fibronectin is available online at https://expertconsult.inkling.com

For instance, one of the type III fibronectin domains contains the ubiquitous cell adhesion domain Arg-Gly-Asp-Ser (RGDS in single-letter amino acid nomenclature), found in many other proteins. Such domains bind to other proteins that traverse the cell membrane, anchoring the cell to the matrix, including the integrins. Multiple forms of fibronectin occur by alternate splicing. Two major forms exist: plasma fibronectin, secreted by hepatocytes, and cellular fibronectin, secreted by fibroblasts and forming the ECM fibrillar network. Fibronectin may be involved in reverse integrin-mediated cell signalling and, through the actin-binding protein profilin, may regulate stress fibre formation in endothelial cells and fibroblasts. Fibronectin also has high-affinity binding sites for fibrin(ogen), thus promoting incorporation of fibrin into the extracellular matrix during wound healing. Fibrin binds growth factors such as fibroblast growth factor 2 and vascular endothelial cell growth factor, providing sources for angiogenic stimulation.

Thrombospondins (Fig. 4.25) are a group of extracellular matrix proteins that were first identified as platelet release proteins but that are now known to be secreted by endothelial cells and other cell types. Together with SPARC (secreted protein, acidic and rich in cysteine), tenascins and CCN proteins they are considered to be adhesion-modulating proteins. Five forms of thrombospondin exist, two of which appear to be alternately spliced forms of the parent molecule. Like extracellular matrix proteins, thrombospondins are composed of building blocks, each with specific cellular and molecular adhesive properties. Thrombospondins are involved in the regulation of angiogenesis, for instance in preventing angiogenesis associated with ocular inflammatory

Fibronectin, the tissue 'glue'

Fibronectin initiates the process of cell and tissue integration by binding integrin receptors on the cell surface and entraining the mechanical forces of the cell through transmembrane connections with the dynamic actin cytoskeleton. Matrix assembly is critically dependent on interactions between matrix proteins and the cells which produce them. Such proteins self-organize (see p. 163) through specific domains on cell surface protein receptors such as $\alpha_5\beta_1$ integrin, which connect to the intracellular cytoskeleton proteins such as actin via linking proteins such as talin and vinculin. Cell contractility then fashions the shape, contour and architecture of the matrix. Once in place, fibronectin then contributes to the assembly of other matrix proteins through its many molecule-specific domains (eFig. 4.11).



eFig. 4.11 The fibronectin molecule: Each subunit comprises three types of repeat involved in Fn assembly (including the cell-binding RGD domain), the collagen-binding domains, and glycosaminoglycan-binding domains (heparin and syndecan). (From Singh et al., 2010.)

disease (see Chapter 7, page 460). In this regard, they form complexes with a ubiquitous plasma protein, histidine-rich glycoprotein (HRG), which has many regulatory functions in wound healing, cell migration and immune cell function. In addition, thrombospondins are better regarded as matricellular proteins that regulate a variety of processes, including cell adhesion and collagen fibrillogenesis. Thrombospondins also have considerable regulatory effects on cells by, for instance, binding to the leucocyte inhibitory receptor SIRP-1α, and to the ubiquitously inhibitory protein TGFβ. SPARC (Fig. 4.25) is a component of basement membranes but appears to be essential only for normal lens physiology since SPARC-deficient mice develop age-related cataract and lens rupture at 6 months of age but have no other phenotype. CCN (for CYR61/CTGF/NOV) proteins (Fig 4.25) are considered a separate family of proteins and include previously described proteins such as CTFG (connective tissue growth factor) and von Willebrand factor (which is involved in haemostasis) and are multifunctional regulatory proteins.

Several other proteins are present in the extracellular matrix, such as proteases (e.g. plasminogen) and their inhibitors (plasminogen activator inhibitor 1 (PAI-1), α_2-macroglobulin, etc.). PAI-1 is present at high concentrations around various cells in the quiescent state and is considered important in the regulation of cell migration by controlling the level of cell-associated plasminogen activator required to initiate the degradation of basement membrane proteins. Other important proteins include the matrix metalloproteinases (MMPs) and their inhibitors (TIMP-1 to -3). The cells themselves secrete some of these proteins, while others are synthesized predominantly in the liver and reach the extracellular matrix of peripheral tissues via the blood.

Some proteins have their action at a distance from the cell, such as fibrillin, a 350 kDa microfibrillar protein that is involved in the assembly of elastin fibrils (Fig. 4.24). Fibrillin is found in the tertiary vitreous and the zonule (see previous section). Mutations in fibrillin are found in Marfan syndrome, a disease of elastic tissue in which dislocation of the lens is a central feature.

Glycosaminoglycans occur in the extracellular matrix bound to core proteins as proteoglycans

Glycosaminoglycans (GAGs) occur in a variety of forms, essentially based on a repeating disaccharide structure (Box 4.5). The prototype proteoglycan was described in relation to cartilage, in which a series of proteoglycans are linked to a hyaluronic acid backbone (Fig. 4.27). This, however, is not relevant to ocular proteoglycans; for instance, hyaluronic acid in the vitreous is not associated with other GAG-containing proteoglycans but is linked to collagen type IX, which in this situation is regarded as a proteoglycan. In addition, keratan sulphate proteoglycan occurs in the cornea in the absence of hyaluronan, where it appears to regulate collagen fibril diameter (see later). In other tissues, hyaluronan has both structural and cell-regulatory activities. For instance, two hyaluronan receptors of the LINK protein family, LYVE-1 and stabilin-1, are important in specific functions of lymphoid cells and endothelial cells, respectively. The main cell surface receptor for hyaluronan is CD44 and after internalization hyaluronan may even have intracellular functions. However, endogenous intracellular hyaluronan is also involved in chromatin rearrangement in the nucleus, especially during late prophase/early prometaphase of mitosis (see Chapter 3, p. 136).

The number of proteoglycans that have been described in relation to other cellular functions has increased greatly (Table 4.3). Proteoglycans (PGs) are thus now described in three families of proteins: transmembrane, hyaluronan-binding and collagen fibril-regulating proteoglycans. Transmembrane PGs have a direct role in specific intercellular interactions, hyaluronan-binding PGs function as links for space-filling effects (e.g. in the vitreous cavity), while collagen fibril-regulating PGs have specific roles, e.g. force-generating in tendons, light transmission in the cornea, mechanical support in bone and deformability in heart muscle. In the cornea, four PGs (decorin, lumican, keratocan and mimecan) are the major keratan sulphate moieties that, together with collagens type VI and XII, are essential for maintaining corneal transparency.

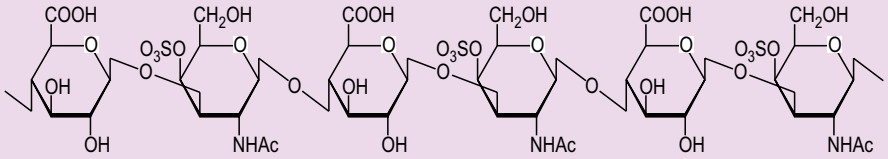
Extracellular matrix molecules are intimately involved in cell adhesion

Some proteoglycans such as syndecans participate constitutively in cellular adhesion and its associated cell signalling, acting as co-receptors with other receptors such as integrins in the receptor cluster, and thus connect through to the cell cytoskeleton (Fig. 4.28). Other proteoglycans interact variably with glycosaminoglycans such as CD44 and neuropilin, both of which are involved in leucocyte–matrix interactions and may be dependent on the state of leucocyte activation.

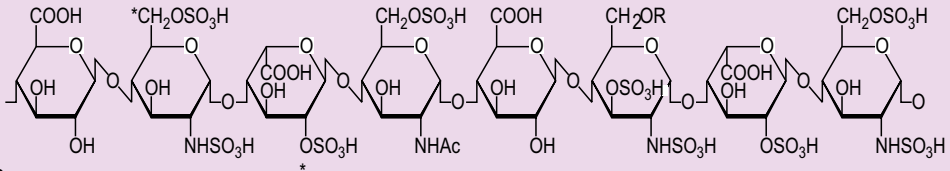
Cell adhesion is fundamental to many biological processes, such as morphogenesis, development, immune reactions to foreign proteins and many other processes. Cells adhere both to the matrix and to other cells and usually do so via specific receptors in the cell membrane, i.e. integrins. Integrins are heterodimeric proteins (they have α and β chains, some of which are common to more than one integrin type) and are described in terms of their chain composition, e.g. $\alpha_3\beta_5$ and $\alpha_1\beta_6$ integrins (Table 4.4; see

Box 4.5 **Structure of glycosaminoglycans**

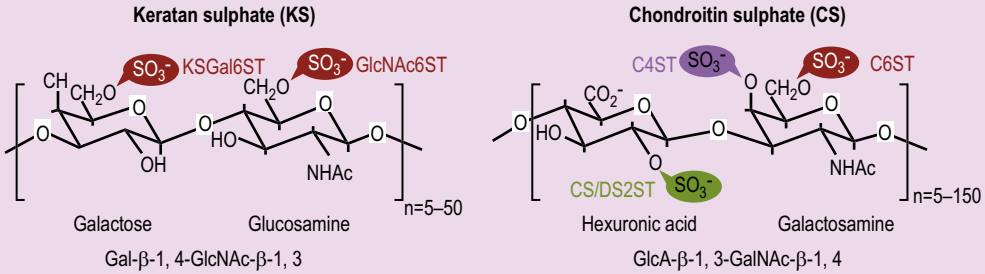
Glycosaminoglycans are long chains of repeating disaccharides based on a common structure: (**A**) heparan sulphate; (**B**) chondroitin 4,6-sulphate. Chondroitin 6- and 4,6-sulphates have additional sulphate groups at the appropriate C atoms (*); dermatan sulphate has the GlcA (glycosamino-) residue replaced in variable lengths of the chain by IdoA (iduronic acid). Heparan has many more sulphate groups and a higher content of IdoA than heparin. Keratan sulphate has galactose instead of glucose, and hyaluronic acid has no sulphate groups.

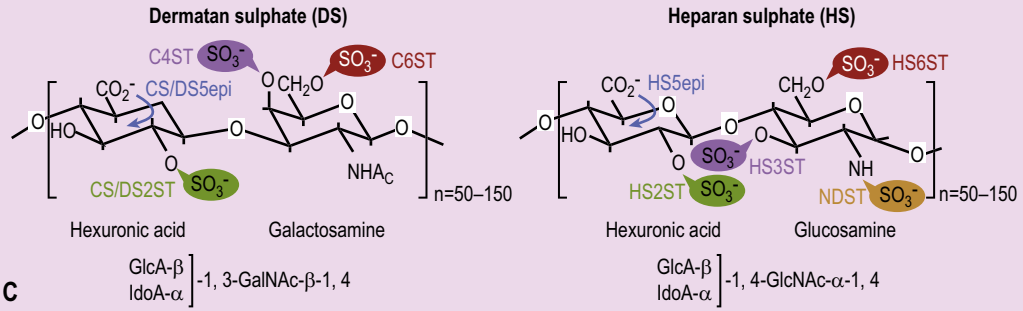


A



B





C

also Chapter 7, Table 7.3, and the role of integrin receptors in leucocyte adhesion). Each cell type adheres preferentially to particular extracellular matrix proteins, depending on the type of integrin receptor it happens to express (see Table 4.4). In addition, cells may express different integrins at different times, depending on their state of activation, as, for instance, with the (myo)fibroblast during wound healing (see later). Differential adhesion by cells is shown clearly in the lens where anterior epithelial cells use the $\alpha_5\beta_1$ integrin to bind fibronectin, while the equatorial and posterior fibre cells express the $\alpha_6\beta_1$ integrin.

Binding of the cell to the matrix via integrin receptors not only has a structural role but also is involved in transmembrane signalling, which may be two-way, i.e. the matrix may modify the behaviour of the cell and the cell may transmit information to other cells via the matrix (inside-out signalling). Integrins particularly regulate adhesion of motile cells to the matrix, but the overall process involved in this form of 'mechanotransduction' (physical force-generated signalling) has many similarities to the cadherin-mediated cell adhesion in sheets of cells such as epithelia via the adhesome.

Further information on insideout signalling and cell adhesion is available online at https://expertconsult.inkling.com

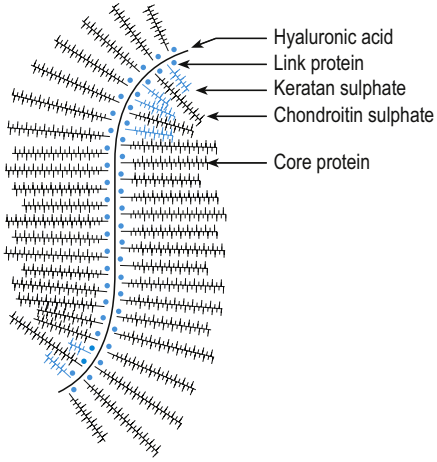


Fig. 4.27 Artist's impression of hyaluronic acid molecule. (From Rosenberg, 1975.)

Epithelial and endothelial cells bind via transmembrane complexes to each other and to the basement membrane

Epithelial and endothelial cells rest on a highly organized basement membrane (see previously). Binding of epithelial-type cells occurs predominantly via adherens-type junctions and desmosomes/hemidesmosomes, in which several distinct proteins have been identified (Table 4.5). Some of these proteins are members of what is known as the cadherin family of proteins (cell adhesion proteins), a group of transmembrane proteins that regulate intercellular adhesion.

Further information on cadherins, filaments and cell junctions is available online at https://expertconsult.inkling.com

These proteins are not only important in the mechanical support of intercellular interactions but also play a role in inter- and intracellular signalling. Gap junctions in particular facilitate and indeed may permit amplification of signalling events that are important in coordinated cell behaviour, such as in an epithelial cell sheet. The mechanistic similarities between integrins and cadherins have been highlighted previously.

Mesenchymal cells bind to the matrix via focal adhesion sites

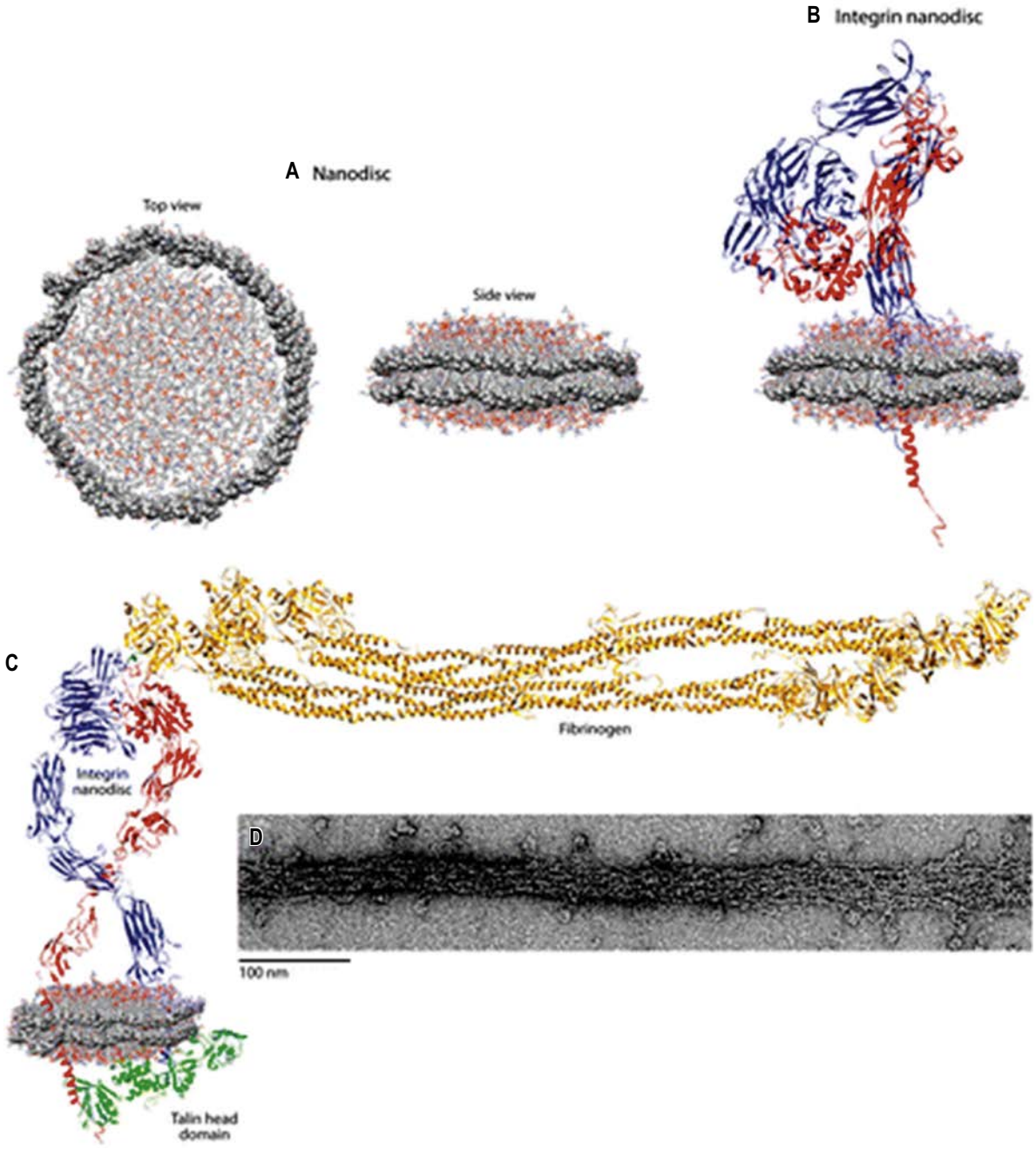
Mesenchymal cells, such as fibroblasts, keratocytes, chondrocytes and stromal cells, generally bind to the extracellular matrix via specific focal adhesion sites that contain

Table 4.3 **Some tissue proteoglycans**

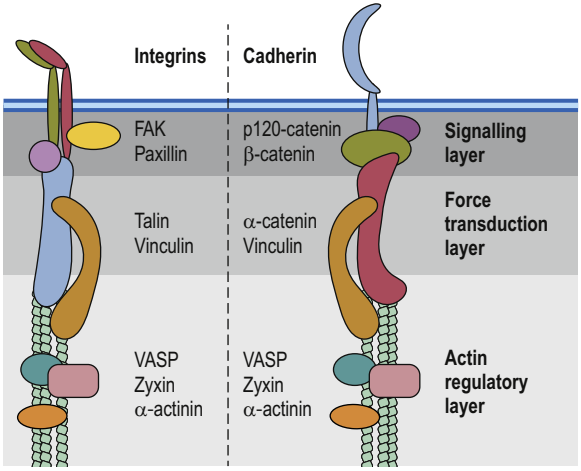
Proteoglycan	GAG	Tissue	Function
Aggrecan	CS, KS	Cartilage	Structural support
Versican	CS	Fibroblasts	Cell migration, support
Decorin	CS/DS	Fibroblasts	Fibrillogenesis
Fibromodulin	KS	? Keratocytes	Fibrillogenesis
α_2 (IX) collagen	CS	Vitreous, cartilage	Collagen binding
Syndecan	CS, HS	Epithelia, fibroblasts	Morphogenesis
Basement membrane	HS	Basement membrane	Support
CD44	CS	Lymphocytes, epithelia, retina	Cell–cell interactions

Inside-out signalling and the adhesome

This is an exquisitely sensitive system and is tunable, i.e. the integrins can vary their affinity for adhesion to matrix proteins, as directed by cellular inside-out signalling processes. The molecular subtlety is so exquisite that integrins embedded in discoid membrane structures (nanodiscs) are bent and become extended when they bind ligand. Integrins are kept in a loosely bound state by 'clasps' on either side to the membrane. The orientation (tilt) of the integrin molecule precisely determines the interaction with ligand such as fibrinogen/fibrin. Intracellular talin mediates inside-out signalling, which is regulated by kindlins under the control of the small signalling molecule Rap1 (eFig. 4.12). The sites of cell adhesion through integrins involve many proteins and similar to the site of cell adhesion is now discussed as an adhesome. The general mechanistic similarity between the integrin adhesome and the cadherin adhesome is remarkable (eFig. 4.13).



eFig. 4.12 (**A**) Nanodiscs with lipid molecules are shown *(coloured)* and membrane scaffolding proteins *(grey)* are shown in solid surface view. The nanodiscs range from 10 to 13 nm in diameter. (**B**) Reconstituted integrin nanodisc containing an integrin molecule (ribbon diagram, α subunit in blue and β subunit in red). (**C**) The head of a talin domain *(green)*, integrin nanodisc *(red* and *blue)*, and an integrin ligand (in this case, fibrinogen, *yellow*). (**D**) A negatively stained electron microscopy image of the reconstituted inside-out integrin activation system using fibrin as a ligand. (From Kim et al., 2011.)



eFig. 4.13 Resemblance between cadherin junctions and integrin adhesion, also suggesting possible similar outside-in signalling mechanisms comprising a signalling layer, a force transduction layer (FTL) and an actin regulatory layer. The open conformations of talin *(purple)* and β-catenin *(light blue)* mediate binding of vinculin *(light brown)* in the FTL, leading to strengthening of the binding. (From Han et al., 2016.)

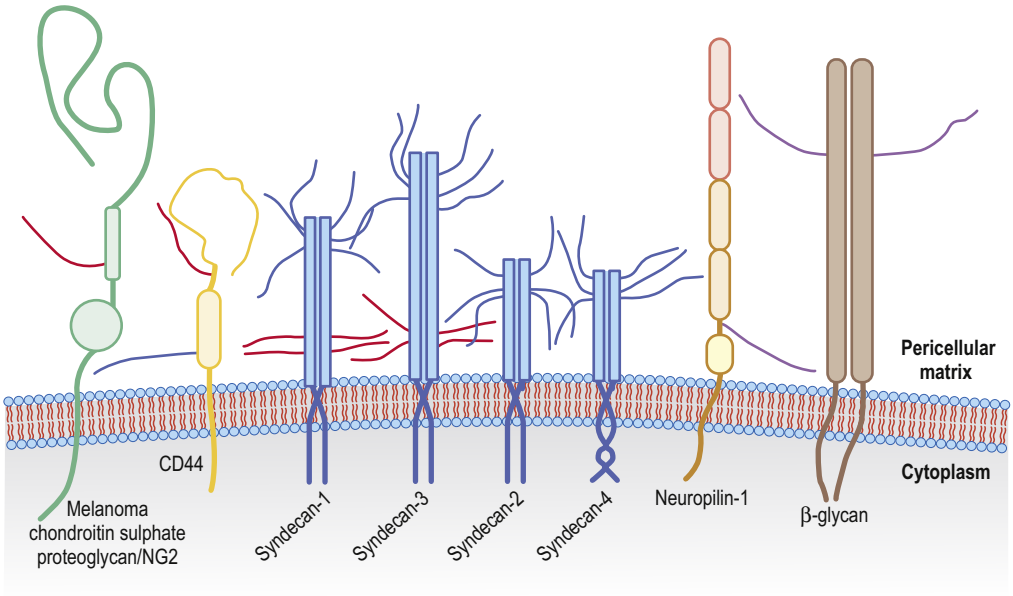


Fig. 4.28 The transmembrane proteoglycans. The four syndecans are usually substituted with heparan sulphate chains *(dark blue)*, while syndecans-1 and -3 have additional chondroitin sulphate chains *(pink)*. The melanoma chondroitin sulphate proteoglycan/NG2 *(green)* has one chondroitin sulphate chain. The other proteoglycans also have other functions and are considered 'part-time' proteoglycans. (From Couchman, 2010.)

Table 4.4 Integrin binding to extracellular matrix proteins

- Integrins $\alpha_1\beta_1$, $\alpha_2\beta_1$, $\alpha_6\beta_1$, $\alpha_7\beta_1$, $\alpha_v\beta_4$, $\alpha_6\beta_4$ bind to laminin
- Integrins $\alpha_4\beta_1$, $\alpha_5\beta_1$, $\alpha_v\beta_1$, $\alpha_v\beta_3$, $\alpha_6\beta_4$ bind to fibronectin
- Only integrin $\alpha_3\beta_1$ binds to both laminin and fibronectin

Table 4.5 Proteins associated with cell junctions

Adherens junction	Desmosome
α, β, γ Catenin	Plakoglobin
Plakoglobin	Desmoplakin I and II
Vinculin	Desmoplakin IV
α Actinin	Desmocollin
Tenuin	Lamin B-like protein
Plectin	Plectin
Radixin	Desmoglein

transmembrane cytoskeletal proteins. Actin stress fibres bind via α-actinin and talin to the cytoplasmic side of the integrin receptor and to vinculin, while the extracellular component of the integrin receptor binds to extracellular matrix proteins such as collagen and fibronectin (see Fig. 4.26).

Further information on insideout signalling and cell adhesion is available online at https://expertconsult.inkling.com

Direct reverse signalling takes place through these sites using the signalling protein, focal adhesion kinase, one of the protein tyrosine kinases. Also involved is a further intracellular protein termed VASP (vasodilator-stimulated phosphoprotein), which integrates profilin–actin binding with talin, vinculin, F-actin and the integrin signalling complex. Together with known ion channel-dependent mechanosensors, focal adhesion proteins such as talin act as mechanosensors responding by induced unfolding of their tertiary structures. Similarly, cofilin acts to sever actin filaments in a negative-dependent fashion based on tension in the actin filament at focal adhesion sites. Certain other mesenchymal cells, such as muscle, have specific proteins in their focal adhesion sites such as dystrophin and paxillin, not only at the site of attachment of the extracellular matrix but also at the neuromuscular junction. A mutation in the dystrophin gene has been identified in patients with muscular dystrophy. An integrin-linked kinase appears

to be central to regulation of outside-in and inside-out signalling in adherent cells.

Biochemical pathways that affect ocular function

The eye, particularly the outer retina, is highly metabolically active and requires large amounts of ATP, the universal energy storage molecule, for this purpose. Cells cash in on this energy bank by hydrolysing ATP and coupling this event to cell-specific enzyme reactions that are otherwise energetically unfavourable. ATP is generated by oxidative metabolism, particularly of glucose but also of other molecules such as fats and proteins.

Hydrolysis of ATP is not without risk to the cell: free electrons are produced (H^+), which are normally mopped up by nicotinamide adenine dinucleotide (NAD). Indeed, NADH is the main electron carrier in the oxidation of glucose and other molecules, while NADPH (see later) is used to reduce certain molecules such as free fatty acids to permit them to enter the metabolic pathways.

Oxidative consumption of glucose requires coenzyme A, which is a carrier of acyl groups. Other coenzymes are also required for active metabolism, many of which are derived from vitamins.

Glucose metabolism and tissue glycation

The conversion of glucose to pyruvate is termed aerobic glycolysis. In the absence of oxygen (anaerobic glycolysis), this process continues with the formation of lactate and is accompanied by a net gain of two molecules of ATP. In the presence of oxygen, however, pyruvate enters the citric acid cycle with the production of 24 molecules of ATP. NAD^+ is regenerated in the mitochondrial electron transport chain. Glucose and other fuel molecules entering the citric acid cycle may thus be consumed to provide energy for synthesis of the building materials for other molecules such as amino acids.

Not all the fuel molecules are utilized immediately, but instead they are diverted to produce molecules that act as sources of power rather than energy. This is achieved via the pentose phosphate pathway, in which NADPH and ribose are produced, the former for use in reductive processes and the latter in the biosynthesis of nucleotides for RNA and DNA. These different pathways are briefly summarized in Fig. 4.29.

The pentose phosphate pathway also permits glucose formation from unrelated precursors (gluconeogenesis) but does not occur in the brain owing to lack of glucose-6-phosphatase; this probably also applies to the retina.

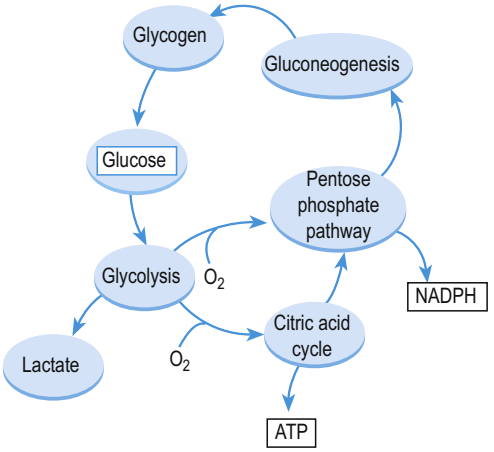


Fig. 4.29 Summary of glucose pathways.

However, the pentose phosphate pathway is active in the cornea and lens (see later).

Finally, glucose may be stored in the liver and muscle as glycogen, whence it can be retrieved for energy purposes by glycogenolysis. The Müller cells in the retina also contain stores of glycogen, which may be essential to maintain retinal function (see Chapter 1, p. 49, Fig. 1.36 and Chapter 5, p. 305).

Energy metabolism is controlled globally by nutrient intake, use and storage and is regulated by a range of signalling pathways targeting food excess such as insulin, insulin-like growth factor-1 and mammalian target of rapamycin (mTOR); food restriction is likewise controlled by a set of pathways including the ATP-consuming anabolic pathways and ATP-generating catabolic pathways, which are under strict regulatory control by a 'master-switch' enzyme system, adenosine monophosphate-activated protein kinase (AMPK), which controls the overall whole body energy metabolism and is involved in loss of neuroprotection in the brain, and possibly the retina, in ischaemic states. In addition, sirtuins, a family of seven histone-deacetylase proteins in mammals, serve as sensors of energy (ATP) levels and also have wider transcriptional and other regulatory roles such as protection from toxicity of reactive oxygen species via upregulation of AMPK.

Excessive glycolysis occurs under certain conditions such as in rapidly dividing tumours and at site of inflammation in myeloid cells. This is known as the Warburg effect and has significant effects on the handling of glucose by the organism, being involved in weight loss and cachexia.

Glucose enters cells by facilitated diffusion

Glucose transporters (GLUTs 1–14) are members of the SLC2 (solute carrier 2) gene family and are membrane proteins, which permit the facilitated diffusion

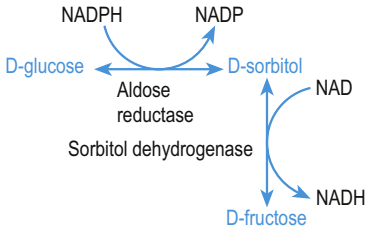


Fig. 4.30 Aldose reductase pathway.

of hexoses and polyols, including myoinositol, glucosamine and ascorbate, into and out of cells down their concentration gradient. Different cells and tissues have different GLUTs, which is relevant to how each tissue handles glucose.

Further information on glucose transporter mechanism of action and tissue distribution is available online at https://expertconsult.inkling.com

Thus, skeletal muscle utilizes GLUT4, which is dependent on insulin for its function in the plasma membrane; in contrast, brain and retina, which are considered to be insulin-independent tissues, express GLUTs 1 and 3. The liver and adipose tissue also have GLUTs 2 and 7, with several of the other transporters. GLUTs 1, 3 and 4, particularly GLUT1, can be upregulated in response to hypoxia via the gene hypoxia-inducible factor-1 alpha (HIF-1α), which also induces other genes such as vascular endothelial cell growth factor, which is important in retinal vessel permeability and new vessel growth. In the retina, this aerobic glycolysis effect is regulated through the Wnt/βcatenin pathway, which induces HIF-1a and promotes GLUT1 in photoreceptors and increased glycolysis through induction of enzymes such as hexokinase 2 (Valle et al., 2017), creating increased risk of AMD. GLUT4 in muscle is critical for whole body glucose homeostasis; GLUT7 is involved in gluconeogenesis and glycogenolysis in the liver, where it is located in the endoplasmic reticulum in association with glucose-6-phosphatase, the essential enzyme for these reactions.

Excessive levels of glucose may impair cellular metabolism

Glucose metabolism is tightly regulated by insulin, IGF-1 and glucagon release from the pancreatic β cells. In diabetes mellitus, in which there is impaired secretion or utilization of insulin, hyperglycaemia leads to excessive uptake of glucose into cells, despite negative feedback on the expression of GLUTs on the cell surface. This has the effect of overloading the metabolic pathways with activation of alternative routes for glucose handling such as the aldose reductase pathway (Fig. 4.30).

In certain tissues, such as the lens and retina, increased sorbitol may cause damage by osmotic dysregulation because sorbitol cannot be transported out of the cell easily. In this case, even high glucose can have direct osmotic damaging effects on the cell owing to raised intracellular $[Ca^{2+}]_i$ perhaps as a result of cell shrinkage with consequent activation of the stretch receptors. Alternatively, the excessive use of NADPH in this pathway might lower the levels of myoinositol, which is required for intracellular signalling, or cause increased generation of reactive oxygen species directly via activation of the phosphoinositol pathway. New data has also linked the sorbitol pathway more directly to inflammation via NAD/NADH and NFkB (see Chapter 7, p. 401) and disruption of the de-acetylase, sirtuin1. In this, via NFkB remains acetylated and thus activated for inducing cytokine production (Fig. 4.30).

Thus, activation of the aldose reductase pathway is not the direct cause of cellular damage but is merely a coincidental perturbation in the cell, which fails to deactivate NfFkB; excessive production of free radicals follows and may be important during oxidation of the high concentrations of glucose. Interestingly, aldose reductase has been implicated in inflammatory processes by promoting lipid peroxidation and free radical generation (see later). Whatever the mechanism, aldose reductase is likely to be involved in the pathophysiology of the lens at least, because inhibitors of this enzyme prevent the development of cataracts.

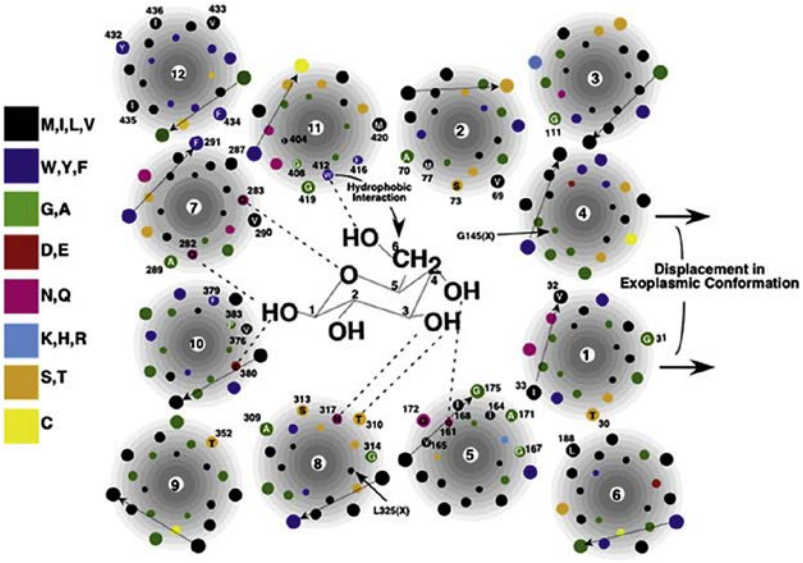
Chronic hyperglycaemia has also been shown to alter cellular metabolism via modulating insulin post-receptor intracellular signalling, involving the cascade insulin receptor substrate 1 (IRS-1)/phosphatidylinositol 3-kinase (PI3K)/Akt, as the result of production of high levels of hexosamine in the cell. This mechanism promotes insulin resistance in diabetes mellitus. A more general mechanism for the effects of high ambient glucose concentration on cell behaviour has been the induction of the cytokine TGF-β with its widespread effects on cell function, particularly increased proteoglycan synthesis and consequent basement membrane thickening. Patients with type II diabetes are known to have increased serum levels of TGF-β and thickened basement membranes.

Glucose metabolism may be modified depending on the cell. Recently, macrophages have been shown to produce itaconic acid (previously identified in the fungus Aspergillus) via the citric acid cycle through activation of the CAD/IRG1 gene (Fig. 4.31). Itaconic acid has been known previously through industrial processing and was not previously thought to be synthesized in mammalian cells. It has antibacterial properties but some bacteria have evolved mechanisms to reconvert itaconic acid to pyruvate.

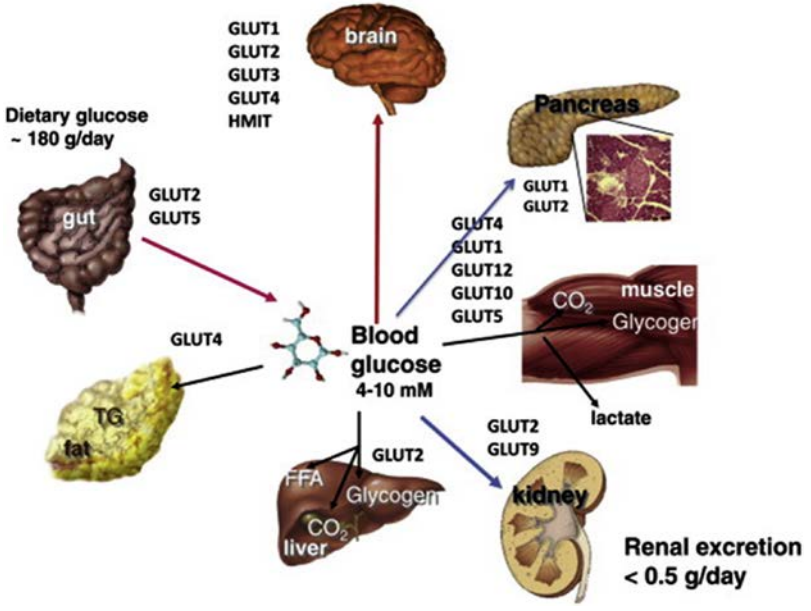
Glucose transporters

Glucose transporters (GLUTs, 14 have been described) are one of three molecular pathways of glucose transport into cells, the others being the sodium-driven glucose symporters (SGLTs) and the SWEETs, which have mostly been described in plants but also occur in mammals. GLUTs have a closely defined mechanism for shifting glucose molecules across the cell membrane (eFig. 4.14) and are specific to tissues and cells (eFig. 4.15). Glucose is transported across the membrane in aqueous phase propelled by hydrophobic interactions between glucose molecules and residues on the amino acids in the transporter protein, which maintain phase separation of the molecules, while other amino acids are dedicated to promoting binding and thus carriage of the molecule across the glucose gradients that exist. About one-third of the glucose is transported through the gut via GLUT2 and GLUT5 (for fructose) and is carried via red blood cells because of their very high concentration of GLUT1. In the resting state, most of the glucose is used by the brain and nervous tissues and during exercise muscle uses most of the circulating glucose. Adipose tissue regulates glucose utilization via production of hormones termed adipokines. Adipokines include molecules such as leptin, resistin, visfatin and adiponectin, which influence insulin signalling pathways, modify GLUT4 translocation and regulate glucose uptake. Other adipokines such as vaspin, chemerin and pref-1 are linked to obesity and insulin-resistant diabetes in rodents. It is not clear whether similar effects occur in humans. Receptors for chemerin are present on dendritic cells and macrophages and contribute to inflammation in obesity. An essential role for AMP-activated kinase (AMPK) in regulating glucose homeostasis of the organism via adipokines such as leptin is now known.

Six SGLTs have been described: SGLT1 in the intestine and SGLT2 in the kidney cortex; SGLT3 is a glucose sensor rather than a transporter; SGLT4 transports mannose and fructose; SGLT5 preferentially transports mannose and SGLT6 inositol. SGLT is active in the lens epithelium. The function of Low molecular weight sugar transporters (SWEETs) in mammalian systems is currently under investigation.



eFig. 4.14 Model of the exoplasmic substrate-binding site of GLUT1. The glucose molecule in the centre of the diagram is not drawn to scale. The helices surrounding amino acids in the glucose transporter are shown in a simplistic fashion for clarity. Amino acid residues that are in contact with solvent in the aqueous cavity are numbered and identified by the single-letter code. *Dotted lines* represent putative hydrogen bonds. (From Mueckler and Thorens, 2013.)



eFig. 4.15 Role of GLUT proteins in the maintenance of glucose homeostasis. The various GLUT proteins expressed by different tissues are shown. (From Mueckler and Thorens, 2013.)

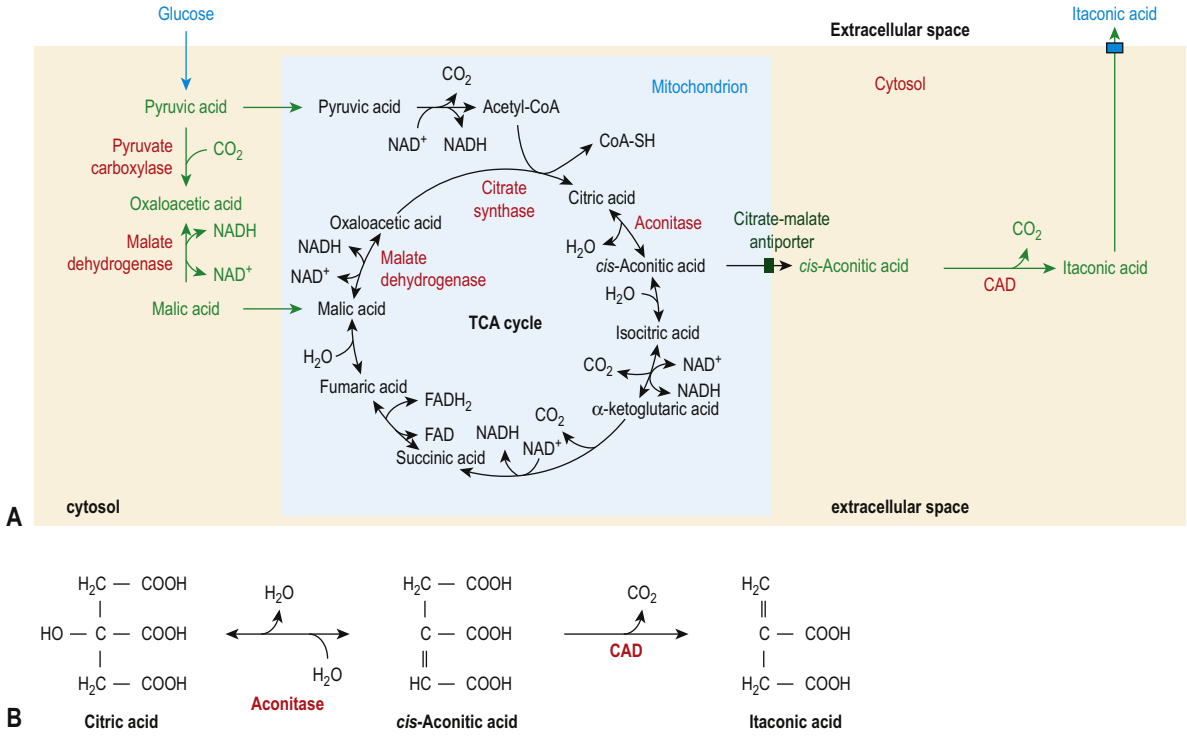


Fig. 4.31 Biosynthesis of itaconic acid production and its subcellular localization in *Aspergillus terreus*. Glucose is converted to pyruvic acid in the cytosol, which enters the tricarboxylic acid (TCA) cycle in the mitochondria either via acetyl-CoA or malic acid. The intermediate cis-aconitic acid is transported from the mitochondria into the cytosol and serves as a precursor for itaconic acid production catalysed by cis-aconitic acid decarboxylase (CAD). Extracellular metabolites are marked in *blue*; cytosolic metabolites, *green*; mitochondrial metabolites, *black*; and involved enzymes, *red*. Abbreviations: *FAD*, Flavin adenine dinucleotide; *$FADH_2$*, reduced form of flavin adenine dinucleotide; *NADH*, reduced form of nicotinamide adenine dinucleotide; *NAD^+*, oxidized form of nicotinamide adenine dinucleotide. (From Cordes et al., 2015.)

Excessive levels of glucose lead to glycation of proteins

Addition of sugar moieties to proteins can occur by enzymatic conjugation (termed glycosylation) or non-enzymatic conjugation (termed glycation). Glycation of proteins occurs in the presence of high concentrations of glucose in two phases: an early reversible phase during which the protein forms a Schiff base and Amadori products, followed by a later irreversible phase in which advanced glycation end products (AGEs) appear (Fig. 4.32). However, AGEs can be produced in the early phase because free glucose degrades to α-oxoaldehydes, such as glyoxal, methyl glyoxal and 3-deoxyglucosone, which are produced as glycolytic intermediates and also during lipid peroxidation (see later) and are potent glycating agents. (Fig. 4.32).

Glycation of proteins occurs at lysine, arginine and cysteine amino acids, thus affecting proteins such as collagen (pentosidine cross-links (Fig. 4.32) and haemoglobin and occurs as part of ageing and in diabetes mellitus; in addition, important extracellular matrix proteins, such as PAI-1 (see previously) may be glycated, leading to defective control of cell behaviour, including cell migration and activation. Glycation of cell membrane proteins such as Ca^{2+} channels in pericytes may impair their function, thus rendering them less responsive to endothelin-1-induced contractility. The two processes of non-enzymatic glycation and auto-oxidation of proteins are also directly interrelated and may be important in glucose-derived oxidative stress. This last effect is a direct inducer of apoptosis and mediates neuronal and endothelial cell death in diabetic neuropathy and vasculopathy, through binding of AGEs to a specific receptor, RAGE. This receptor mediates inflammation and binds other proteins such as the nuclear protein high mobility group box-1 (HMGB-1) released from damaged glial cells and neurones and is implicated in Parkinson's and

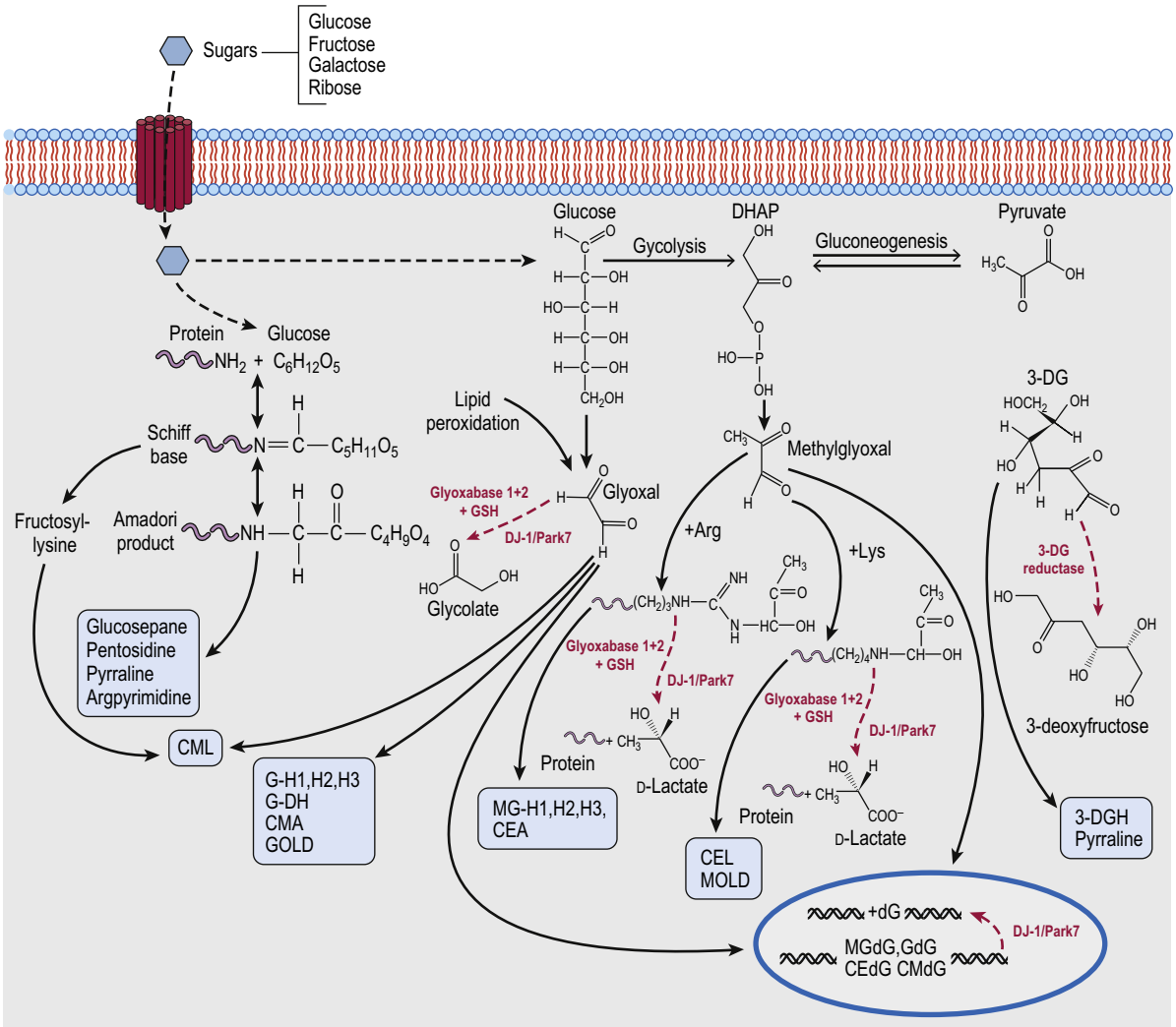


Fig. 4.32 Biochemical pathways involved in the formation of advanced glycation end-products (*AGEs*) from dietary sugars. AGEs are derived from several pathways including the Maillard reaction which generates Schiff bases and Amadori products as well as several reactions which involve glyoxalase enzyme (in *red*). (From Rowan et al., 2018.)

Alzheimer's disease as well as diabetes and ageing generally. RAGE also has a positive function because it is involved in dDNA repair (Kumar, et al., 2017).

A number of enzymes are known to repair proteins in the early stages of glycation, including fructosamine-3-kinase and glyoxalase and the term 'carbonyl stress' is used to describe imbalance in glycation versus deglycation.

Interestingly, the free radical scavenger vitamin B_6 may inhibit AGE production via the inhibition of oxidative degradation of Amadori intermediates and the trapping of reactive oxygen products (see later).

Glucose and lipid metabolism

Glucose and lipid metabolism are also intimately linked via the metabolite acetyl coenzyme A. In the presence of excess glucose, production of fatty acids and cholesterol through acetyl coenzyme A is increased with consequent increases in phospholipids (see later) and circulating levels of very low-density lipoproteins. While this has well-known implications for the development of atherosclerosis, more recently this form of metabolic dysfunction has been suggested to underlie pathologies such as age-related macular degeneration, not simply through the production of abnormal lipid

deposits in areas where their removal is difficult, such as the subretinal space, but also by a concomitant increase in lipid-based free radicals (see later).

Oxidation/reduction and free radical production

Oxidative metabolism and the generation of ATP (energy storage) are conducted via the mitochondrial cytochrome enzyme system. However, during the process of oxygen consumption, a small amount (<5%) is metabolized by alternative pathways. Univalent reduction of oxygen produces highly reactive free radicals, namely the superoxide anion and the hydroxyl radical, and the toxic molecule hydrogen peroxide (Fig. 4.33).

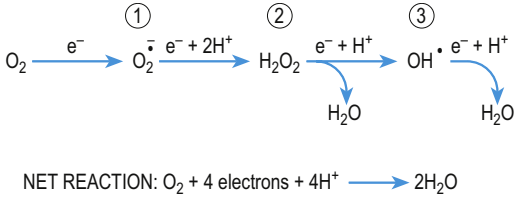


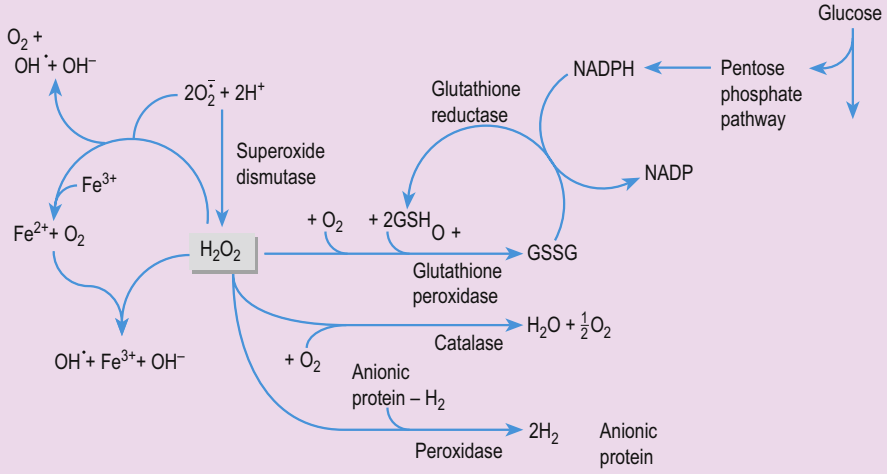
Fig. 4.33 Free radical production: *(1)* superoxide anion; *(2)* hydrogen peroxide; *(3)* hydroxyl radical.

The superoxide anion is generated in mitochondria in the ubiquinone cytochrome b system (see Fig. 4.7) and in other membranes by auto-oxidation via cytochrome P_{450}-linked reductases (also part of the drug-metabolizing system), which are rich in the vitamin flavoprotein. Superoxide is then converted by superoxide dismutase to hydrogen peroxide (H_2O_2), which diffuses into the cytosol via the smooth endoplasmic reticulum. Superoxide is also generated in the cytosol itself via recycling of redox enzymes such as xanthine oxidase; conversion to H_2O_2 in the cytosol is rapid. Xanthine, in particular, accumulates in certain tissues during ischaemia and subsequent reperfusion leads to massive release of free radicals, producing extensive tissue damage. Free radical scavengers can prevent this.

Free radicals can be generated by other mechanisms. Reduction of Fe^{3+} to Fe^{2+}, by reducing agents such as ascorbate, catalyses the conversion of H_2O_2 to $OH^\bullet$, and auto-oxidation of other compounds such as thiols and catecholamines produces free radicals. The Fe^{2+} ion has particular relevance to the problem of retained intraocular metallic foreign bodies. Tissue damage during inflammation is in a large part attributable to the production of free radicals by phagocytic cells during the respiratory burst (see Chapter 7, p. 396). These interactions are illustrated in Box 4.6.

Box 4.6 Generation of free radicals from hydrogen peroxide

The oxidation of hydrogen peroxide (H_2O_2) leads to glutathione consumption via the pentose phosphate pathway and requires a supply of glucose to maintain homeostasis.



Free radicals are detoxified by enzymatic and non-enzymatic mechanisms (Table 4.6). Melatonin and N-acetyl serotonin also act as antioxidants in the brain. A critical redox regulatory system involving glutathione and the enzyme glutathione S-transferase is central to cellular homeostasis (Fig. 4.34). Superoxide dismutases require divalent ions such as Mn^{2+}, Cu^{2+} and/or Zn^{2+} for normal activity; some of which occur in ocular tissues. H_2O_2 is further reduced both by enzymatic or non-enzymatic means, while several other mechanisms target removal of other reactive oxygen species (see Box 4.6).

The mechanism of cell damage varies with each molecular species. Chelated metal ions are important in non-enzymatic degradation of H_2O_2, producing the highly reactive $OH^{\bullet}$ radical, which is particularly damaging to cell membranes. H_2O_2 causes damage by inhibiting glycolysis and glucose uptake and is the most stable of the reactive oxygen species as well as being generated by several intracellular processes.

Table 4.6 Detoxification of free radicals

Mechanism	Agent
Superoxide anion degrading enzymes	Superoxide dismutases
Antioxidants	Ascorbate
Free radical scavengers	Vitamins A and E
H_2O_2 degrading enzymes	Catalase, peroxidase

Further information on mechanisms of cell damage by free radicals is available online at https://expertconsult.inkling.com

Oxidative stress is considered one of the major mechanisms underlying DNA damage and has been correlated with changes in telomere length associated with ageing. Cells respond to DNA damage using a protein termed protein kinase ataxia telangiectasia mutated (ATM), which mediates DNA repair, cell cycle arrest and apoptosis. Oxidative stress can directly activate dimeric ATM (usually the monomer is the

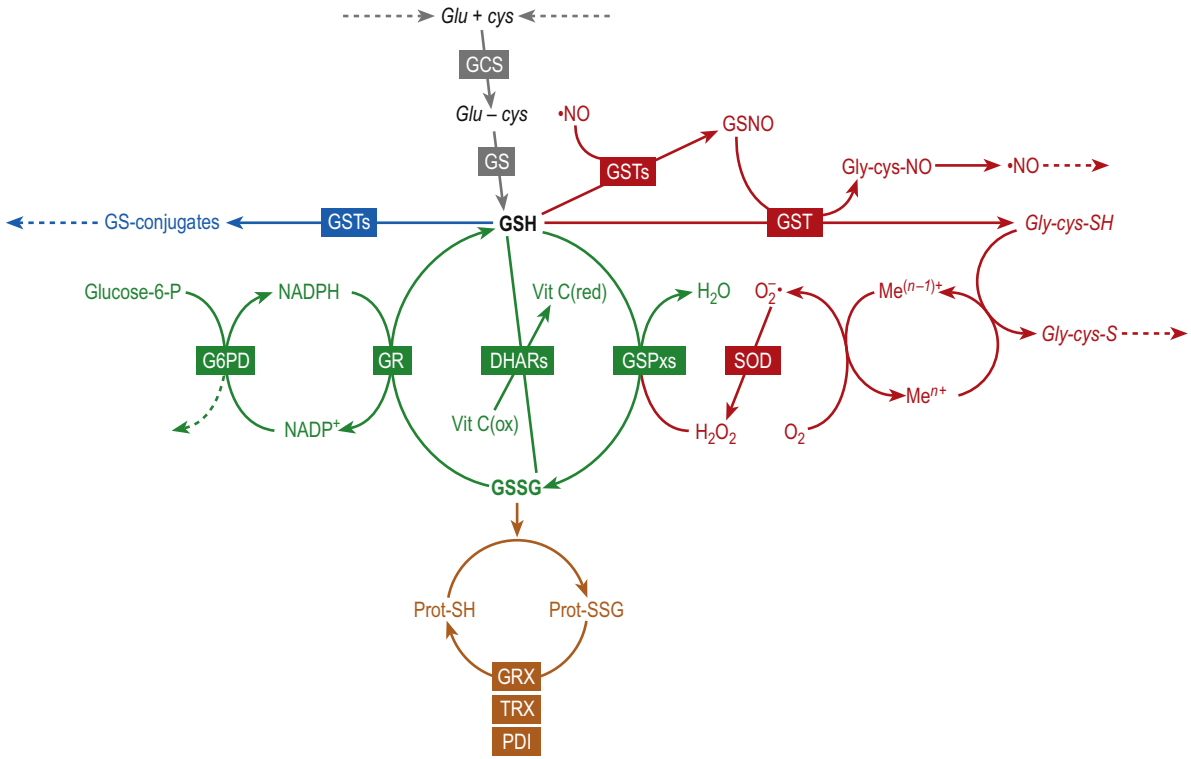


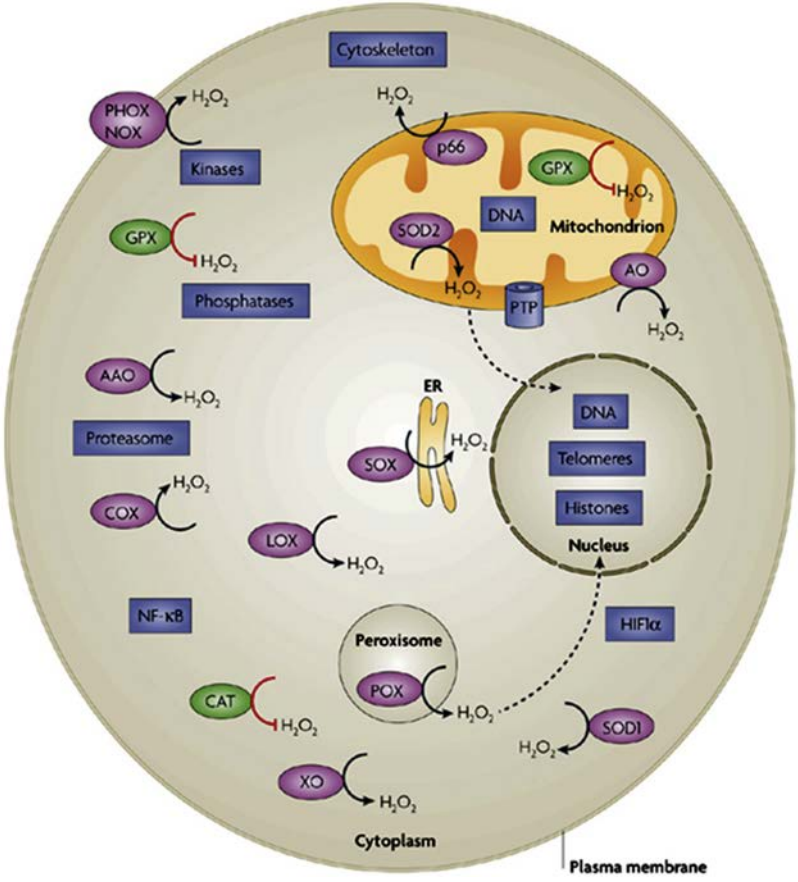
Fig. 4.34 Outline of the interrelations connecting the various roles played by GSH in cellular homeostasis: antitoxic *(blue)*, antioxidant *(green)*, pro-oxidant *(red)*, modulator *(orange)*. *γ-GCS*, γ-glutamyl-cysteine synthetase; *DHARs*, dehydroascorbate reductases; *G6PD*, glucose-6-phosphate dehydrogenase; *GPxs*, glutathione peroxidases; *GR*, glutathione reductase; *GRX*, glutaredoxin; *GS*, glutathione synthetase; *GSNO*, S-nitrosoglutathione; GSTs, glutathione S-transferases; *Me*, metal; *PDI*, protein disulphide isomerase; *SOD*, superoxide dismutase; *TRX*, thioredoxin. (From Pompella et al., 2003, with permission from Elsevier.)

Mechanism of cell damage by free radicals

Free radicals, and in particular H_2O_2, are involved in cell damage in various ways. They are not only a cause of acute damage and the stress response but also have a cumulative effect that may contribute to ageing. This may be more than simply a by-product of damage: some of the mitochondrial production of H_2O_2 is generated by an enzyme involved in the apoptosis pathway, which induces oxidative stress and may accelerate ageing (eFig. 4.16). H_2O_2 is produced by several enzymatic systems (indicated in purple), which generate H_2O_2 in different cellular compartments including phagocytic oxidases (PHOX) and NADP/H oxidases (NOX) in the plasma membrane; superoxide dismutase (SOD2), mitochondrial p66Shc (p66) and amine oxidase (AO) in mitochondria; peroxisomal oxidases (POX) in peroxisomes; sulphydryl oxidase (SOX) in the endoplasmic reticulum (ER); and amino acid oxidases (AAO), cyclooxygenase (COX), lipid oxygenase (LOX), xanthine oxidase (XO) and superoxide dismutase (SOD1) in the cytosol.

H_2O_2 may also disrupt cellular homeostasis by targeting regulatory macromolecules (in blue) including actin, myosin, tubulin (cytoskeleton), different kinases, several Ser/Thr and Tyr phosphatases, the proteasome, the mitochondrial permeability transition pore (PTP), mitochondrial and nuclear DNA, transcription factors such as HIF1 or nuclear factor (NF)-B, and histones and telomeres. Catalase (Cat) and glutathione peroxidase (GPX) enzymes buffer H_2O_2 to avoid cumulative oxidative events.

Thus, the genes that control production of H_2O_2 may have a say in determining lifespan of the cell (Giorgio et al., 2007).



eFig. 4.16 (1) Several enzymatic systems (indicated in *purple*) generate H_2O_2 in different cellular compartments (2) H_2O_2 disrupts cellular homeostasis by targeting regulatory macromolecules *(blue)* (see text for details). (From Giorgio et al., 2007.)

active agent), which seems to act as a cellular sensor for potential oxidative damage. As indicated above the ligand binding to RAGE may have a beneficial effect in this response (Kumar).

Lipids and lipid peroxidation

Lipids are the structural basis on which cell membranes are built and are generally composed of a hydrophobic tail of two fatty acid chains bound to a hydrophilic (polar) head (Box 4.7). Fatty acids, thus, have at least three roles:

- as integral components of phospholipids and glycolipids
- as functional molecules, e.g. hormones and second messengers
- as energy stores in the form of triacylglycerols.

In addition to phospholipids and glycolipids, cholesterol is the third major lipid component of eukaryotic cell membranes; cholesterol is predominantly hydrophobic but has a hydrophilic group on carbon 3 (Fig. 4.35). In an aqueous medium, lipids arrange themselves in such a way that the polar (hydrophilic) groups face the aqueous phase, whereas the hydrophobic groups face each other (a micelle). However, owing to the bulky nature of the two chains of fatty acids, membrane lipids do not readily form micelles but group together as a lipid bilayer (Fig. 4.35). Cholesterol inserts itself into the lipid bilayer where it conveys membrane stability to the fluid bilayer (Fig. 4.36).

Phospholipid bilayers can reach macroscopic dimensions and are barriers to the diffusion of aqueous solutes but remain quite fluid themselves. Thus, they form an ideal material to act as biological membranes. The fluidity of lipid membranes is related to the length of the fatty acid chains, the number and nature of double bonds in the chain and, in eukaryotes, the content of cholesterol. Proteins suspended in lipid bilayers, such as rhodopsin, have very rapid lateral mobility within the membrane unless they are anchored across the membrane to cytoskeletal and/or matrix proteins; in contrast the polarity of the protein in the membrane is fixed.

The formation of lipid bilayers is a spontaneous event, due to the physical hydrophobic interactions between the hydrocarbon tails, assisted by polar interactions between the aqueous phase and the hydrophilic head, and between adjacent polar groups. Furthermore, lipid bilayers form closed compartments in which defects are self-sealing, even after disruption. In many membranes, however, proteins, especially receptors and co-receptors, are distributed in clusters while the lipids form lipid rafts which differ in their lipid composition from adjacent regions (see p. 167).

Fatty acids undergo physiological degradation during energy consumption by oxidation and polyunsaturated fatty acids (PUFAs) are particularly susceptible. Similarly, lipids in cell membranes can be oxidized. Indeed, peroxidation of lipids is one of the main sources of cell damage and polyunsaturated fatty acids are particularly susceptible to this form of non-enzymatic damage.

Lipid peroxides oxidize neighbouring fatty acids and set up chain reactions in the membrane. Phospholipase A_2, by releasing free fatty acids from the membrane, detoxifies this reaction. However, the arachidonate released is then converted to prostaglandins and other eicosanoids (Box 4.8). Arachidonates are also modified by free radicals in a non-enzymatic reaction. This produces a series of prostanoids, prostaglandin-like chemicals termed isoprostanes. In fact, isoprostanes are produced in much larger quantities than prostaglandins and are an *in vivo* biomarker of lipid peroxidation and, indirectly, tissue damage. Isoprostanes occur at low levels in normal healthy individuals, thus reflecting a level of oxidative injury not completely suppressed by natural antioxidants. However, hydroxyoctadecadienoic acids (HODEs), hydroxyeicosatetraenoic acids (HETEs) and hydroxycholesterols are sufficiently stable end products of lipid peroxidation to act as biomarkers of tissue damage as are isoprostanes and neuroprostanes. Isoprostanes have biological and potentially pathological effects on cell function, thus directly contributing to disease.

Release of carbonyl compounds (molecules containing a C=O double bond) from both carbohydrates and lipids can lead to protein modification by glycation (AGEs, see earlier) and by lipoxidation (ALEs, advanced lipo-oxygenation end products). These are seen especially in uraemic patients on dialysis. Vitamin B_6 (pyridoxamine) may also inhibit ALE production as well as AGE production (see earlier).

Lipid peroxidation can be limited by vitamin E, a lipid-soluble free radical scavenger that is present within the bilayer (especially photoreceptors) and reacts directly with the lipid free radical to produce the phenoxy radical, or with ascorbic acid in the aqueous phase (see earlier).

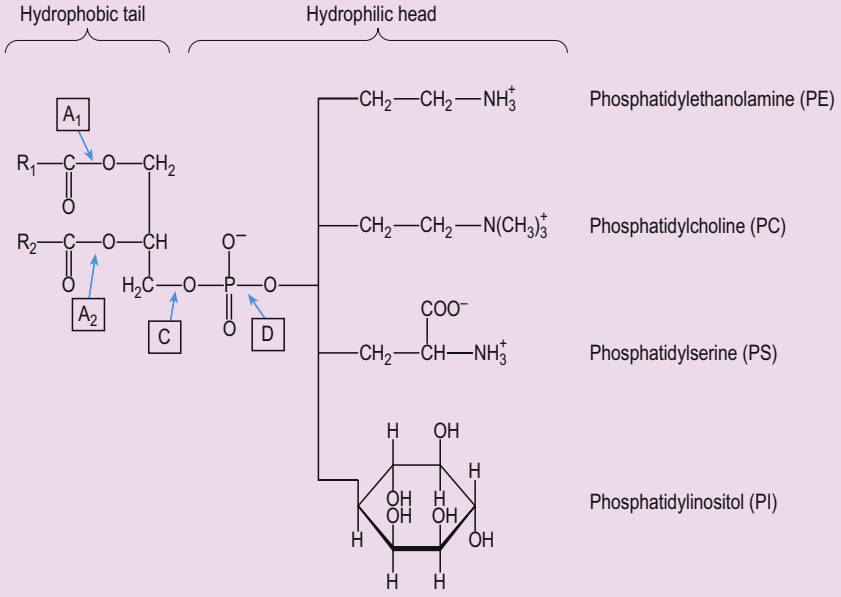
Lipids also occur extracellularly and are of particular importance as lipoprotein carriers in the circulation, and also as lubricants on surfaces. This is particularly important for the ocular surface.

The ocular surface

The ocular surface comprises the conjunctival and corneal non-keratinized epithelium and is bathed by the tear film. The integrity of the surface depends on true apposition of the eyelid margins and secretions from the meibomian and lacrimal glands. Collectively, this is termed the lacrimal functional unit (LFU). Evaporation of the tear film is clinically measurable as the tear film break-up time and blinking restores tear film integrity. Blink rate and break-up time are therefore linked. Tears provide lubrication for lid closure, assist in smoothing out the irregularities in the ocular surface (which otherwise may have transient effects on light transmission) and have an antibacterial function.

Box 4.7 **General structure of phospholipids**

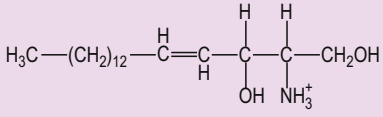
The general composition of lipids is a hydrophobic tail consisting of fatty acids bound to a hydrophilic head, which characterizes the molecule (**A**). Glycolipids differ from phospholipids in having a sugar moiety rather than a phosphorylcholine linked to the fatty acid chain. Sphingomyelin is intermediate between these (**B**).



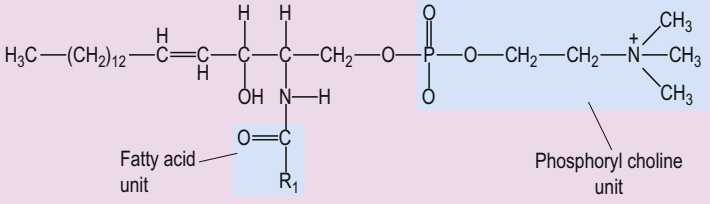
R_1 and R_2: Fatty acids (usually saturated and unsaturated, respectively)

A1 A2 C D : Sites of phospholipase activity

A



Sphingosine



Sphingomyelin

B

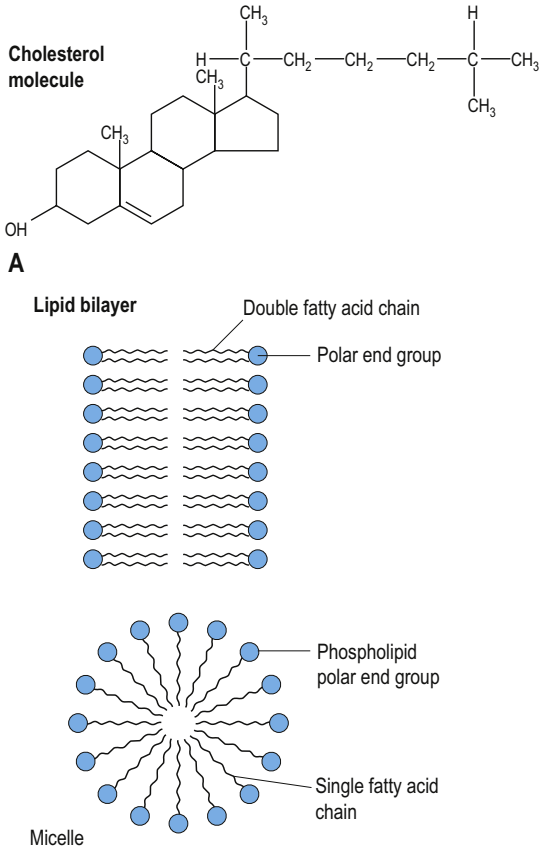


Fig. 4.35 (**A**) Structure of the cholesterol molecule and (**B**) the lipid bilayer.

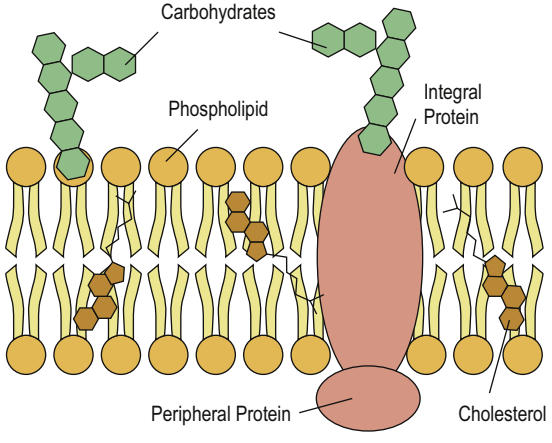


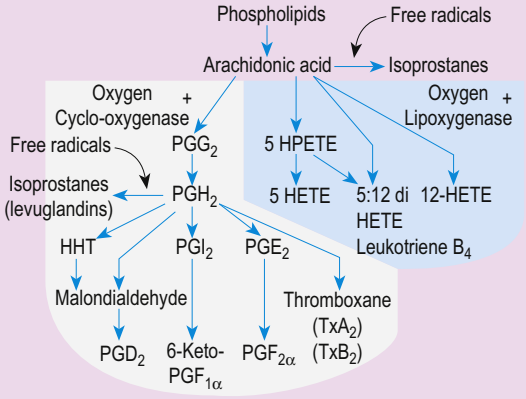
Fig. 4.36 Cholesterol embedded within the lipid membrane where it provides stability. (From Swafford, 2015.)

Box 4.8 **Oxidation of lipid membranes**

Phospholipids, especially those containing arachidonic acid (20:4) and docosahexaenoic acid (22:6), a fatty acid abundant in photoreceptor cells, are oxidized to lipid free radicals, which are then conjugated to form dienes. These are then converted to lipid peroxides.

$$RH \rightarrow R^{\bullet} + O_2 \rightarrow ROO^{\bullet}$$

Peroxidation of lipids in cell membranes leads to the release of large quantities of arachidonate, which acts as a substrate for eicosanoid production.



The tear film

The tear film is a protective covering for the cornea composed of three layers: a surface oily layer, an aqueous layer and a deep mucous layer. Previous estimates of tear film thickness suggested that the aqueous layer comprised more than 95% of the tear film, but this has been reduced to 60%, with a larger component being provided by the mucous layer based on evidence from *in vivo* studies using the confocal corneal microscope. In addition, clinical evaluation of the tear film involves an assessment of the tear meniscus height (TMH). The TMH reflects the volume of tears that collects at the contact line between the eyelid margin and the bulbar conjunctiva and has several other names, including the inferior marginal strip and the tear prism or rivus. It is possible to estimate TMH and depth (TMD) using optical coherence tomography (OCT)(see p. 227).

The cornea presents a hydrophobic non-wettable surface, which is made wettable by a layer of mucus on its surface. This is overlaid by the aqueous component of the tear film and lastly by a layer of lipid, which prevents evaporation of the tears (Fig. 4.37). Soluble mucus in the aqueous layer and meibomian lipid combine to ensure stability of the tear film by lowering the surface tension and

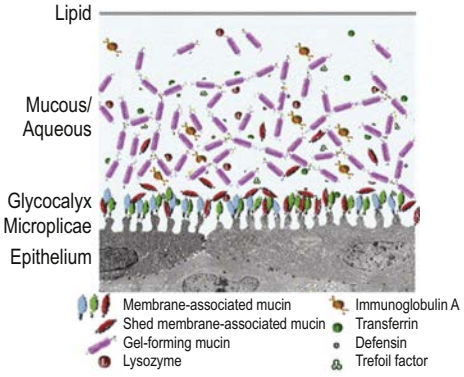


Fig. 4.37 Diagram of the tear film–glycocalyx interface demonstrating tear components and the apical surface glycocalyx with its membrane-tethered mucins. (From Govindarajan and Gipson, 2010.)

permitting spreading of the tear film on blinking. The tear break-up time (TBU) represents the time taken for aqueous to evaporate and meibomian lipid to come in contact with the hydrophobic epithelial cell layer. Sophisticated measurements of the TBU have been made using mathematical modelling techniques (Fig. 4.38).

The blink reflex removes up to 70% of the surface aqueous–mucous layer while the remaining 30%–40% thick mucous layer in contact with the cornea is not wiped away on blinking. Instead, it is thinned by blinking and may take 30 minutes to reconstitute itself. The mucous layer is composed of the glycocalyx of the epithelial cells (an intrinsic layer of carbohydrate-rich glycoproteins and glycolipids covering the cell surface) and an additional layer of tear mucins produced by the conjunctival goblet cells. Reduced TBU time in dry eye disorders of various types therefore reflects a disturbance in tear mucin/aqueous protein

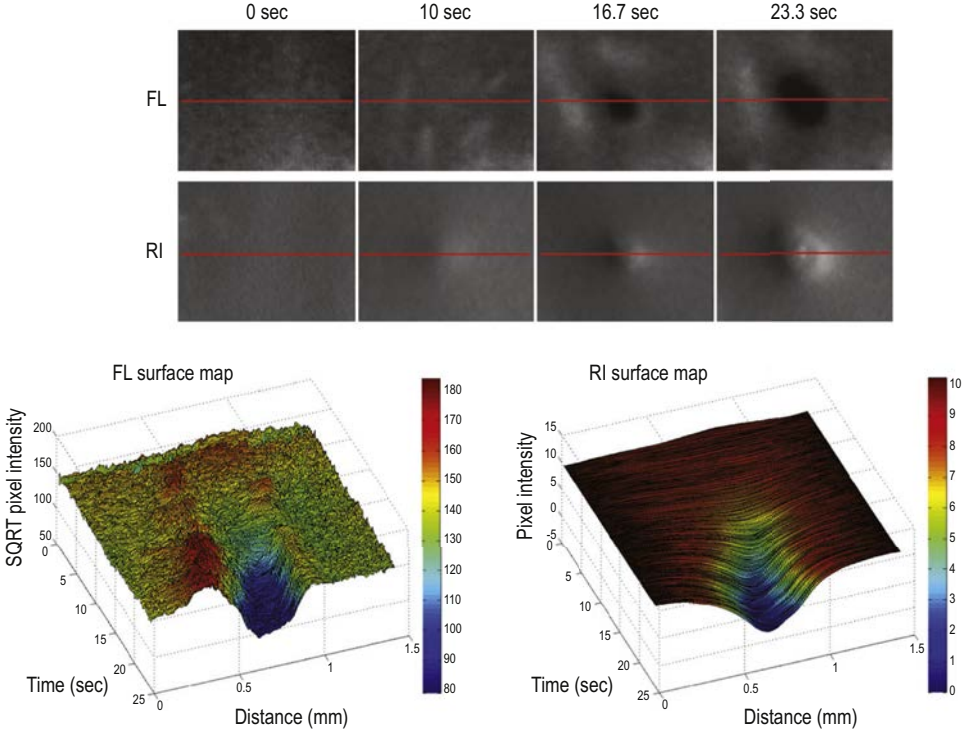


Fig. 4.38 Tear film breakup (*TBU*): dynamics and imaging: novel imaging techniques reveal the disruption of the tear film over time using simultaneous lipid interferometry and fluorescence intensity imaging. (From Braun et al., 2018.)

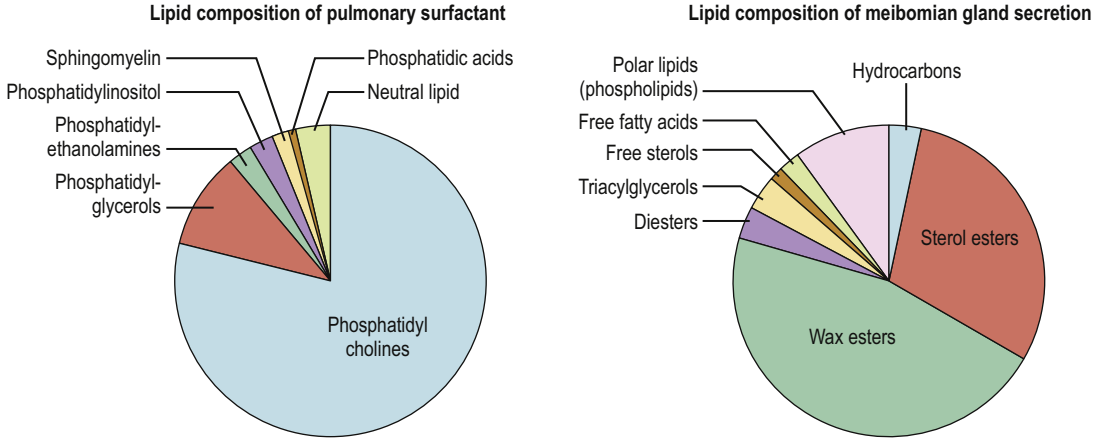


Fig. 4.39 Proportion of different lipid moieties compared between composition of pulmonary surfactant, a lipid-rich surface layer covering the lung mucosa, and the meibomian gland secretion, the major contribution to tear film lipid. Note the diversity of lipids in the tear film, particularly the wax esters. (From Panaser and Tighe, 2012.)

interaction and may be associated with a reduction in goblet cell density.

Tear film lipids have unique characteristics

The lipid layer of the tear film is derived from meibomian gland secretions and is very thin (0.1 μm). It is composed of a mixture of polar and neutral lipids (Box 4.7) with a melting point (35°C) that ensures it is always fluid on the ocular surface. The polar lipids are in contact with the aqueous phase of the tear film and provide structural stability to the tear film, while the non-polar lipids are at the air-tear film interface and provide barrier function and thixotropic properties. The lipids include unsaturated and branched-chain fatty acids and alcohols, 8–32 carbon chains in length. Tear lipids are under study by way of lipidomics and different sets of wax esters, oleates and cholesterol esters have been identified. Interestingly, in comparison with lung surfactant, there is a more heterogeneous composition with prominence of wax esters (Fig. 4.39).

Tear lipids promote movement of water into the aqueous phase during formation of the tear film, where they bond to tear lipocalins. The functions of the meibomian lipid layer are:

- to prevent evaporation of tears
- to prevent spillover of tears at the lid margin (stabilizes the tear film
- to prevent migration of skin lipid onto the ocular surface
- to provide a clear optical medium.

Tear film lipids can be evaluated clinically by meibography (Fig. 4.40) and its interaction with the aqueous layer by tear interferometry. Disturbance of Meibomian gland lipid is the single major cause of dry eye syndrome.

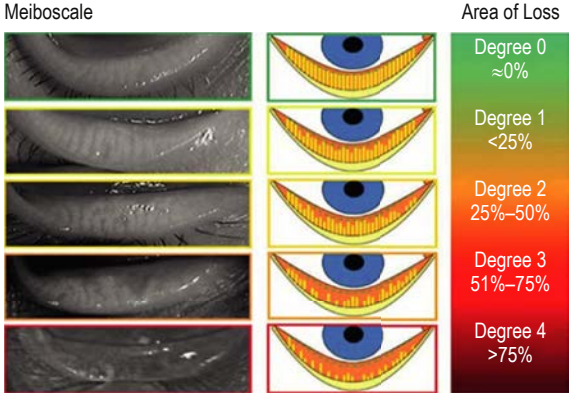


Fig. 4.40 Meibography using imaging techniques to evaluate semi-quantitatively the appearance and function of the meibomian gland using a clinical scale shown on the *left*, the meiboscale. (From Elder and Srinivasan.)

Lacrimal gland secretion provides the aqueous component of tears

The lacrimal gland and its accessory glands (see Chapter 1, p. 91) are classic exocrine acinar glands secreting a dilute aqueous solution containing proteins and small molecular weight components and electrolytes (Table 4.7). Although over 1500 proteins have been identified in tear fluid by proteome analysis, the principal proteins in tears

Table 4.7 **Composition of the tear film**

Physical properties	Solutes (μmol/L)	Proteins	Enzymes/inhibitors
98% H_2O_2	Na 120–160	Lysozyme	Glycolytic
6–9 μL volume	Cl 118–135	Lactoferrin	Amylase
pH 7.5	HCO_3 20–25	Tear-specific prealbumin (lipocalin)	Plasminogen activator
310–334 mOsm	K 20–42	G protein (IgA, IgG)	(α_2-macroglobulin) (α_1-antitrypsin)
	Mg 0.7–0.9	(Ceruloplasmin)	
	Ca 0.5–1.1	(Albumin)	
	Glucose 0.5–0.7	(Orosomucoid)	
	Retinol		
	Urea		

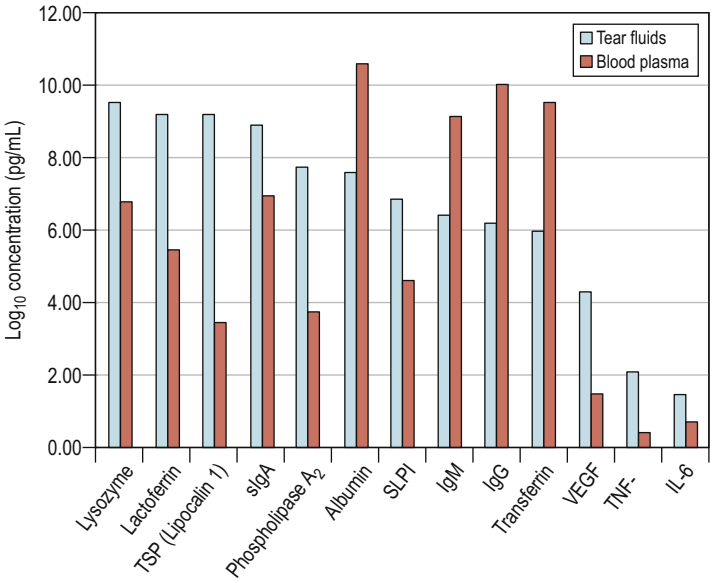


Fig. 4.41 Comparison of protein concentrations in human tear fluid and plasma to show the remarkable difference between tear fluid and plasma. The proteins shown in this figure represent a wide dynamic range (from mg/mL to pg/mL). (From Zhou and Beuerman, 2012.)

are immunoglobulin A (IgA), lactoferrin, G protein, tear-specific prealbumin (lipocalin) and lysozyme (Fig. 4.41). Tear-specific prealbumin is a member of the lipocalin superfamily and, together with a further lipocalin, apolipoprotein D, is secreted by the lacrimal gland. The function of tear lipocalin is to interact with meibomian gland lipid and induce surface lipid spreading. It may also have a role in removing harmful lipophilic molecules. Most lipids bind to lipocalin apart from some di- and tri-glycerides, squalene and omega fatty acids, which reside close to the air-tear interface (Glasgow, 2018).

Tear secretion by the lacrimal gland is under neural control by the autonomic nervous system (Fig. 4.42). Basal tear secretion occurs at a rate of 1.2 μL/min but can be massively increased by a variety of mechanical and psychophysical stimuli.

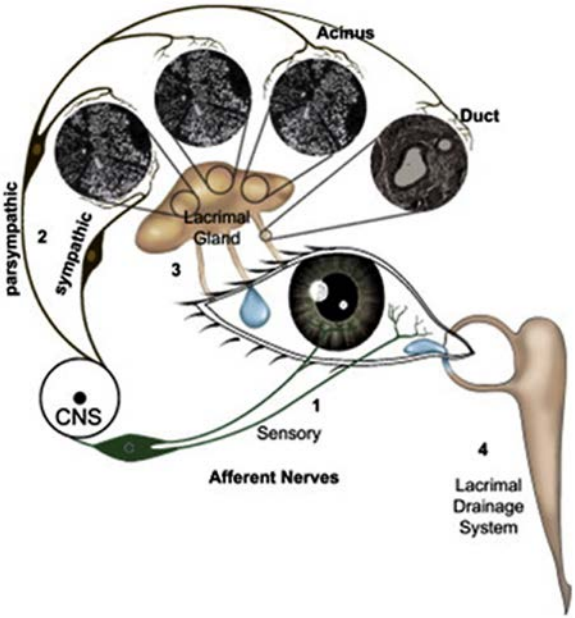
Further information on control of lacrimal gland secretion is available at https://expertconsult.inkling.com

A pheromone, exocrine gland secreting protein (ESP22), has been described in tears of prepubertal mice which inhibits mating behaviour in adult mice (Ferrero, 2013).

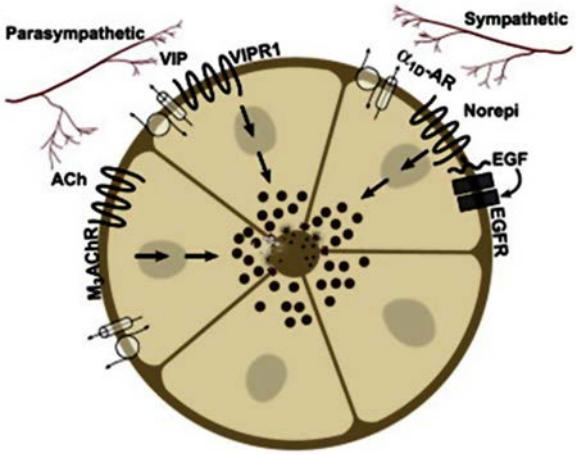
Several tear proteins have antibacterial activity (e.g. lactoferrin and lysozyme), and the high levels of immunoglobulins may have a protective function (see Chapter 7, p. 406). Lactoferrin synergizes with lysozyme in its antibacterial action by

Lacrimal gland secretion

Tight neural control of lacrimal secretion is essential to healthy functioning of the tear film. The three major components of the aqueous secretion (water, electrolytes and proteins) are each secreted through different mechanisms, but there is overall neural regulation of the final secreted product so that a coordinated response can be achieved. Sensory input comes via the trigeminal nerve endings in the cornea and conjunctivae to induce efferent parasympathetic and sympathetic nerves that innervate the lacrimal gland (eFig. 4.17) to release neurotransmitters which effect the release of the tear fluid (eFig. 4.18).



eFig. 4.17 Diagram showing the neural anatomy and physiology of lacrimal gland tear secretion.



eFig. 4.18 Release of acetyl choline occurs from parasympathetic nerve endings as well as vasoactive intestinal peptide (VIP), while norepinephrine is released by sympathetic nerves. Additional mediators such as EGF and a1d-AR contribute to the overall response.

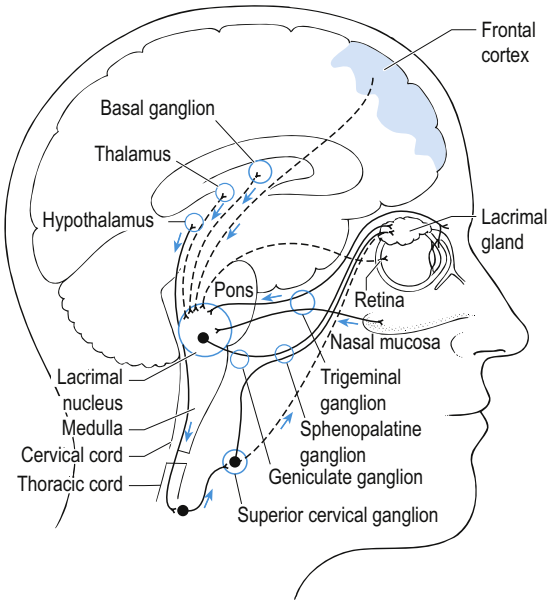


Fig. 4.42 Neural control of tear secretion.

binding to lipoteichoic acid on the bacterial surface and allowing access of lysozyme to the peptidoglycan. The aqueous component also has anti-adhesive and lubricant properties possibly attributable to lipocalin, thus ensuring that protein and debris generally do not adhere to the corneal surface (e.g. during prolonged periods of lid closure while sleeping).

The mucous layer stabilizes the aqueous layer by providing a hydrophilic contact surface

The mucous layer is composed of the glycocalyx of the epithelial cell surface and a layer of tear-specific mucoproteins (Fig. 4.37). Further tear mucins are secreted by the conjunctival goblet cells.

Mucus imparts viscosity to the tear film as a result of their extremely high levels of glycosylation. Mucus glycoproteins (termed mucins; ~20 different mucins are known, of various size up to 50 × 10^6 kDa) have two characteristic features: (a) multiple tandem repeats of serine and threonine, which link N-acetyl-galactosamine (GalNAc) to galactose via O-glycosylation, and (b) a strong negatively charged hydrophilic surface resulting from extensive O-glycosylation. The molecular structure of mucin has been likened to a bottle-brush, the hairs of the brush are representing the multiple *O*-glycosylated short oligosaccharide chains on an elongated protein 'handle' containing threonine and serine residues (Fig. 4.43). This structure has the capacity to form many interactions with other hydrophobic (lipid) as well as hydrophilic (protein) and charged molecules, thus forming mucus.

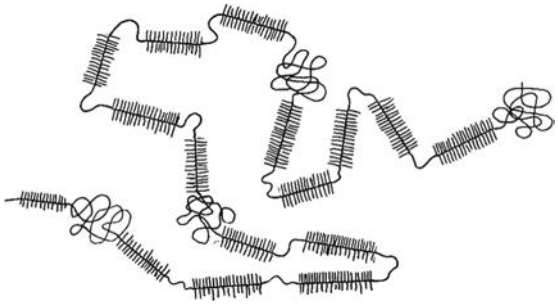


Fig. 4.43 Bottle-brush mucin. (Courtesy J Tiffany.)

There is some tissue specificity concerning the type of mucin on mucosal surfaces. Mucins are either secreted or membrane-tethered: in tears, the main secreted mucins are produced by goblet cells (MUC5AC), by lacrimal acinar cells (MUC7) and by corneal and conjunctival epithelium (MUC1, 4, 16 and 20). Thus, MUC5AC is secreted along with trefoil peptides TFF1 and TFF3. In contrast, MUC1, MUC4 and MUC16 and the sialomucin complex are associated with the epithelial glycocalyx and, although regarded as membrane mucins, they may be released into the aqueous phase of tears through activity of ADAMTS-1 (a disintegrin-like metalloproteinase with thrombospondin type 1 motif) metalloproteases (see Chapter 7, p. 394, Table 7.2). The ocular surface membrane-tethered mucins are considered to have anti-adhesive properties, particularly MUC16 for bacterial adherence. Mucins, especially MUC5AC give a gel-like quality to tears.

Marx's line, a lissamine green and fluorescein staining border of the tear meniscus at the mucocutaneous junction between the conjunctiva and the eyelid skin, an important clinical feature representing a zone of tear hyperosmolarity. It is thought that conjunctival epithelial cells have a higher cell cycle rate at this site, because of the mechanical force of repeated eyelid closure and produce less tear mucins such as MUC16, allowing greater penetration by dyes such as lissamine green. The placement of this line changes in dry eye disease and other pathologies.

Tear secretion is under psychoneuroendocrine control

As indicated previously (Fig. 4.42), control of tear secretion is mostly through the autonomic system and thus tear secretion is affected by instillation of drugs that modulate this system (e.g. pilocarpine, atropine). Apart from direct effects on the acinar cells (see earlier, intracellular second messengers), the parasympathetic system can markedly increase tear flow via its effects on the myoepithelial cells surrounding the acinar cells (see Chapter 1, p. 92 and p. 209).

Further information on control of lacrimal gland secretion is available online at https://expertconsult.inkling.com

Hormonal control of tear secretion is well recognized but not clearly defined. Reduction in tear flow occurs in women after the menopause, while testosterone stimulates secretion of certain tear components, such as IgA.

The conjunctiva

The conjunctiva contains a variety of specialized cells that allow it to act as the binder for and the source of many tear film constituents. These include immune cells (T and B cells, mast cells, dendritic cells; see Chapter 7, pp. 425-450) and specialized mucus-secreting goblet cells.

The epithelium is more than a simple covering layer for the conjunctiva

The conjunctival epithelium can be regarded as intermediate in type between the keratinized squamous epithelium of skin and typical non-keratinized columnar mucosal epithelia, for instance of the respiratory and gastrointestinal tracts. However, it resembles the latter more closely and may require the multifunctional protein clusterin to inhibit keratin production. Clusterin is on the surface of cells lining body cavities, is highly expressed in the ocular surface and the concentration in tears has a critical threshold for development of dry eye disease. It is involved in transport of lipoproteins, inhibiting complement-mediated lysis and in modulation of cell–cell interactions.

Although the conjunctiva is non-keratinized, keratins are expressed by the conjunctival epithelium in a typical paired combination (K3/K12). However, there is much more K3/K12 in corneal than in conjunctival epithelium. Several other cytokeratins present in non-keratinized stratified (K4 and K13) or simple (K8 and K19) epithelia are also found.

The epithelium contains numerous innate immune cells (macrophages, mast cells, NK cells and dendritic cells) but also has a rich population of intraepithelial and stromal T cells organized to form the conjunctiva-associated lymphoid tissue (CALT), similar to the mucosa-associated lymphoid tissue (MALT) of other organs. Even within the CALT, the distribution of lymphocyte subsets varies from tarsal to bulbar to forniceal conjunctival regions (see Chapter 1, p. 88). In addition to its function as a barrier to external organisms, the epithelium is a major source of tear mucins derived from the intraepithelial goblet cells, which, like the lacrimal gland, are under neuroendocrine control (see earlier).

Further information on control of lacrimal gland secretion is available online at https://expertconsult.inkling.com

The turnover and health of the epithelial and goblet cells are markedly dependent on vitamin A and the retinoids and become abnormal in vitamin deficiencies, leading to a severe form of dry eye syndrome. In addition, certain accumulations of goblet cells occur, e.g. on the tarsal (lid) conjunctiva where they occur in crypts (Henle's crypts) and on the bulbar conjunctiva a few millimetres nasal of the limbus (Manz's glands).

In addition to goblet cells, there are other types of conjunctival epithelium, distinguished by ultrastructural appearances: secretory epithelia appear in mature and immature forms depending on their content of secretory granules and the presence of a Golgi complex; in contrast, other cells with a high content of RER and/or mitochondria are presumably involved in transport and epithelial regeneration. Differences in regional cell cycle kinetics are considered to underlie Marx line (see earlier).

The conjunctival stroma is highly vascular and contains aqueous veins

The stroma of the conjunctiva has a superficial lymphoid layer and a deep layer containing a rich plexus of vessels, including lymphatic vessels, that acts as a watershed between the intraocular circulation and the external circulation of the eye and lids. Through these vessels, waste materials from the anterior chamber of the eye are transported to the pre-auricular draining lymph nodes and venous drainage systems in the neck.

The lids

The lids function to protect the cornea and adnexal structures and have a highly specialized structure that ensures they are properly apposed to the surface of the globe (see Chapter 1, p. 85). Indeed, defects in lid apposition, as occur in diseases such as trachoma where there is scarring and deformation of the lids, lead to significant corneal exposure, ulceration and blindness (see Chapter 7, p. 398).

Closure of the lids leads to compression of the lipid layer of the tears such that it increases in thickness to about 1.0 mm at the lid margin. When the eyes open, the lipids are dispersed to form a lipid bilayer from the lower lid upwards as the eyelid opens, with the hydrophilic groups on the phospholipid molecules interacting with the aqueous compartment. A 'lid wiper' region in the upper eye-lid is now recognized and extends from the line of Marx to the upper subtarsal fold and disturbances of normal spreading are associated with 'lid-wiper epitheliopathy', which can be suspected by measuring the exerted lid pressure during blinking using a blepharon-tensiometer (Yamaguch and Shiraishi, 2018).

Control of lid movement is via the VII nerve for motor function and the V nerve for afferent input via mechanisms such as the blink reflex (see Chapter 1, pp. 73-81 and 85-93).

The lid also contains accessory lacrimal glands (glands of Krause in the fornix and glands of Wolfring in the tarsal lid margin), specialized mucus-secreting glands (they secrete MUC4, important in eyelid opening during development, rather than goblet cell-type MUC5AC see earlier) in the conjunctiva and Meibomian glands that contribute to the makeup of the tear film. Lashes and their hair follicles, in addition to sweat glands, are also important specialized structures in the lid, which protect the eye from foreign particles.

Incomplete eyelid closure may be physiological during sleep. Control of eyelid closure is under both reflex supranuclear and learned/conditioned (cerebellar) control.

Cornea and sclera

The outer coat of the eye (cornea/sclera) is a tough non-compressible layer of connective tissue that can withstand considerable deformation and pressure.

The cornea

The transparent cornea functions as the main optical lens of the eye, focusing transmitted light rays onto the retina. Its cellular and extracellular matrix components are of the same basic chemical composition as other opaque tissues in the body that scatter, rather than transmit, light rays. Scattering of light in opaque tissues is the result of the large disparity in refractive index (RI) between matrix components such as collagen (refractive index, RI=1.55 in the dry state) and glycosaminoglycans (RI=1.35). The cornea's ability to transmit light is, in essence, a function of how the cells and matrix components are organized within the tissue to reduce this disparity in RI.

Corneal transparency is a function of its relative acellularity and matrix structure

The epithelium. The six-cell-thick stratified layer that is the corneal epithelium (see Chapter 1, p. 15) presents the first refracting interface to transmitted light. Most of the light-absorbing properties of the cornea take place in this layer, mainly for short-wavelength light. However, electromagnetic radiation on the visible spectrum is transmitted through the epithelium.

The cells are typical keratin-expressing epithelial cells containing integrin receptors for basement membrane components such as fibronectin, laminin and collagen. Corneal epithelial cells express a particular combination of paired 54/64 kDa keratins (keratin 3/keratin 12). The 54 kDa protein may have a role in inflammatory eye disease, while the 64 kDa protein appears to be useful as a marker for differentiating cells of the central cornea from limbal stem cells (see later). Keratin 12 may be important in corneal epithelial junctions since K-12 knockout mice are prone to recurrent erosion (epithelial cell loss).

The cells are organized to present few interfaces, the most prominent being that between the basal cells and the basement membrane. Hemidesmosomes bind these cells to the basement membrane of the anterior corneal stroma through a band of anchoring fibrils, which pass through the lamina densa of the basement membrane proper into a densely woven network of collagen fibres (see Box 4.4 and Fig. 4.26) known as Bowman's layer (see Chapter 1, p. 16) (Fig. 4.44). Bowman's layer is about 12 μm thick and composed of types 1, and III with some Types V and V11 collagens. In addition, transplasma membrane collagen type XVI supports firm adhesion in these basal cells.

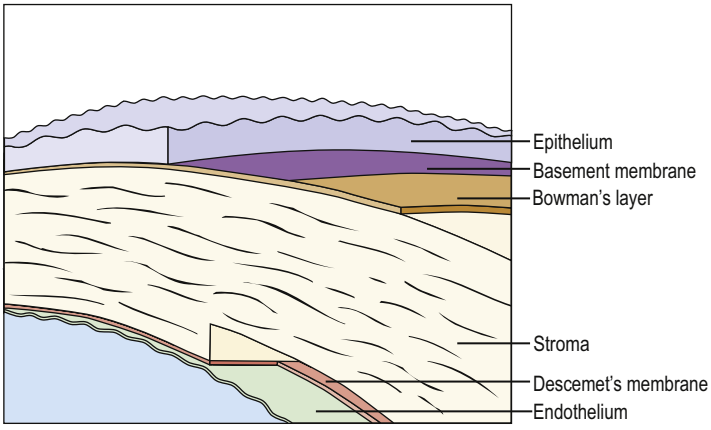


Fig. 4.44 Diagram demonstrating the layers of the human cornea: the anterior basement membrane, Bowman's layer, the stroma, Desçemet's membrane and the endothelium. (From Last et al., 2009.)

The epithelium presents an effective barrier to fluid transport, which is achieved by extensive close contacts and tight junctional complexes.

Further information on cadherins, filaments and cell junctions is available online at https://expertconsult.inkling.com

The site of the barrier is in the suprabasal epithelium and is mediated by high expression of the tight-junction protein claudin. Spot desmosomes are numerous and differences in the content of desmosomal proteins have been observed depending on the site. For instance, desmoglein and desmocollin are absent from basal limbal epithelial cells, which may have functional significance in their role as putative stem cells. Limbal stem cells arise from pouches or stem cell niches in the peripheral corneal epithelium.

What is a stem cell niche? Further information is available online at https://expertconsult.inkling.com

Matrix factors affecting transparency: collagen. The stroma in the human cornea accounts for 90% of the corneal thickness and thus contributes most to the optical function of the tissue. Several different types of collagen are present in the cornea. In addition to normal basement membrane types IV and VIII collagens at the epithelial and endothelial cell layers, the two specialized corneal regions, Bowman's layer and Descemet's membrane, contain collagens not normally found in other matrices. Bowman's layer is a condensation of type I and V/VII with a high proportion of type III in a matrix containing chondroitin and dermatan sulphate, while Descemet's membrane contains high levels of novel collagens (types V, VIII, IX and XII) organized in a lattice arrangement. This provides elasticity and deformability to the cornea while maintaining high levels of light transmission. Descemet's membrane also imparts strength and resilience to the corneal stroma and is the main resistance to normal intraocular pressure. Recently, a further layer has been identified anterior to Descemet's layer (Dua's pre-Descemet's layer) as a result of advances in lamellar keratoplasty using an air bubble to separate the stroma from the endothelium (Dua, 2013).

Further information on Dua's layer is available online at https://expertconsult.inkling.com

The orthogonal lamellar arrangement of stromal type I collagen (which accounts for 50%–55% of stromal collagen) fibrils is considered to be important in determining corneal transparency. Transmission of light by the cornea is a result of the small fibril diameter (<30 nm) and low interfibrillar distance (~55 nm) preventing light scatter ('destructive interference' of Maurice). It is only when the distance between the regions of different refractive index becomes greater than 200 nm that light scatter occurs, for instance when corneal swelling occurs. The Type I/V/III randomly oriented collagen fibres in Bowman's are also very thin and allow light transfer. Type III collagen is normally laid down irregularly in the early phases of matrix deposition, e.g. in wound healing, which might account for the random arrangement of collagen fibres in Bowman's layer.

Type V (approximately 10%) and some type III (1%–2%) collagen also exist in the corneal stroma, while the remainder is made up of type VI collagen. Types I, III and V collagen are fibrillar collagens, but type VI has large nonhelical globular polypeptides at both the C- and N-termini of the helical protein backbone (see Table 4.2). Collagen fibrils are heterotypic (i.e. they contain both type 1 and V collagen) and type V is thought to initiate fibril formation and regulate fibril thickness (Fig. 4.45). However, fibril thickness is also dependent on the nature of the stromal glycosaminoglycans and particularly on the specific proteoglycan (see later).

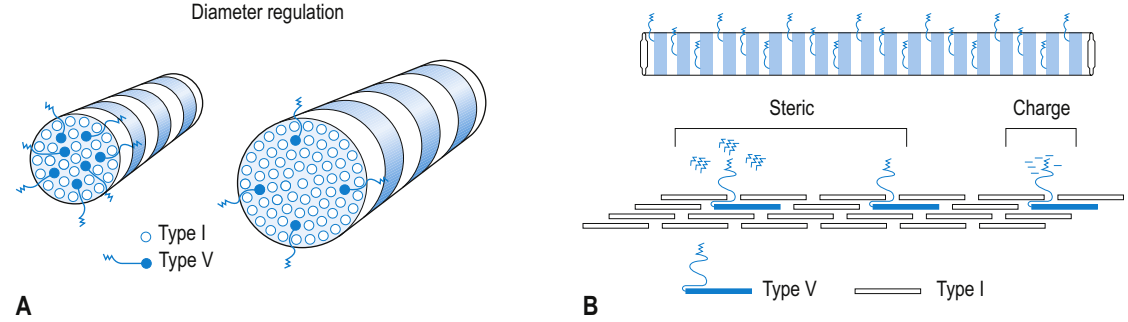
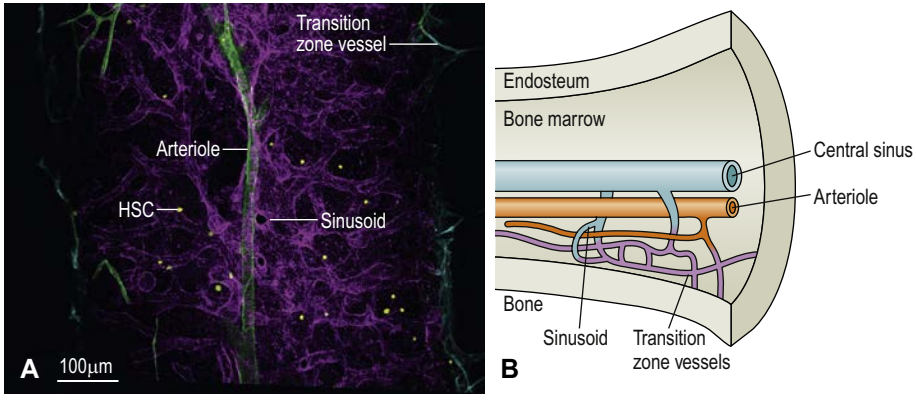


Fig. 4.45 Model for type V regulation of collagen fibril diameter. (**A**) Collagen fibrils within the cornea contain a high percentage of type V collagen and have small diameters and fibrils in other tissues possessing low levels of type V collagen have a large diameter. The experimental reduction of type V collagen generates fibrils that have the characteristics of those found in tissues with low type V levels. A mechanism by which type V collagen may limit fibril growth is shown in (**B**). Type V N-terminal domains project onto the fibril surface, and when sufficient numbers have accumulated, they block further accretion of collagen monomers and thereby limit growth in diameter. The N-terminal domain is large and possesses a number of acidic residues, and so may affect this block using steric and/or electrostatic hindrance. (Courtesy David E. Birk.)

The stem cell niche

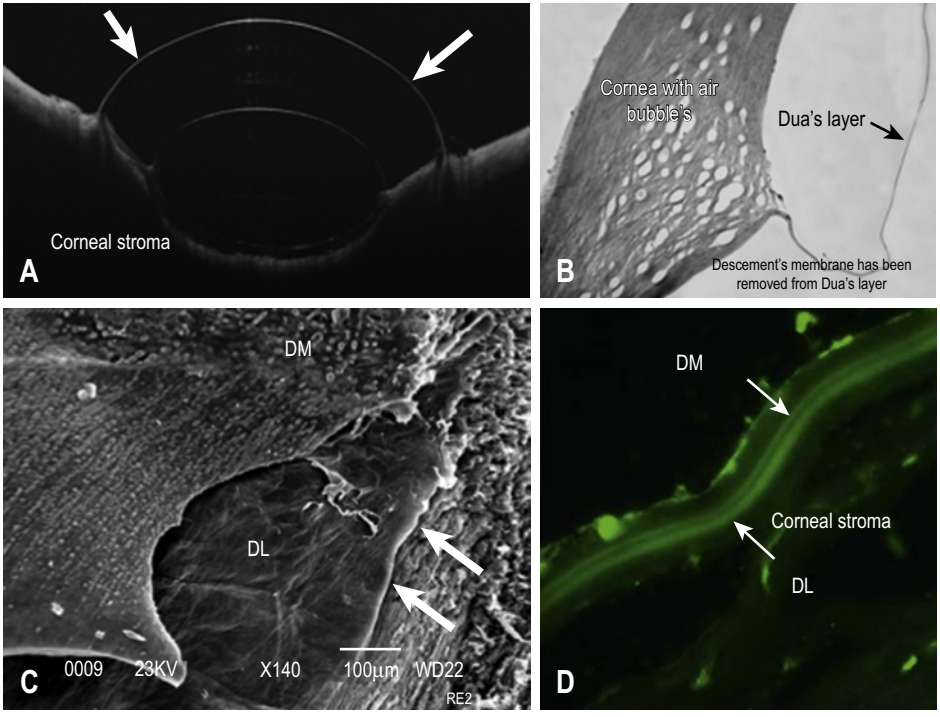
A niche can be described as a protected physical environment. Stem cell niches exist for most if not all differentiated tissues. The most studied niches include the stem cell niche in the intestinal crypt and the hematopoietic stem cell (HSC) niche in the bone marrow. The stem cell niche in the intestine is a readily identifiable region of the crypt, protected by Paneth cells of the gut and a specialized metabolic microenvironment. The HSC niche in the bone marrow is less well defined and is a complex network of interacting cells (eFig. 4.19). Hematopoietic stem cells generate the many types of myeloid, erythroid and lymphoid cells as well as platelets. It is separate from the mesenchymal stem cell niche which also exists in the bone marrow and produces mesenchymal stem cells which have immunosuppressive properties (see Chapter 7, p. 41). The limbal stem cell niche is situated at the bottom of the corneal palisades of Vogt (see p. 219).



eFig. 4.19 *Left panel:* An image of the vasculature in the bone marrow showing the distribution of α-catulin⁺ haematopoietic stem cells (HSCs, *yellow*) relative to sinusoids *(purple)*, arterioles *(green)* and transition zone vessels *(blue)*. The HSC niches are randomly distributed throughout the bone marrow in small foci. *Right panel:* Diagram to show relationship of vasculature to the HSCs in left panel. (From Crane et al., 2017.)

Dua's layer

Advances in corneal transplantation in which specific layers of the cornea are selectively grafted required development of techniques to separate the endothelium from the deep stroma. These new techniques allowed endothelial keratoplasty (EK) and deep anterior lamellar keratoplasty (DALK) to be performed and required the inspissation of an air bubble into the stroma. Harminder Dua found that the air bubble naturally developed in certain planes, one of which was a layer immediately anterior to Desçemet's membrane which retained most of the bubble but was leaky at certain points in the periphery (eFig. 4.20). Dua's layer's ability to retain air in this fashion facilitated the DALK procedure which has now become a preferred technique for many corneal surgeons. However, penetrating keratoplasty remains the main option for patients with extensive stromal disease and endothelial loss as occurs in developing countries but is rarely performed because of a shortage of donor grafts and difficulties in post-operative patient care.



eFig. 4.20 Dua's layer can be revealed by air injection at different angles in the stroma of a human eye. (**A**) Optical coherence tomography of Dua's layer (*DL*) *(lower arrows)* separated by a bubble from Desçemet's membrane (*DM*) *(upper arrows)*. (**B**) Light micrograph of DL from which the DM has been peeled off. DL forms the posterior wall of a type-1 big bubble created by injection of air in the human corneal stroma. (**C**) Scanning electron microscopy of DM and DL (*arrows* show edge of DL). (**D**) Immunohistochemistry of DM and DL to show presence of elastin. The anterior surface of DM shows a distinct band of elastin, apposed to the DL, which shows as a thicker band of elastin separated from the DM by a narrow non-staining band of interfacial matrix. High elastin content is a distinct feature of the DL. (Courtesy H. S. Dua.)

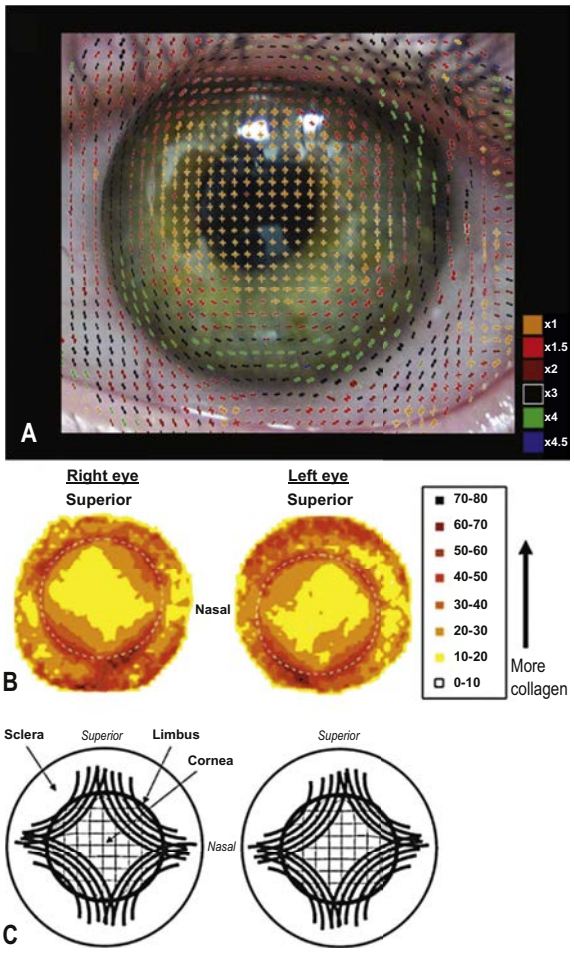


Fig. 4.46 (**A**) X-ray diffraction data showing the preferred directions of stromal collagen lamellae across the human cornea, limbus and anterior sclera. The scale of the X-ray vector plots (indicated in the *colour key*) reflects the degree of collagen alignment. (**B**) Maps of X-ray scatter intensity showing the distribution of preferentially aligned lamellae in a pair of eyes from the same human donor. The limbus is denoted by a *broken line*. Note the symmetry between left and right eyes. (**C**) Theoretical model showing the net course of lamellae and based on the data shown in (**A**) and (**B**). Lamellae are thought to change direction in the corneal periphery by a process of splitting and interweaving. (Modified from Meek and Boote, 2009.)

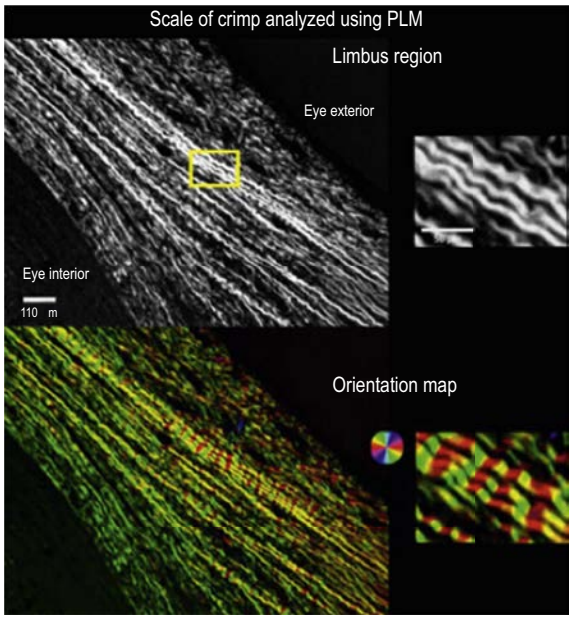


Fig. 4.47 The scale of collagen crimp analysed using PLM. Wide-field views of the limbus region *(L)* show wavy patterns in the collagen fibres emphasized using a colour map of orientation. Higher magnification images *(R)* show the periodic waviness of the fibres. (From Jan et al., 2018.)

Stromal collagen fibril lamellae are arranged in parallel, running at oblique angles to each other (see Chapter 1, p. 18). Apart from their small size, their uniformity of thickness is likely to be a major factor in light transmission. The parallel arrangement of the central corneal fibrils extends to the periphery where the fibrils adopt a concentric configuration to form a 'weave' at the limbus with some transversely running fibrils fusing with the circumferential collagen fibrils. Scleral collagen fibrils take a curved orientation across the peripheral cornea, thus acting as anchoring fibrils which flatten the peripheral cornea at this point (Fig. 4.46). This imparts considerable strength to the peripheral cornea and permits it to maintain its curvature and thus its optical properties.

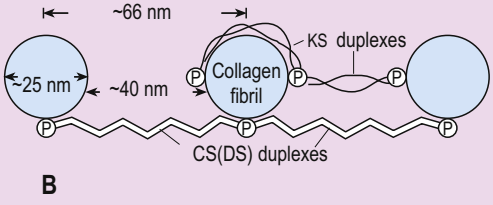
Collagen fibres are arranged in bundles or groups as well as sheets (e.g. in tendons). In their natural state they are wavy or 'crimped' and this applies to collagens in the eye. Crimped collagen can be visualized by polarized light microscopy (Fig. 4.47) and is a measure of its overall biomechanical strength and ability to resist stretching or force. In the eye this is reflected in the stiffening of the globe in response to high intraocular pressure. Crimp parameters such as waviness and tortuosity are significantly different for the peripheral cornea and limbus compared with the central cornea, indicating marked differences in biomechanical properties of these regions.

Glycosaminoglycans. The corneal stromal matrix is unusual in that it contains no hyaluronan, except at the limbus where there is a gradual increase in concentration towards the sclera. The major corneal glycosaminoglycan (GAG) is keratan sulphate; in the central cornea non-sulphated chondroitin is also present, while towards the periphery chondroitin sulphate is the second major GAG. Chondroitin-4-sulphate and dermatan sulphate are almost

Box 4.9 **Organization of tissue proteoglycans**

The organization of proteoglycans in the collagen matrix is probably as shown below: KS-PG (keratan sulphate proteoglycan) double complexes maintain the lateral interfibrillar distance, while CS-PGs (chondroitin sulphate proteoglycans) regulate the overall longitudinal arrangement of the fibrils by spanning three fibrils (**A**). Proteoglycans bind directly to the collagen fibrils, probably through the 'minor' collagens such as V or VI that co-distribute with the type I fibrils and have non-collagenous polypeptide domains to interact with the proteoglycans (**A** and **B**).





(Figure courtesy J Scott and IRL Press.)

identical (see Box 4.5) and many believe that the second major GAG is not chondroitin but dermatan sulphate.

Corneal GAGs exist in the native state as proteoglycans (PGs); four major forms exist in the cornea: decorin, lumican, keratocan and mimecan. Decorin- and lumican-deficient mice have reduced corneal transparency as a result of disorganization of fibril arrangement, while keratocan (and less so mimecan) maintains overall corneal thickness. Both corneal dermatan sulphate and keratan sulphate proteoglycans belong to the class of small non-aggregating PGs (small leucine-rich PGs known also as SLRs).

The critical regions in proteoglycans are their linkage sites; for keratan sulphate the link is an *N*-glycosidic bond between *N*-acetylglucosamine and asparagine in the core protein. The terminal sites of the branched oligosaccharide

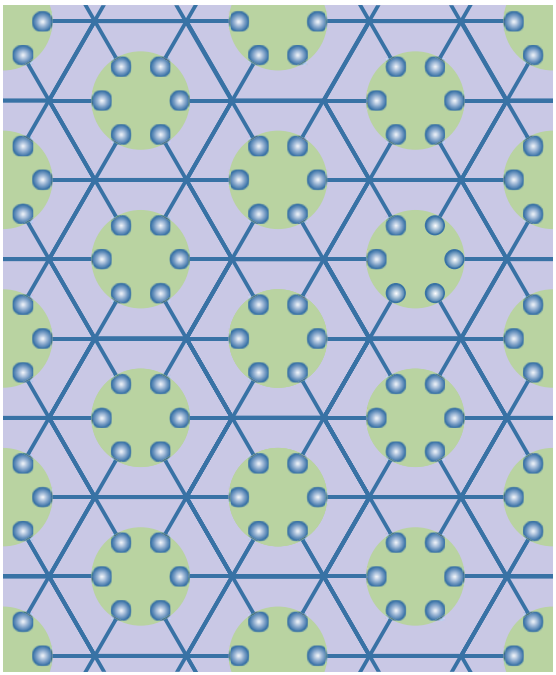


Fig. 4.48 Schematic visualization of the basic components of a new model for the corneal collagen lattice. At equidistant sites along their circumference, six core proteins of proteoglycans are attached to the hexagonally arranged collagen fibrils. Glycosaminoglycans (GAGs) of the proteoglycans as stained by CB connect their next nearest neighbour collagen fibrils and form a ring-like structure around each collagen fibril. (From Muller et al., 2004, with permission from Elsevier.)

structures contain fucose or mannose, while chondroitin sulphate contains xylose residues.

The interaction between proteoglycans and collagen fibrils has been elegantly demonstrated using cupromeronic blue and $MgCl_2$ at a tightly controlled (3.0 mM) concentration (the 'critical electrolyte concentration'). Both dermatan sulphate proteoglycan and keratan sulphate proteoglycan bind to the collagen arrays at specific binding sites (one proteoglycan to one binding site), suggesting that these sites are essential to the spacing of the fibrils and to the thickness of the interfibrillar space. Keratan sulphate proteoglycans appear to bind to the step regions of the fibrils while dermatan sulphate proteoglycans bind to the gap (Box 4. 9).

Different binding 'maps' occur in different species, such as the mouse. Therefore, several variations on the theme of proteoglycan–collagen interaction appear to be compatible with transparency. However, SLRs appear to be essential. Studies in human tissue developed this model to reveal a regular hexagonal arrangement of six proteoglycans per collagen fibril interacting with the 'next but one' fibril (Fig. 4.48).

Swelling pressure versus hydration. The cornea is about 80% hydrated. This is higher than other tissues such as the sclera, which is about 70% hydrated. Despite this, the corneal stroma 'imbibes' water if it is placed in a solution of saline (this is well demonstrated by the injured, lacerated cornea which swells and becomes opaque); the water-attracting 'hygroscopic' properties of the cornea are the result of its high content of GAGs. The cornea has been described as 'a slice of water stabilized in three dimensions by a meshwork of fibrils and soluble polymers'. The cornea thus has a swelling pressure and a metabolic pump (the endothelium) designed to maintain it. The swelling pressure generates a level of interfibrillar tension and this may be the biophysical mechanism whereby the fibrils are maintained in their normal arrangement. In addition, the swelling pressure itself may reciprocally activate chloride channels and other transporters that maintain a balance of excess ions in the aqueous humour.

The swelling pressure of the cornea is an important clinical concept and attempts have been made to measure swelling pressure using instruments such as the Ocular Response Analyzer (ORA), an instrument similar to an applanation tonometer used for measuring intraocular pressure. The ORA measures corneal hysteresis, which is a function of biomechanical strength of the corneal tissue related to the viscosity of the matrix and its response to deformation. However, whether the ORA truly measures hysteresis independently of ocular pressure is still unclear, despite its considerable relevance for corneal refractive surgical procedures. The swelling pressure, however, affects overall corneal thickness, which indirectly influences measurement of intraocular pressure. Other factors also affect corneal thickness including conditions such as diabetes, thus complicating interpretation of data.

Cellular factors affecting transparency: (1) keratocytes. Corneal fibroblasts (keratocytes) are important in maintaining transparency because they are the source of stromal collagens and proteoglycans. Although most of the changes that occur in the assembly of the matrix are post-translational, the genes and enzymes that regulate these changes are activated in the keratocytes (specific enzyme defects are associated with corneal opacification as in the mucopolysaccharidoses). Keratocyte dysfunction may also underlie the corneal haze seen after corneal refractive surgery.

Collagen turnover in early postnatal life takes about 24–50 hours, but there is little information on adult collagen metabolism. *In vitro* studies on cultured keratocytes have not been very informative because these cells produce a range of collagens and GAGs not found *in vivo*. In contrast, organ cultures of cornea produce a panel of GAGs more akin to that found *in vivo*. Hypoxia and anaerobic glycolysis favour the production of KS-PG over CS/DS-PG (dermatan sulphate proteoglycan). In rabbit cornea, the development of transparency correlates with a dramatic increase in the concentration of KS-PG in the early postnatal period.

Keratocytes also express a range of corneal 'crystallins', so-called because they play a role in light transparency by reducing light scatter and not because of homology to lens crystallins (see later). Corneal crystallins are chaperones and include transketolase and aldehyde dehydrogenase. They may have a role in minimizing corneal scarring (Torricelli et al., 2016).

Cellular factors affecting transparency: (2) leucocytes. The cornea also contains a population of resident stromal and intraepithelial leucocytes which are more frequently found in the peripheral cornea (see Chapters 1, p. 18 and 7, p. 456). These cells include macrophages and dendritic cells, which are thought to be important as immune sentinels in sensing danger (see Chapter 7, page 425), and rare haemopoietic stem cells. These are distinct from limbal corneal stem cells of mesenchymal origin, which reside in epithelial stem cell niches (see later).

What is a stem cell niche? Further information is available online at https://expertconsult.inkling.com

Overall cells comprise about 15% of the stromal volume.

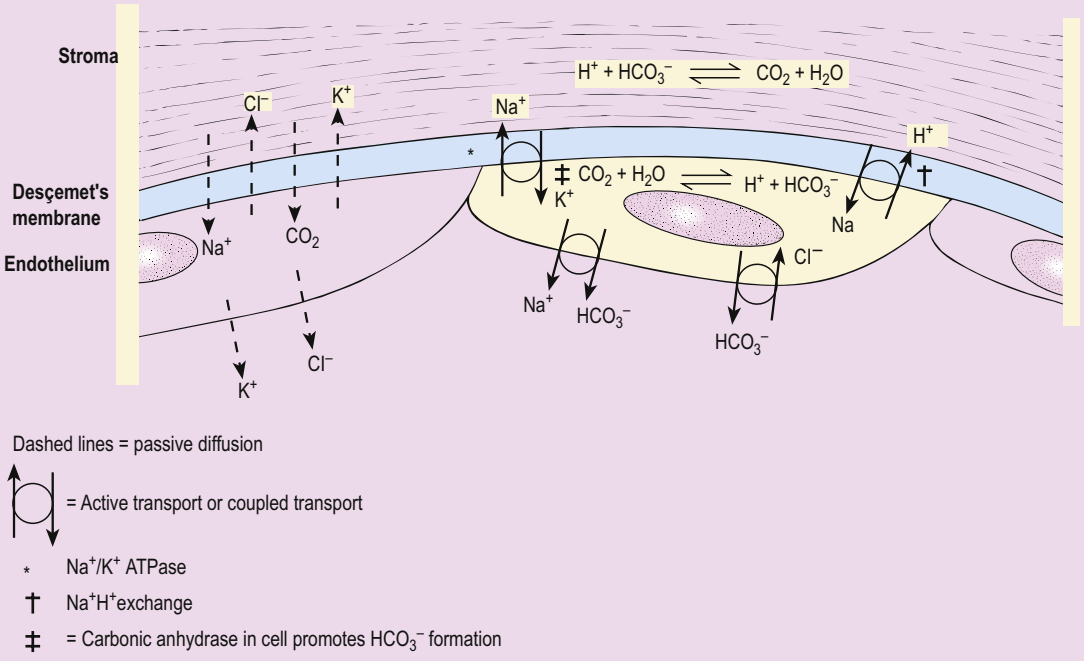
Cellular factors affecting transparency: (3) the endothelium. The endothelial pump determines the level of hydration of the GAGs and thus transparency. Despite the greater than normal level of hydration of the corneal stroma compared with other tissues, its water binding capacity is unsaturated because of the endothelial pump (Box 4.10), which transports water out of the cornea towards the anterior chamber. This is known as the pump–leak mechanism and is an energy-driven ATPase-dependent mechanism (see Chapter 4, Box 4.10, p. 215). Transport of Cl^- ion here depends on the cystic fibrosis transmembrane conductance regulator (CFTR), a Cl^- channel which is expressed in many fluid transporting epithelia in the body.

Other mechanisms in the transport of fluid across the endothelium include a Na^+/H^+ exchanger protein, which drives an electrogenic coupling between Na^+ and HCO_3^- ions. This 'antiport' is essential for maintenance of the intracellular pH by exchanging Na^+ (in) with H^+ (out) in the cell (Box 4.10). This may be one component of the electro-osmosis concept for fluid transport in the cornea. In addition, the corneal endothelial cell membrane contains a water transporter/channel, namely aquaporin 1, which is involved in bulk transport of water molecules (see the section 'The lens') and perhaps also in CO_2 transport. This, however, seems to be mainly a passive transporter, responding to sudden changes in stromal water content and endothelial cell volume rather than having a constitutive physiological role.

Box 4.10 **Corneal endothelial cell pump**

The corneal endothelium transports water out of the stroma by an ATP-driven ion pump mechanism.

Na^+ and HCO_3^- are transported across the endothelium from the stroma to the aqueous, mediated by a Na^+/K^+ -dependent ATPase and a HCO_3^- -dependent ATPase, probably involving carbonic anhydrase. The Na^+/K^+ ATPase is located in the plasma membranes but the HCO_3^- -dependent ATPase is in the mitochondria, where its major role in ion transport may be to generate the ATP required for the Na^+/K^+ ATPase. Carbonic anhydrase may also be involved in the Na^+/H^+ antiport for maintaining intracellular pH.



Fluid transport across the cornea may also be subject to changes in metabolism such as occur with a steep gradient of lactate from the stroma to the anterior chamber in which bulk movement of lactate and water is facilitated by transcellular monocarboxylate co-transporters in the endothelium.

Transportation of water out of the cornea is a dual process. In addition to the transport of water from the stroma to the aqueous humour, there is a net flux of ions and water towards the epithelium and the tears. For instance, an epithelial Cl^- pump is modulated by several receptors including β-adrenergic and serotonergic receptors coupled to the adenylate cyclase and Ca^{2+} second messenger systems and receptors that involve protein kinase C. Dopamine and α-adrenergic receptors may also be involved in Cl^- transport, but importantly many of these receptor systems are coupled to the CFTR Cl^- channel expressed in corneal and conjunctival epithelia.

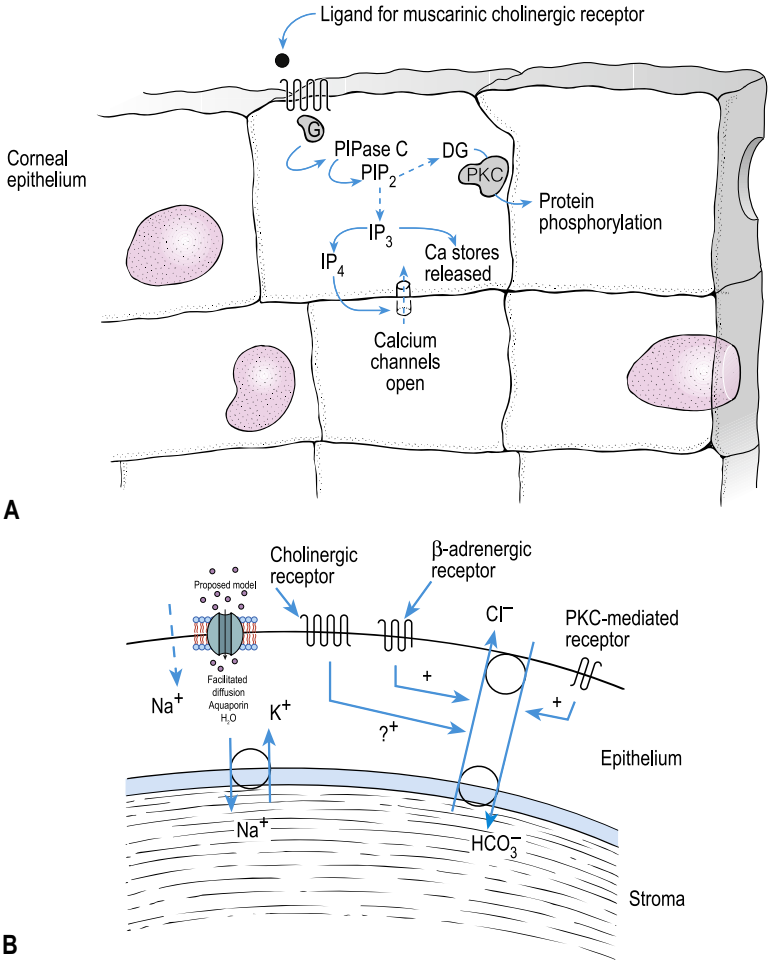
Further information on fluid dynamics at the corneal epithelium is available online at https://expertconsult.inkling.com

Optical factors affecting image formation. The curvature of the cornea plays a major part in refracting and focusing light to produce an image on the retina. Even with a completely clear cornea, the image can be distorted by abnormalities in curvature, producing the various forms of regular and irregular astigmatism. A certain degree of astigmatism is present in many eyes since the corneal curvature is rarely perfectly spherical. These account for some refractive errors common in the healthy population. Refractive errors can also be caused by variations in the overall ocular dimensions of the eye.

A brief introduction to the optics of the eye is available online at https://expertconsult.inkling.com

Corneal epithelial fluid transport

The corneal epithelial barrier is very tight and it does not spontaneously transport fluid into the tears; aquaporin 5 is expressed in the conjunctival and corneal epithelium and is involved in water transport. Ion transport into the cornea is mediated by Na^+/K^+ ATPase and Ca^{2+}/Mg^{2+} ATPase in the basolateral plasma membrane of the epithelium (eFig. 4.21). Na^+ movement from the tears into the epithelium is by passive diffusion down a concentration gradient, but from the epithelium into the stroma active transport is required via the Na^+K^+ ATPase to which Cl^- transport in the opposite direction is coupled, again by CFTR channels. This transport system is sensitive to the eicosanoid metabolite 12(R)HETE (compound C) (see Box 4.8). In addition, this electrolyte flux across the epithelium accounts for the electrical potential difference from (–) outside to (+) inside the cornea of ~25–40 μV.

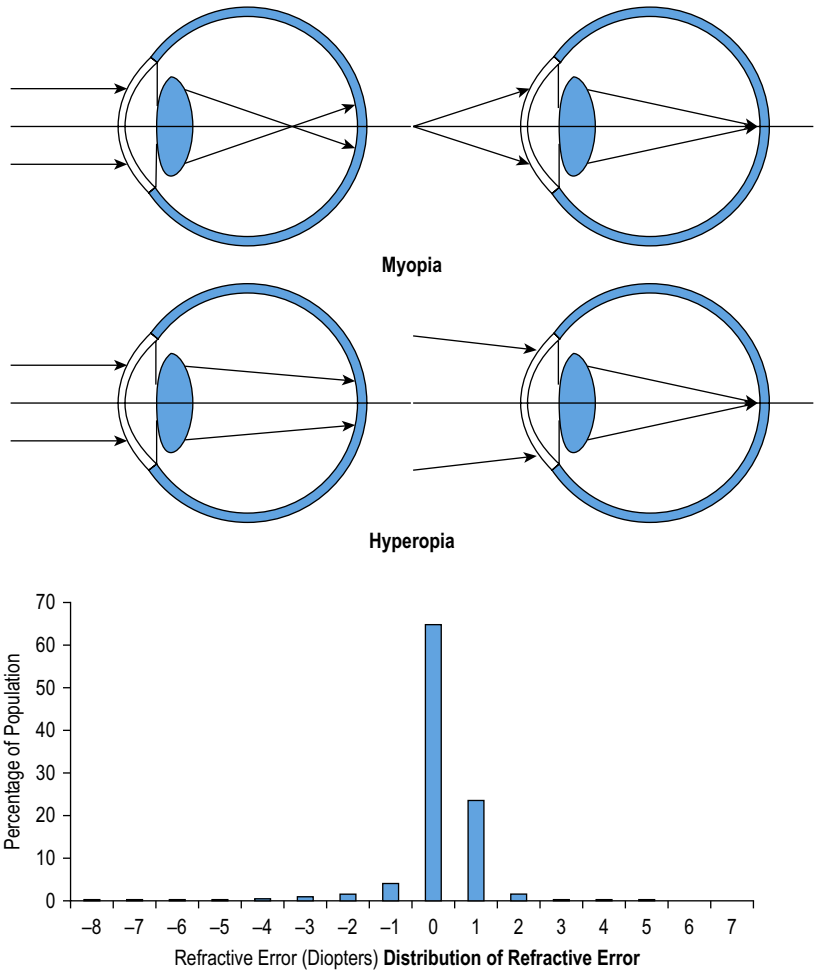


eFig. 4.21 Adrenergic and cholinergic receptors are present in the corneal epithelium and are coupled to a variety of second messenger systems including cAMP, phosphatidyl inositol phosphate 2 (*PIP_2*), and protein kinase C (*PKC*) (**A**). They are involved in the transport of fluid out of the cornea in an anterior direction by enhancing the function of the HCO_3^-/Cl^- pump (**B**). Active pumps for outward extrusion of Cl^- (activated by β-adrenergic and other receptors) and inward transport of Na^+ and HCO_3^- are present. *Dashed line*, passive transport; *circles*, coupled/active transport. There is also a separate water transporter, aquaporin 5, in the corneal epithelium, which is important for the maintenance of normal corneal stromal thickness.

Emmetropia, myopia and hyperopia

Briefly, a normal-sized eye, anteroposterior dimension 22–24 mm, is termed an emmetropic eye and is one in which the image is focused sharply on the retina by the combined refracting properties of the cornea and the lens (eFig. 22). Myopia or short-sightedness occurs in eyes with above average length and the image therefore is focused in front of the retina; conversely, long-sighted eyes have a shorter than normal length and the image is focused behind the retina. Myopia can be corrected by placing a convex lens in front of the cornea, which thus causes some divergence of the light rays and with the correct lens, the image can be formed on the retina correctly. Myopia can also be corrected by 'reshaping' the cornea (see pp. 218-219) to cause some degree of 'flattening' of its curvature and thus reduce its refractive power. Conversely, hyperopia can be corrected using a convex lens and similarly corneal surgery can be performed to correct hyperopia, although it is generally less successful.

Further information on refraction with links to optometry literature is given in Chapter 5, p. 285



eFig. 4.22 Emmetropia is a state of refraction where a point at an infinite distance from the eye is conjugate to the retina. Ametropia is a state where a refractive error is present or when distant points are no longer focused properly to the retina. Myopia or near-sightedness (short-sightedness) is one form of ametropia where the eye is effectively too long or has too high a power. Consequently, a point at infinity focuses in front of the retina. In myopia, a point lying between infinity and the eye is conjugate to the retina. Hyperopia or far-sightedness is a form of ametropia where the eye's power is too weak or the eyeball too short. In this case, a point at infinity focuses behind the retina. A point behind the eye is therefore conjugate to the retina. Astigmatism is a form of ametropia in which refractive error changes with the optical meridian. (From SPIE Field Guide to Visual and Ophthalmic Optics. www.spiedigitallibrary.org, ebooks, eISBN-9780819478191.)

For personal use only. No other uses without permission. Copyright ©2020. Elsevier Inc. All rights reserved.

Severe degrees of astigmatism can be caused by disease of the cornea such as keratoconus (cone-shaped cornea), and the late effects of scarring from wounds such as corneal incisions following cataract surgery, when it can be very difficult to restore preoperative curvature. Although these conditions do not directly affect the transparency of the cornea, they can increase the amount of spherical and chromatic aberration and diffraction, thus degrading the image.

Metabolism of corneal cells

Oxidative metabolism and glucose utilization. The epithelium takes up most of its glucose from the stroma and converts it to glucose-6-phosphate, after which 85% is metabolized via the glycolytic pathway to pyruvate. The bulk of this is then metabolized to lactic acid, but some is diverted into the citric acid cycle to produce ATP as an energy store. The pentose phosphate pathway accounts for the remainder of glucose utilization by the epithelium, producing an important resource for free radical control, namely NADPH (see Fig. 4.29). This is the main mechanism of generation of reducing agents such as glutathione and ascorbic acid. The corneal epithelium has also developed a unique nuclear ferritin-based mechanism to minimize DNA damage from ultraviolet (UV) light-induced free radical damage. Lactoferrin in tears may also assist in this process.

The metabolism of keratocytes is mostly concerned with generating sufficient energy to produce and maintain stromal components (as mentioned previously). In the steady state, these cells are not highly metabolically active and there is a high stromal lactate production, possibly contributing to fluid transport across the endothelium (see previous text).

The endothelium has large energy requirements to sustain its ATPase-dependent pump mechanism and is about five times as active as the epithelium. The major metabolic pathway in the endothelium is aerobic glycolysis, with the citric acid cycle and the pentose phosphate pathway also playing a significant role.

Oxygen handling by the cornea. The epithelium obtains its oxygen from the preocular tear film at a rate of 3.5–4.0 μL per cm^2 per hour. The endothelium, however, and the keratocytes in the deep stroma receive their oxygen supply from the circulation via the aqueous humour. Corneal function and health are therefore dependent on local conditions at the surface of the eye and on systemic factors such as cardiopulmonary capacity.

Oxygen is consumed in the citric acid cycle, generating 36–38 molecules of ATP per molecule of glucose. The utilization of the citric acid cycle versus the glycolytic pathway is determined by the energy demands of the tissue, specifically the need for ATP. Thus, the endothelium makes greater use of the citric acid cycle than the epithelium. In addition, oxygen consumption by the cornea increases almost twofold when acidosis prevails, as occurs in contact lens wear. This is in part the result of the activation of pH regulatory mechanism, including Na^+/H^+ exchange, which then stimulates Na^+/K^+ ATPase activity.

Excess oxygen can be detrimental to the organism if it is converted to the superoxide radical and then to hydrogen peroxide (see Box 4.6 and Fig. 4.33). In both the epithelium and the endothelium, redox systems involving glutathione and its two enzymes (glutathione reductase and glutathione peroxidase) depend on the generation of NADPH and thus on a supply of glucose. When the intracellular levels of glutathione are reduced in the cornea by one-third, the clarity of the cornea and its ability to pump fluid decline dramatically. Remarkably, the oxidative stress induced by UV light or alkali damage can be prevented by solutions of molecular hydrogen applied topically (Cejka, 2017).

The aqueous humour contains high levels of H_2O_2, perhaps by virtue of the reduction of oxygen in the aqueous humour via ascorbate usage. Free radical damage, oxidative stress and ER stress to the corneal endothelium may all combine to induce apoptosis and thus may account for the progressive endothelial cell loss associated with age. Small GTPases (see p. 165, Fig 4.3 Ras, Rac and Rho small GTPases in integrin signalling) have been implicated in various ocular diseases including corneal endothelial cell death and Rho kinase inhibitors are under evaluation as novel therapies.

Free radical damage to the cornea can also be induced by therapeutic interventions such as that incurred during phacoemulsification for cataract and by UV radiation used to perform corneal collagen cross-linking during refractive surgery (see p. 220).

Several receptors for neurotransmitters and other agonists are present in the corneal epithelium involved in nutrient handling and regulation. In addition, since the cornea is richly innervated, they contribute to a healthy nutritive neurogenic environment, preventing neuropathic keratopathy.

Effects of contact lens wear on corneal physiology. Contact lenses are common optical devices that correct the refractive errors of the eye to achieve emmetropia. The tear film bathes both surfaces of the contact lens (Fig. 4.49), but corneal epithelial function is still relatively compromised. The corneal epithelium layer receives its oxygen from the tears and its glucose from the circulation via the aqueous humour and the limbal vessels (see earlier). Contact lenses reduce the direct availability of oxygen to the epithelium, thus shifting the balance from aerobic to anaerobic metabolism. The already high lactate levels in the cornea are doubled with contact lens wear and carbon dioxide production is increased. The induced acidosis has a direct effect on stromal hydration by impairing deturgescence mechanisms (see earlier).

Hard (rigid) contact lenses are usually made from polymethylmethacrylate (PMMA) and have the greatest effect on corneal function; in addition to restricting oxygen

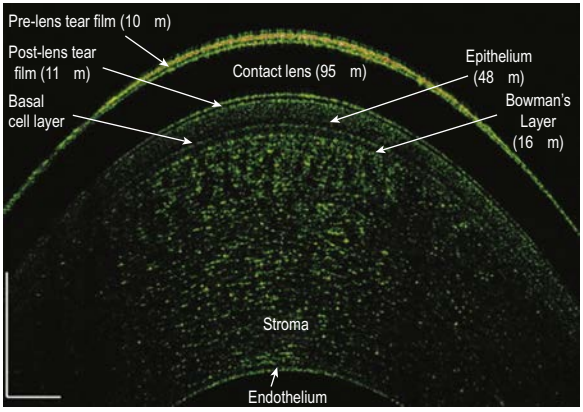


Fig. 4.49 Ultra-high resolution optical coherence tomography image of the central cornea with a PureVision (Bausch & Lomb, Rochester, NY) lens after instillation of artificial tears. The central cornea was imaged with 6 mm scan on the horizontal meridian. The image was taken immediately after lens insertion and instillation of one drop of artificial tears. The epithelium, including the basal cell layer, and Bowman's layer are evident in addition to the pre-lens and post-lens tear films. Total corneal thickness was measured at 526 μm. Bars = 250 μm. (Reprinted with permission from Wang J, Jiao S, Ruggeri M, Shousha MA, Chen Q. In situ visualization of tears on contact lens using ultra-high resolution optical coherence tomography. *Eye & Contact Lens* 2009;2:44–49. (Wang, 2011))

availability, hard lenses deplete glycogen stores, even though the level of glucose availability is not reduced. Inhibition of hexokinase reducing glucose utilization by the cornea prevents prolonged wear of hard contact lenses, owing to the damaging effect on corneal transparency induced by the disturbed metabolism. Soft contact lenses are made from polymers of hydroxyethyl methacrylate (HEMA), poly(HEMA/vinylpyrrolidones), silicone or other similar materials and permit extended wear of the lens owing to their permeability to oxygen and carbon dioxide. However, there is still some degree of lactate accumulation with soft lenses and prolonged use disturbs the endothelium. HEMA-based lenses are hydrophilic, while silicone-based lenses are hydrophobic: accordingly, there is less protein deposition on the latter but greater levels of denatured proteins. However, corneal inflammatory episodes are if anything more frequent with silicone lenses. Manufacturers of contact lenses continually produce new 'biomimetic'-type lenses with increasing water content (up to 59%) in attempts to support normal corneal physiology (hydrogel lenses). In addition, incorporating other material into contact lenses such as polyethylene glycol and cross-linked hyaluronan are other possibilities under consideration to improve biocompatibility.

A popular compromise in contact lens type is the gas-permeable rigid lens, which combines the reduced toxicity of PMMA with high gas-transfer capability. The wide variety of lens types and materials has led to their being characterized on the basis of their oxygen flux, defined as the *DK* value:

$$\text{Oxygen flux} = \text{DK}/\text{L} \times \Delta\text{P}$$

where D is the diffusion coefficient, K is the solubility and L is the thickness of the lens material. ΔP is the change in the partial pressure of oxygen across the material. HEMA and PMMA have a low oxygen flux, while hydrogels and silicones have a high flux. Both the thickness of the lens and the *DK* value determine its suitability for use in terms of its gas permeability. The actual amount of oxygen that reaches the cornea is the most important factor in the design of a contact lens and most practitioners describe contact lenses in terms of their equivalent oxygen performance (EOP).

Contact lenses may have deleterious effects on the epithelium, causing thinning, reduction in the hemidesmosome density and the number of anchoring fibrils and reduced adhesion of the epithelium to the basement membrane. This may be a direct effect of low O_2-transmitting lenses on basal epithelial cell proliferation. This is especially true of extended-wear hydrogel lenses. In severe cases, excessive use of contact lenses produces epithelial oedema and keratopathy in the form of punctate epithelial erosions. Rigid contact lenses also produce tear film instability by causing damage to the epithelium and the mucin layer in particular. Contact lens wear may also induce changes in the corneal stroma (thickening) and the endothelium (polymegathism). Since contact lenses are foreign material they are intrinsically pro-inflammatory (Efron, 2017).

Despite these caveats, innovations in contact lens manufacture and wear are constantly being suggested. These include the incorporation of UV filters to protect limbal stem cells (Notara et al., 2018), biosensors for detection of diseases including cancer (Tseng et al., 2018), and the corneal refractive therapy (orthokeratology) where CLs are worn overnight in an attempt to alter the refraction of the eye.

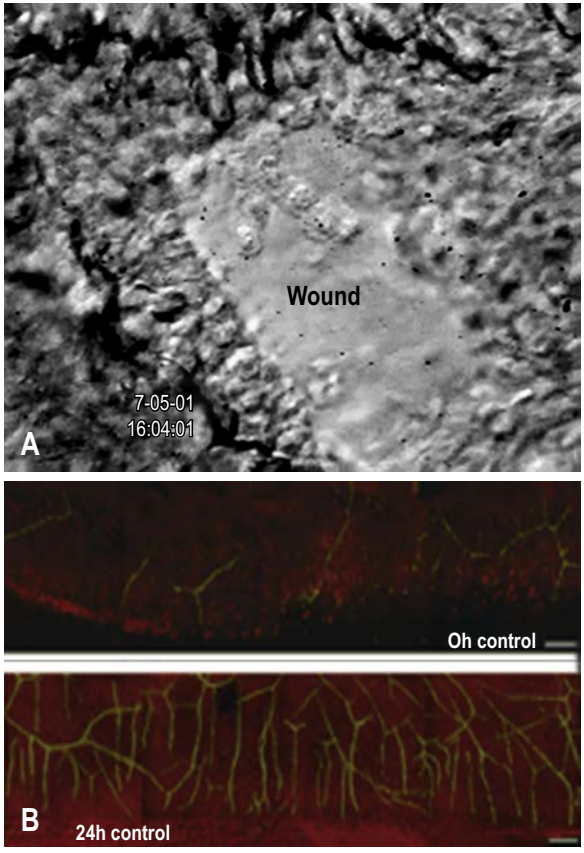
Cell turnover and wound healing in the cornea

The epithelium. The epithelium is constantly being regenerated by mitotic activity in the basal layer of cells. However, after epithelial abrasion, the initial response of the basal epithelium at the edge of the defect is to migrate (slide) as a flattened sheet of single-cell thickness across the stroma. These cells adhere by forming hemidesmosomes and intercellular contacts and more superficially placed cells can be seen to tumble over the adherent cells into the available space, leading to rapid filling of the epithelial defect.

Further information on epithelial healing in corneal abrasion is available online at https://expertconsult.inkling.com

Corneal epithelial wound healing

This has been visualized directly in cultures of corneal epithelium (see Video 4.1) in which superficial epithelial cells appear to rush to fill the defect ahead of the basal more slowly centrally migrating epithelial cells. In addition, the differentiation of the corneal epithelium into a multilayered structure after wounding requires neural input from the corneal nerve endings, which orient themselves perpendicular to the wound edge after the initial single layer of cells has closed the defect and are probably required to promote full differentiation of the six to ten layers of cells (eFig. 4.23). Thus, in conditions of neuropathy or when the nerves fail to extend their axons, as after herpes infections of the cornea, full epithelial healing frequently fails and recurrent breakdown of the defect occurs.



eFig. 4.23 (**A**) Time-lapse video sequence of corneal epithelial wound healing showing 'tumbling' of superficial cells to fill the defect; (**B**) orientation of regenerating nerve endings towards the wound edge (dark area in lower section) as the wound closes. (From Song et al., 2004.)

However, *in vivo* studies using 'confetti' mice indicate that proliferation of the basal cells may be the major mechanism whereby the corneal defect is filled (Park, et al., 2019). This, combined with reorientation of nerve endings into the wound edge, leads to differentiation of the newly divided cells into the five- to seven-cell layered mature epithelium. Closure of corneal epithelial defects occurs in a predictable manner, which can be visualized clinically using fluorescein staining of the cornea; here, the initial wound edges, in geometric patterns, are filled in and meet in the centre of the cornea.

Migration of epithelial cells is achieved by cytoskeletal and cell-shape changes involving redistribution of actin–myosin fibrils and changes in actin-binding proteins (e.g. the cytoskeletal scaffold protein, fodrin) and cell adhesion molecules (e.g. E-cadherin), under genetic regulation via growth factors. Migration of the cells also depends on outside-in intracellular signalling via matrix components such as fibronectin/fibrin, laminin and collagen peptides through cell surface integrins (see pp. 165). The role of fibronectin/fibrin in corneal epithelial resurfacing may be to facilitate healing where the normal basement membrane and, in particular, its laminin component have been lost.

Adhesion of epithelium to the basement membrane and Bowman's layer is normally mediated via hemidesmosomes, the lamina densa and the anchoring type VII collagen fibrils (see Fig. 4.26). However, while hemidesmosomes form during the early stages of re-epithelialization (18 hours), many days elapse before anchoring fibrils reappear and many months pass before full ultrastructural integrity is restored. This may explain in part the phenomenon of recurrent erosion where there has been damage to Bowman's/superficial stromal layers of the cornea. Proteolytic activity in repairing epithelial defects is also important—both urokinase-type plasminogen activator and matrix metalloproteinases have been implicated.

Limbal corneal stem cells. The epithelium is constantly being renewed (in mice, the entire corneal epithelium is renewed in 7 days) under the control of stem cells located in an epithelial 'stem cell niche' at the limbus in the palisades of Vogt (Fig. 4.50). While some renewal of epithelial cells takes place from the basal layer to the surface throughout the cornea, the main source of cell renewal is limbal stem cells located in the upper and lower cornea. Patients who have severe damage to the limbal region of the cornea, for instance with chemical burns, suffer from stem cell deficiency and are prone to conjunctivalization of their cornea in which the conjunctival epithelium migrates onto the cornea to fill the defect, bringing with it stromal tissue and blood vessels. The outcome is corneal opacification and blindness. Attempts to promote corneal epithelial wound healing using growth-promoting agents such as epidermal growth factor and retinoic acid have not met with great success. Limbal stem cell grafts have been used to treat ocular surface disease with limited success. The major difficulty is in isolating a sufficiently pure population of viable limbal stem cells that do not have a specific cell marker. A current panel of surface markers used includes Wnt7A, CDH3, CK14/15, ABCB5 and CD200 (Bojic et al., 2019). Despite difficulties, a multicentre clinical trial has been registered for recruitment (Allogeneic ABCB5-positive Limbal Stem Cells for Treatment of LSCD @clinicaltrials.gov).

The stroma. Incisional wounds of the cornea that involve the stroma may be accidental or intentional. The immediate effect is to cause wound gape and imbibition of water from the tears by corneal GAGs (see p. 192). This causes localized opacification (light scatter) and initiates a series of events in the cornea directed at closing the wound. These include deposition of fibrin within the wound, rapid epithelialization of the wound incision and activation of the keratocytes to divide and synthesize collagen and GAGs. During the early phase of corneal wound healing, epithelial cells invading the stroma undergo epithelial-to-mesenchyme transformation (EMT) with high expression of α–smooth muscle protein, and lay down collagen and GAGs found in any typical wound: e.g. hyaluronan acid, type I and type III collagen and matrix glycoproteins. The size and arrangement of the fibrils are not regular, further contributing to the corneal opacity. In extensive wounds, this opacification remains permanently; however, in smaller, well-defined wounds there is an attempt by the cornea to restore clarity by producing normal corneal matrix components.

Surgical wounds to the cornea aim to minimize the risk not only of causing wound-related corneal opacities but also of inducing shape change to the cornea and thus astigmatism. Accordingly, cataract surgeons are using increasingly small wounds to remove the lens, the most recent innovation being OCT-guided femtosecond laser surgery, in this case where the laser is used to incise the lens capsule. Since the corneal curvature is a function of tension in its circumferential fibres (see Fig. 4.46), restoration of the normal curvature will not be achieved unless the edges of the wound are precisely apposed by surgical reconstruction. Indeed, this is the basis of refractive surgery which attempts to reshape the curvature of the cornea. Early practice of partial thickness radial keratotomy was performed by making 'relaxing' incisions in the peripheral cornea to release the circumferential tension and thus flatten the corneal curvature. The wounds are intentionally left to heal in a gaping configuration. Various types of laser surgery, such as argon-F1 (excimer = excited dimer), ultraviolet laser energy and, more recently, femtosecond laser, are used to produce precise customized ablation of the anterior stroma performed directly on the exposed tissue after lifting a central, hinged flap of stroma. Ablation is thought to be caused by photon–photon interactions derived from thermal reactions or directly by photoablation, whereby covalent cross-links in the collagen fibrils are disrupted. This is termed laser *in situ*

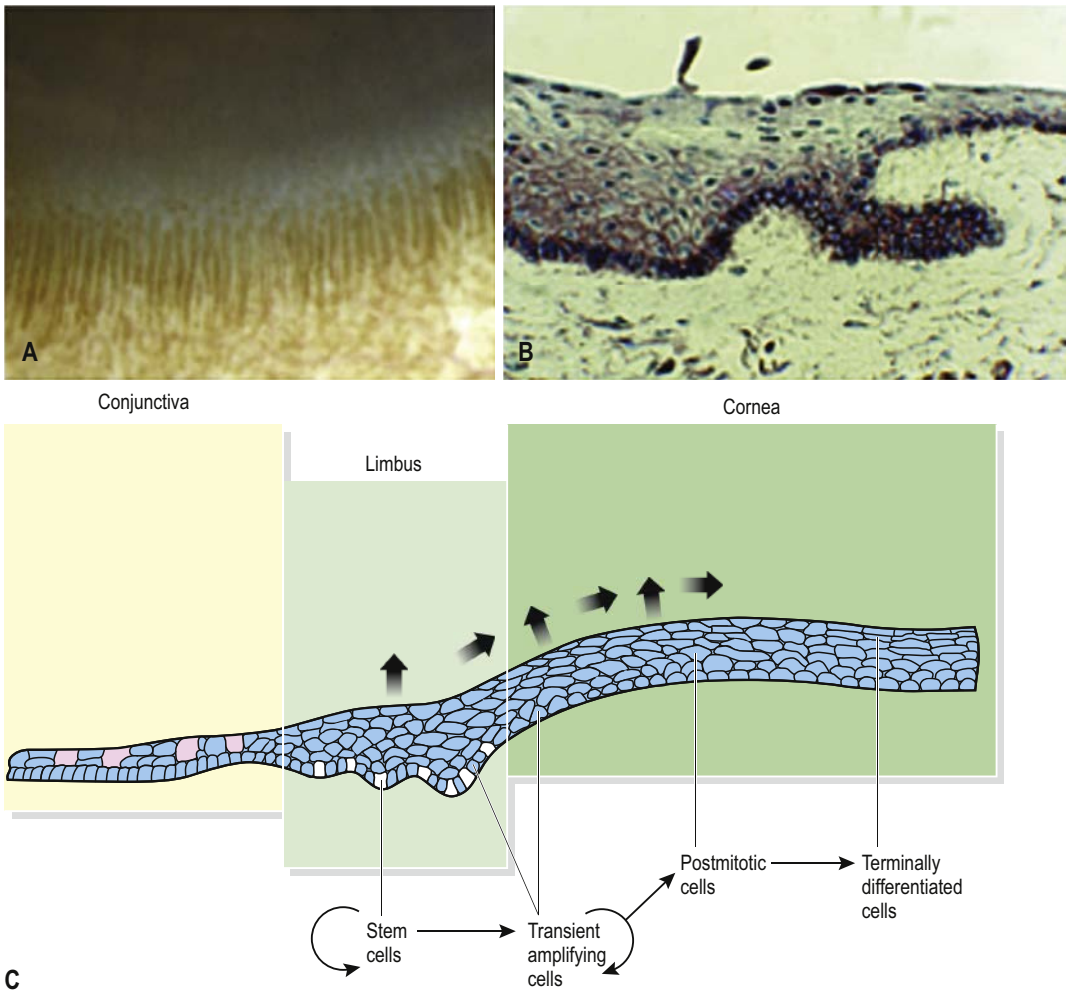


Fig. 4.50 The corneal stem cell niche: a limbal location for corneal stem cells and transient amplifying cells. Stem cells (*white*) are exclusively located in the basal limbal epithelium at the bottom of the epithelial papillae forming the palisades of Vogt. Transient amplifying cells occur in the basal epithelia of the limbus and peripheral cornea. Postmitotic and terminally differentiated cells make up the suprabasal and superficial layers. (From Schlötzer-Schrehardt and Kruse, 2005, with permission from Elsevier.)

keratomileusis (LASIK) in which the surface of the cornea is reshaped (usually flattened) by raising a corneal flap, laser treating the exposed stromal bed and restoring the corneal flap without sutures. Both these and conventional surgical corneal incisions are fully epithelialized in the normal manner, with epithelial migration into the depths of the wound sometimes producing excessive layers of cells.

Refractive surgery continues to expand both in quantity and in novelty of technique, combined with greater precision using methods such as eye-tracking to control the incision. Wavefront-guided surgery combined with anterior segment OCT imaging are further innovations in this ever-expanding field. Wavefront technology is a term used to describe the fundamental quality of an optical system, for instance in the field of astronomy, in terms of wavefronts of higher and lower order, and the eye lends itself well to such analysis. For instance, images formed by the healthy eye may be blurred for three main reasons: light scatter, diffraction and optical aberrations, e.g. chromatic or spherical. Using wavefront technology, optical aberrations from monochromatic sources such as myopia are selectively addressed and treated with wavefront-based LASIK.

Refractive corneal surgery, including photorefractive keratometry (PRK), LASIK and full-thickness penetrating corneal graft, has complications including sometimes severe residual astigmatism and even progressive ectasia (widening) of the

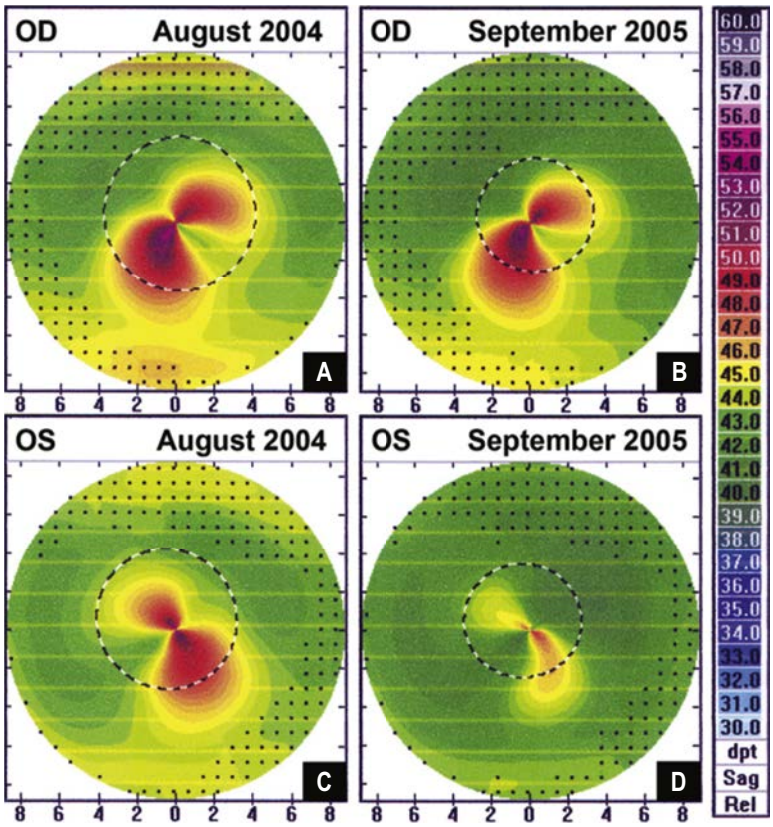


Fig. 4.51 (**A-D**) Corneal topographic maps showing the stabilizing effect of corneal cross-linking on corneal curvature (compare lower right and left panels). (From Hafezi et al., 2007.)

initial wound. Alternative approaches have been tried to restore normal corneal curvature; these include topography-guided laser surgery using keratometry, sometimes including restoration of collagen cross-linking induced by riboflavin and UV light. Collagen cross-linking helps to 'stiffen' the stromal bed (Fig. 4.51) and has been used to treat conditions such as keratoconus, corneal ectasia and some forms of keratitis. The standard (Dresden) protocol involves removing the corneal epithelium while retaining the corneal epithelium during the procedure has less effect. Long-term outcomes for this procedure are unknown.

The endothelium. The corneal endothelium does not normally undergo mitosis in humans even after direct injury as in a perforating corneal wound. With age there is a decline in the number of endothelial cells with an increase in their size and variable morphology (polymegathism). The response to direct wounding is to undergo enlargement and 'cell slide', as occurs in the epithelium in the early stages of migration. If sufficient numbers of endothelial cells are lost, the cell layer cannot perform its pumping action and the cornea imbibes water (decompensates) and becomes opaque. Inhibition of the small GTPase Rho kinase may promote corneal endothelial repair.

Vascularization. The normal cornea is avascular, although a few vessels may be found at the limbus. Blood vessels from the conjunctiva or the deep episcleral plexus may invade the periphery of the cornea beyond the limbus during healing of wounds or infected ulcers. When corneal epithelial or stromal defects fail to close promptly, often as a result of infection or during the severe inflammatory response of chemical injury, the continued release of proteolytic enzymes from inflammatory cells and damaged corneal cells causes degradation of the stroma and increases the risk of spontaneous perforation. Matrix metalloproteinases such as matrilysin (MMP7), stromelysin (MMP3) and MMP-9 as well as plasminogen activators (uPA and tPA) and pro-inflammatory cytokines, such as interleukin 1 (IL-1), IL-6 and IL-8, tumour necrosis factor-α and TGF-β, macrophage inflammatory proteins (MIP) 1α and β (CCL3 and 4), and granulocyte–macrophage colony-stimulating factor (GMCSF/CSF-2), stimulate further ingress of inflammatory cells and initiate a vascularization

response (see Chapter 7, page 402). Growth factors such as VEGF, FGF and HGF (hepatocyte growth factor) are also released, of which VEGF is considered the dominant angiogenic facto. Vessels advance across the cornea to the site of injury or infection and contribute to the eventual pearly white, opaque leucoma of the healed cornea.

Inflammation also induces new lymphatic vessels from the limbal regions, from lymph vessel precursors in the conjunctiva, especially after herpes simplex virus infections of the cornea, which participate in the overall immune response. Lymphatic vessels are induced by the VEGF isoform, VEGF-C. Inhibitors of angiogenesis are co-released during the process of inflammation and include endostatin, a degradation product of collagen XVIII, in addition to other products of collagen, fibronectin and even KI5 (Kringle domain 1–5), which is a fragment of plasminogen itself. In addition, thrombospondins (1 and 2) in synergy with a scavenger receptor CD36 are important anti-angiogenic factors present in the normal corneal stroma. There are many more players in this process in which the essential concept is one of inflammation causing damage to tissues, which releases tissue products, which in turn loop back to regulate and limit the inflammation and prevent further damage. Similar processes occur during embryogenesis and account for the lack of corneal vascularity and the development of a normal anterior segment. Development overall is under the control of the transcription factor FoxC1, mutations which underpin certain types of glaucoma and other ocular anomalies such as Axenfeld-Rieger's syndrome and Peters' anomaly (see Chapter 9, p. 495, Box 9.2).

Vitamin A and the cornea. Deficiency of vitamin A leads to impaired corneal and conjunctival epithelial function, with loss of corneal lustre, Bitot's spots, punctate erosions and xerophthalmia, partly as a result of loss of goblet cells in the conjunctiva. Vitamin A (retinol) is required for control of epithelial keratin expression and the synthesis of cell surface glycoproteins involved in the glycocalyx. Deficiency of vitamin A leads to a form of keratinization of the corneal epithelium.

Vitamin A is also essential for normal corneal wound healing. A simple abrasion or ulcer that would be dealt with rapidly by a healthy cornea is likely to be complicated by a stromal melting response (keratomalacia) in vitamin A-deficient humans and animals. Experimentally it has been shown that topical retinoic acid can reverse the effects of vitamin A deficiency. In addition, retinol promotes the synthesis of the α-1 proteinase inhibitor (alph-1 antitrypsin, AAT), which inhibits a wide range of proteolytic enzymes. Vitamin A has also been shown to have protective antioxidant effects on corneal endothelial cells in culture while retinoic acid is important in promotion of T regulatory cells for control of immune, especially autoimmune, responses.

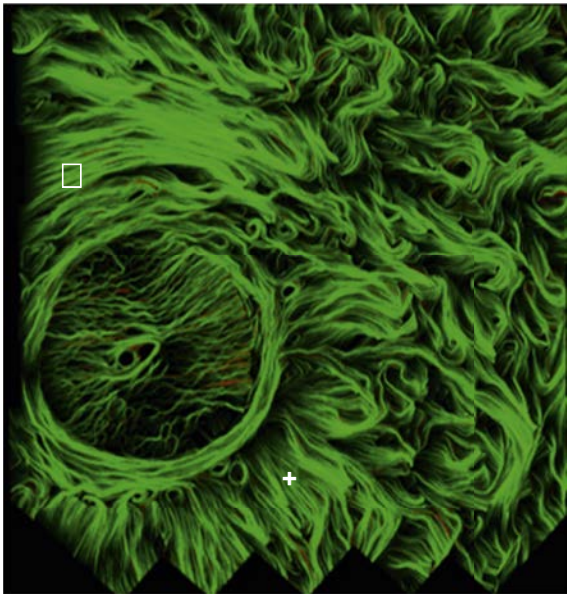


Fig. 4.52 Collagen fibre traces in the optic nerve head of a human eye. The fibre traces help to emphasize the fibre continuity and organization. *(+)* Radial, *(~)* circumferential regions and the less aligned interweaving regions on the right hand side are shown, as well as the fibrous lamina cribrosa, within the canal. (From Gogola, 2018.)

Vitamin A can be converted from traditional β-carotenoids taken in developing countries but is less efficiently converted to vitamin A than from carotenoids in other material such as lycopene, lutein and zeaxanthin.

The sclera

The sclera is non-transparent and tough because of its acellularity and variably sized matrix components.

Matrix factors

The sclera is essentially acellular, containing only a few fibroblasts and non-branching traversing vessels. Recent studies have shown that there are some contractile fibroblasts (myofibroblasts) in the sclera and choroid that may have a role in refractive properties of the eye. It is opaque for the opposite reasons that the cornea is transparent, i.e. that the type I and III collagen fibres are of variable diameter and their distribution is irregular. There are also several other minor collagens in sclera (types V, VI, VIII, XII, XIII). There is, however, some degree of organization of scleral collagen fibres. Studies using polarized light microscopy (PLM) have shown that there are three fibre arrangements in the sclera: an innermost radial and outermost circumferential

and an inter-weaving set of fibres (Fig. 4.52). It has been suggested that the long-running interwoven fibres inserting tangentially into the scleral canal provide a biomechanical advantage to cope with increases in intraocular pressure.

The proteoglycans in sclera are predominantly proteodermatan and proteochondroitin sulphate of the SLR type (see p. 222) and they are localized to the collagen fibres in a similar manner to corneal proteoglycans. However, there are no proteokeratan sulphates. Other proteoglycans present in sclera include aggrecan, PRELP (proline-arginine-rich and leucine-rich repeat), decorin and biglycan among others. In addition, the amount of proteoglycan in the sclera is considerably less than that in the cornea, with the effect that it is much less hydrated (70%). The sclera, unlike the cornea, also contains elastic fibres around a fibrillin core, accounting for about 2% of total fibril content in the adult.

Certain large aggregating proteoglycans, such as versican, neurocan and brevican, combined with hyaluronan are present in the sclera. There is a considerable turnover of extracellular matrix constituents in the sclera and this may determine the shape and size of the eye and thus refraction itself.

Increases in the anteroposterior dimension of the globe occur with age and the adult globe size is reached around the age of 7–10 years. Continuing increases in the antero-posterior dimensions of the globe underlie the development of myopia (short-sightedness) but the cause is unknown. In pathological myopia, a staphyloma may develop (localize out-pouching of the sclera at the posterior pole) where the sclera becomes very thin. The choroid may also be thin or absent in severe cases. The choroid may serve as a paracrine tissue, regulating collagen and proteoglycan synthesis, particularly hyaluronic acid, in the sclera and consequently its overall dimensions.

Bulk fluid transport and the uveal effusion syndrome

Although most of the bulk transport of fluid out of the eye takes place through the anterior chamber drainage angle and/or the uveoscleral meshwork (see Chapter 1, p. 26), there is appreciable transretinal transport of fluid towards the choroid. Some of this is drained via the normal choroidal vessels (10% of the intraocular fluid is drains into the vortex veins), but a proportion is drained directly transsclerally. The effect of this trans-scleral flow is to 'suction on' the retina to its adjoining RPE layer and maintain retinal apposition.

Fluid flowing across the sclera is absorbed by the matrix proteoglycans. Thus, the sclera is maintained in its normal state by having proteoglycans with a low water-binding capacity. In some conditions, such as the rare uveal effusion syndrome, and in nanophthalmia, the sclera contains high levels of abnormal proteoglycans, especially dermatan sulphate-containing proteoglycans, which bind and trap large volumes of water. Thus, the sclera thickens and may secondarily obstruct the choroidal venous drainage, causing further swelling and water retention.

The uveal tract

The uveal tract, comprising the iris, the ciliary body and the choroid, is a continuous layer of which the major functions are to regulate the pupil size for optimal visual function and to act as the lymphovascular tissue of the eye. Each component, however, has several other functions.

The iris

Physiology

The iris is derived from neuroectodermal and mesodermal tissue and is designed to function as the lens aperture of the eye. This is achieved by the opposing actions of the sphincter pupillae and the dilator pupillae muscles (see Chapter 1, p. 26). The sphincter is an annular band of true smooth muscle that inserts close to the pupil margin at the pigment epithelium. The dilator is a highly unique series of myoepithelial cells representing the continuation of the outer layer of ciliary body pigmented epithelial cells. Some consider these cells to be myofibroblasts, which do not have a full complement of muscle proteins including desmin. The non-pigmented ciliary body epithelial cells are continuous with the posterior pigmented iris epithelium (see Chapter 1, p. 30). Between the origin at the ciliary body and the insertion near the sphincter pupillae, the dilator has several side insertions into the stroma of the iris which allow it to mobilize the iris during dilation.

The functions of the pupil are:

- to regulate the amount of light entering the eye (it increases 16-fold on dilation of the pupil from 2 to 8 mm)
- to increase the depth of focus for near vision
- to minimize optical aberrations.

These functions are mediated by light and near reflexes, whose neural pathways involve autonomic parasympathetic (constriction) and the sympathetic (dilation) mechanisms (see Chapter 1, p. 32 and Chapter 5, p. 292, Box 5.5). Unusually large pupil diameters have been linked to the development of myopia through increased optical aberrations.

The neuromuscular junctions of the iris are susceptible to direct pharmacological manipulation by agents that induce miosis (cholinergic agents, sympathetic antagonists) and mydriasis (anticholinergic agents, sympathomimetic agents) (see Chapter 6, pp. 365-366). In addition, pharmacological agents that induce the release of neurotransmitters from synaptic terminals (e.g. substance P is released after nitrogen mustard exposure) can have a marked effect on pupil responses. There is some evidence

that dual innervation exists for iris muscles with excitatory and inhibitory input to each; thus, the action of any individual drug may not be entirely predictable, depending on the state of activity at the time of administration of the drug.

Sympathetic activity in the dilator muscles is mediated by α-adrenergic receptors (the effects can be blocked by phenoxybenzene), but some β activity also exists. In some species, such as the cat, the action of sympathetic agents on the sphincter muscle appears to be mediated by β receptors only; in the monkey the action appears to be strongly α-mediated in the sphincter and dilator, thus producing antagonism. Studies on isolated human iris dilator muscle indicate that the excitatory innervation is α-adrenergic while the inhibitor is cholinergically mediated.

Pigmentation of the iris contributes to iris colour, as does backscatter of short-wavelength light reflected into the iris stroma; thus, iris thickness is a significant determinant of iris colour.

Blood flow in the iris

About 5% of the total ocular blood flows through the iris. Iris blood vessels, derived from the major vessel circle, are contained as radial coils within tube-like formations of the stromal tissue, an arrangement that allows them to remain patent when the iris is fully dilated. Iris vessels have tight junctions and lack fenestrations; this renders them relatively impermeable to large molecules, as demonstrated by anterior segment fluorescein angiography. They constitute a second component of the blood–aqueous barrier (see later). Imaging techniques using high-resolution ultrasound allow visualization of iris and ciliary body vessels and indicate that flow velocities as low as 0.6 mm/second can be detected (Fig. 4.53).

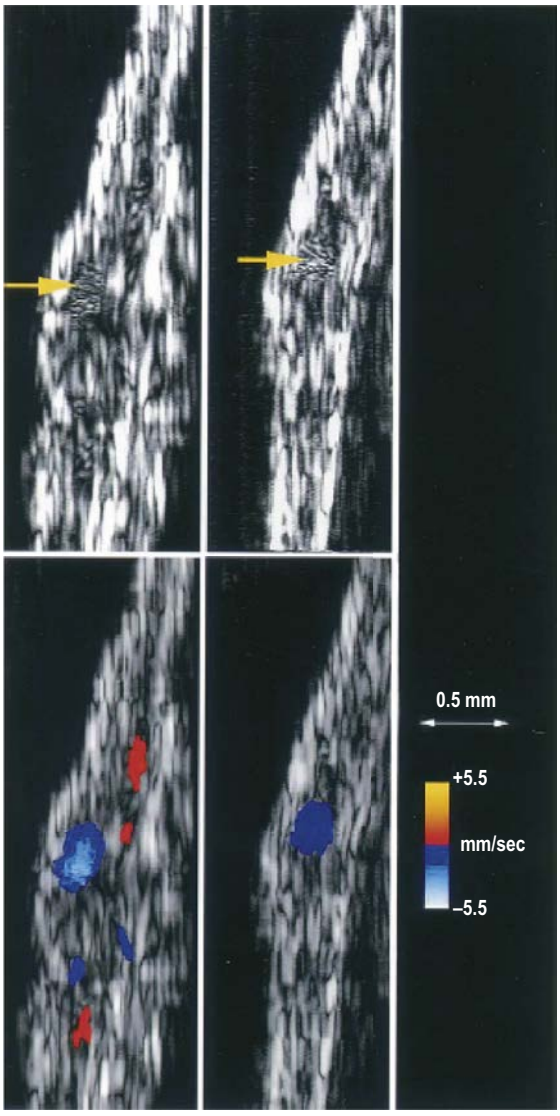


Fig. 4.53 *Top panels*: swept-mode images of iris and ciliary processes at 37°C *(left)* and 4°C *(right)*. *Arrows* indicate position of major arterial circle. *Bottom panels*: colour flow images derived from swept-mode data shown in top panel. (From Silverman et al., 2002, with permission from the World Federation for Ultrasound in Medicine and Biology.)

The ciliary body

Functions of the ciliary body

The ciliary body has multiple functions:

- It provides the blood and nerve supply to the anterior segment.
- It maintains intraocular pressure by secretion of aqueous.
- It constitutes the major portion of the blood–aqueous barrier.
- Its musculature underpins the process of accommodation.

Blood flow in the ciliary body

The ciliary body receives blood vessels from the long posterior ciliary arteries and the major iris circle (see Chapter 1, p. 29 and Fig. 1.21). Blood flow through the ciliary body is about 7% of total ocular flow. The vessels are highly fenestrated, leaking most of their plasma components into the stroma. Blood flow in the iris and ciliary body is autoregulated like that of the retina (i.e. it does not alter significantly with changes in perfusion pressure), but it is also under autonomic control and can be modified by a variety of adrenergic and muscarinic inputs. Under normal circumstances, aqueous production is independent of ciliary body blood flow until the latter declines to <75% of normal.

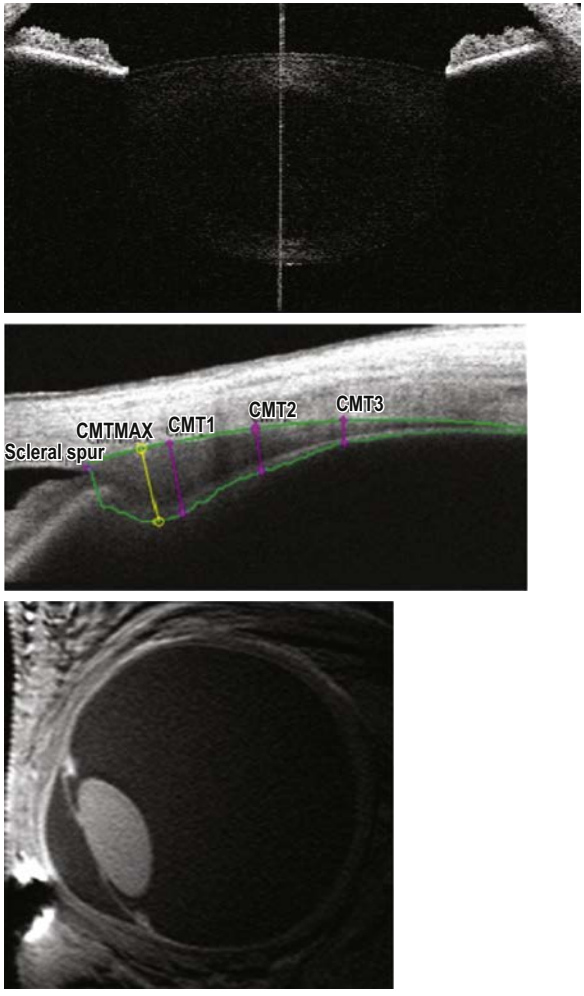


Fig. 4.54 Optical coherence tomography images of the crystalline lens *(top)* and ciliary muscle *(middle)*, and MR imaging of eye (*bottom*). Ciliary muscle image analysis shows cross-sectional ciliary muscle thickness *(CMT)* at 1, 2, and 3 mm posterior to the scleral spur as well as maximum thickness. (From Richdale et al., 2013.)

The ciliary muscle and accommodation

In humans, relaxation of the zonule is induced by contraction of the ciliary muscle, which moves forwards, thereby allowing the lens to adopt a more spherical shape owing to the elasticity of the lens capsule (Fig. 4.54). This is the process of accommodation, which produces greater ability to focus on near objects owing to the increased refractive (dioptric) power of the lens. This has been quantified in a group of young males using a Shack-Hartmann wavefront aberrometer (Ke, 2017). For 2 dioptres of accommodation there is a significant increase in lens thickness as well as anterior ciliary muscle thickness (CMT1, Fig. 4.54) and reduction in anterior chamber depth with CMT3 (Fig. 4.54). Accommodative range in emmetropic, non-presbyopic (see next paragraph) individuals is 5–6 dioptres. Contraction of the ciliary muscle may also steepen the corneal curvature, thus increasing its refractive power, as occurs in lower vertebrates. The parasympathetic neurones mediating this response course within the III cranial nerve bundle in the long ciliary nerves (see Chapter 1, p. 67). It is not clear which of the three sets of ciliary muscle fibres is responsible for the major action in inducing this forward movement but, as for the iris muscles, anticholinergic drugs can block the effects. There is also a small sympathetic inhibitory component and this is increased in late-onset myopes.

Changes in refraction occur throughout life. For instance, intense near work for prolonged periods disrupts emmetropization associated with eye growth and induces myopia. In addition, young children sleeping in dimly lit, as opposed to completely dark, rooms are at risk of developing myopia because of the persistence of poorly focused images through thin eyelid skin. With age, accommodative ability declines resulting from 'hardening' of the lens, while the ciliary muscle remains fully contractile. Reduced lens deformability is caused by increases in lens size in addition to changes in crystallin protein, combining to produce a 'hardened' lens (first stage of cataract formation). The lens equatorial diameter decreases by 0.055 mm for every dioptre decline in accommodative ability.

Blood–aqueous barrier

Aqueous humour, secreted by non-pigmented ciliary epithelial cells, is derived from plasma but contains different concentrations of electrolytes and other small molecules and a restricted set of proteins in low concentration. Thus, a filter barrier exists between the ciliary body stroma and the posterior chamber of the eye, which prevents free diffusion of molecules over a certain size from gaining access to the posterior and anterior chambers, which freely communicate via the pupil. Tight junctions between non-pigmented ciliary epithelial cells create this barrier (see Chapter 1, p. 30, Fig. 1.22). In contrast, extensive gap junctions between pigmented and non-pigmented cells allow the two layers of the ciliary epithelium to act as a metabolic and transport syncytium. However, in the iris, where tight junctions do not exist on the anterior surface, tight junctions between the iris vessel endothelial cells may act as a barrier, unlike the fenestrated vessels in the ciliary body. However, this concept has been challenged. The differences in protein composition between aqueous and plasma are not absolute and thus the barrier is 'leaky', since some high molecular weight proteins are present in aqueous humour. Instead, the protein transudate in the ciliary body stroma diffuses anteriorly into the iris stroma and thence into the anterior chamber down a concentration gradient (Fig. 4.55). Thus, aqueous humour in the anterior chamber always contains a small amount of protein, while aqueous humour in the posterior chamber is protein-free

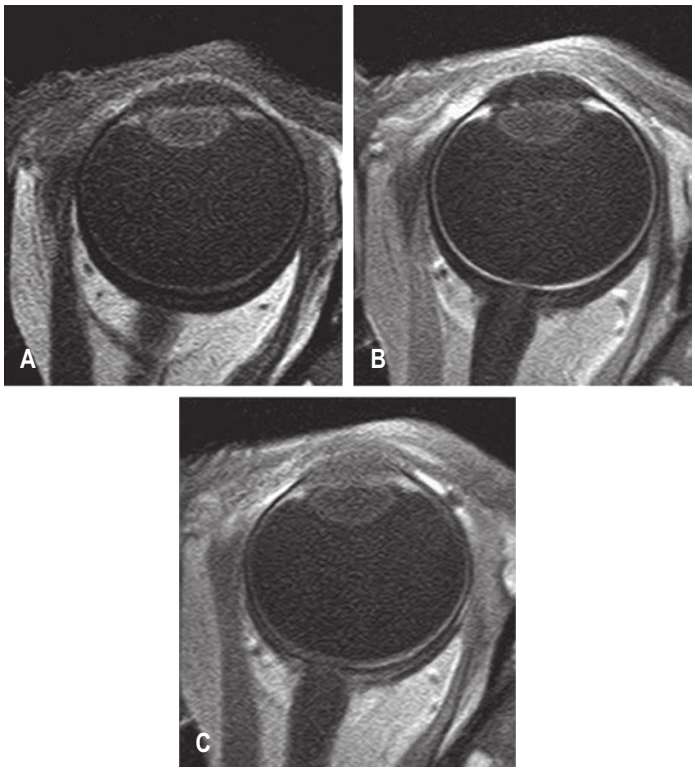


Fig. 4.55 Contrast-enhanced MRI images of the eye. (**A**) Pre-contrast image shows details of tissues in the anterior segment of the eye, including iris, ciliary body and both the anterior and posterior chambers. (**B**) Within 2 minutes of contrast infusion, there is clear enhancement of the ciliary body and the choroid, but the anterior chamber, posterior chamber and vitreous body show no enhancement. (**C**) After 90 minutes, the enhancement in the ciliary body and choroid has begun to diminish. There is clear enhancement in the anterior chamber but the posterior chamber and the vitreous body remain unchanged from (**A**) and (**B**). (From Freddo, 2013.)

(Fig. 4.56). Thus, the anterior chamber (iris, ciliary body and cornea) is not protected immunologically as previously thought (Forrester, et al., 2018) (see Chapter 7, 459).

Breakdown of the blood–aqueous barrier occurs in many conditions, including inflammation and vascular disease. The ciliary body/iris vessels become highly permeable and the stromal transudate increases with a great excess of proteins diffusing into the aqueous. The aqueous humour becomes visibly cloudy (seen as 'flare' by slit lamp biomicroscopy) because of plasma proteins in anterior chambers; it may even become a 'plasmoid' owing to the presence of fibrinogen and other proteins. Inflammatory cells in the ciliary body and iris stroma also migrate into the anterior chamber and traffic through the trabecular meshwork into the conjunctival vessels. If clotting occurs, as in severe uveitis, the aqueous humour becomes 'plastic'.

Eicosanoids in the iris/ciliary body

Prostaglandins were first discovered in the eye in 1957 by Ambache, who demonstrated this biological activity in aqueous humour and named it 'irin'. Eicosanoids is the generic term to describe prostaglandins (PGs) and leukotrienes, both of which are metabolites of arachidonic acid (see Box 4.8, p. 204). Prostaglandins are synthesized from arachidonic acid (AA) released from membrane phospholipids. This may occur in large amounts after trauma or inflammation involving the iris/ciliary body. Other neuropeptides are involved in this process. For instance, release of substance P from the iris leads to receptor-mediated breakdown of PIP_2, which in turn generates AA in the iris sphincter and leads to synthesis of PGE_2. In the ciliary body, the cyclo-oxygenase pathway is also active in microsomes (small vesicles derived from damaged ER). PGE_2 is involved in miosis, while $PGF_{2\alpha}$ is involved in the control of intraocular pressure. PGE2 promotes aqueous flare and is prevented in its absence (e.g. in GE2 receptor deficient mice). $PGF_{2\alpha}$ is also known to increase vasodilatation and capillary permeabilization in the anterior segment of the eye. Many other peptides are present in the iris/ciliary body and the aqueous humour, including neuropeptide Y, vasoactive intestinal peptide, somatostatin and calcitonin gene-related peptide (CGRP) many of which have an immunoregulatory role (see Chapter 6, p. 359,

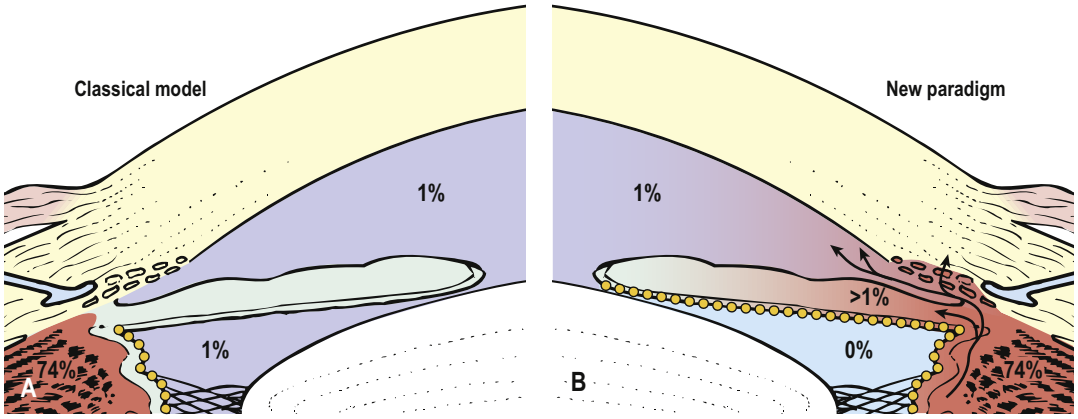


Fig. 4.56 (**A**) In the classic model, the tight junctions of the non-pigmented ciliary epithelium and of the iris vascular endothelium are considered to be the main components of the blood–aqueous barrier, by preventing the passage of molecules greater than 10 kDa into the anterior and posterior chambers. The iris stroma is presumed to be free of plasma proteins and the concentration of plasma proteins in the aqueous is uniform throughout the anterior and posterior chambers. Increases in aqueous humour plasma protein concentrations are considered the result only of an increase in barrier permeability. (**B**) In the new concept (paradigm) of blood–aqueous barrier physiology, the small amount of plasma-derived protein present in aqueous humour diffuses from the ciliary body stroma to the root of the iris, accumulates in the iris stroma and is then released into the aqueous humour of the anterior chamber only *(arrows)*. Thus, the posterior chamber is free of protein. Some of the protein delivered to the iris root immediately enters the trabecular outflow pathways (uveo-scleral outflow, *arrows*). The tight junctions of the non-pigmented ciliary epithelium and of the iris vasculature endothelium still provide the main barrier function. However, an additional key element becomes the tight junctions of the posterior iris epithelium, which prevent the protein in the iris stroma from diffusing posteriorly, when combined with the one-way valve created by the pupil resting on the anterior lens capsule, and the continuous forward flow of aqueous humour through the pupil. (From Freddo, 2013.)

Box 6.9). Nitric oxide (NO) is also released during activation of iris/ciliary body tissues. Many of these mediators modulate normal iris/ciliary body functions such as miosis and aqueous humour production. For instance, CGRP relaxes iris dilator smooth muscle via cAMP mechanisms.

Drugs that inhibit the cyclo-oxygenase pathway (see p. 210), such as indomethacin and aspirin, may be useful in ocular inflammation. However, steroids act at the level of phospholipase A_2 and may have a more global effect on the response (see Chapter 6, p. 374). In addition, the lipoxygenase pathway is active in the anterior uvea with synthesis of leukotrienes B_4, C_4 and D_4, and the chemotaxis of polymorphonuclear leucocytes (see Chapter 7, p. 393). Penetration of drugs into the intraocular environment after topical application occurs more readily through the conjunctiva and the sclera than the cornea, partly as a result of the numerous transport mechanisms available in the conjunctival epithelium but also because of the greater transepithelial permeability of the conjunctiva compared to cornea, which is impermeable even to low molecular weight compounds (<1000 kDa).

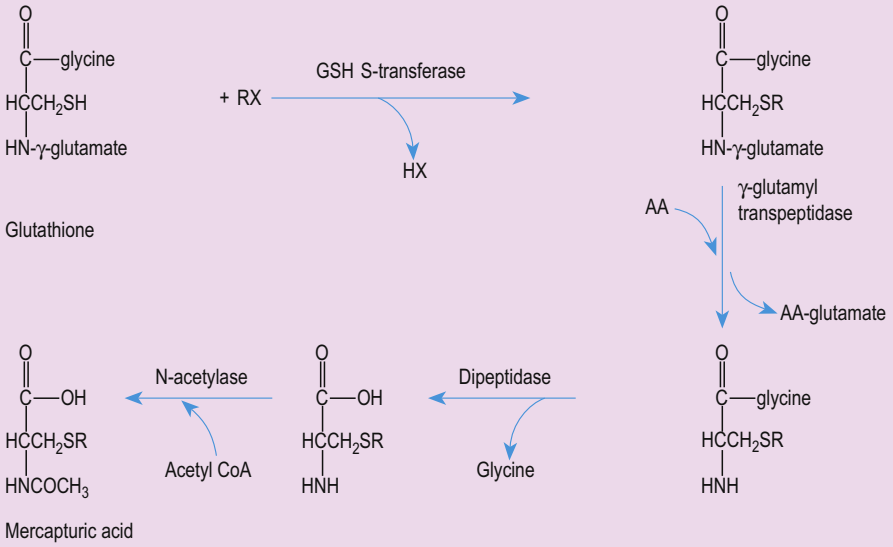
Detoxification and antioxidation in the anterior segment

The cytochrome P_{450} system is the major drug detoxification (CYP) system in the eye. The ER contains a group of proteins known as the cytochrome P_{450} proteins, which catalyse the transfer of a single oxygen atom to endogenous and exogenous substances destined for excretion and/or detoxification, such as steroids, phenobarbital, etc. (see Chapter 6, p. 359, Box 6.9). Their main effect is to convert hydrophobic compounds to hydroxylated hydrophilic compounds, which are then more easily metabolized. The CYP system is usually studied in microsomes isolated from liver ER.

The CYP system is present in the ciliary body (at about 5% of the concentration in liver) where it acts to detoxify many compounds. It does this by either converting the hydroxylated, highly reactive compound to a glucuronide via UDP-glucuronyl transferase or by conjugating it to glutathione via glutathione *S*-transferase (Box 4.11). Several CYP family enzymes have been identified and/or purified from the ciliary body, in particular the non-pigmented epithelium. For instance, the enzyme CYP2D6 is important for metabolizing the topical anti-glaucoma β-adrenergic blocker timolol. The CYP system is regulated by the aryl hydrocarbon receptor (AhR), a transcription factor, first discovered as a xenobiotic sensor but now recognized to be essential for several critical cell biological life processes. ArH is bound to a chaperone in the cytoplasm in an inactive state, but on binding a ligand, e.g. a toxic chemical or cell waste product, free ArH translocates to the nucleus

Box 4.11 **Cytochrome P_{450} and Drug Detoxification**

The cytochrome P_{450} system detoxifies compounds by utilizing the glutathione *S*-transferase system and degrading compounds to mercapturic acid.



bound to the ArH nuclear translocator (ArHNT) where it dimerizes and activates a wide range of genes including CYP genes. Activation of the ArH receptor can suppress inflammation and neovascularization suggesting that this enzyme and the CYP proteins have a major homeostatic role in ocular health.

The ciliary body is the main source of antioxidant systems in the anterior segment. Although antioxidant systems exist in the lens (see later) and the cornea (see earlier), the ciliary body is especially the richest ocular tissue in antioxidant systems with the highest concentrations of catalase, superoxide dismutase, and glutathione peroxidase types I and II. Type I is selenium-dependent and type II is selenium-independent. Type I is closely linked to glutathione reductase whose main function is the reduction of oxidized glutathione (GSSH) produced by the detoxification of peroxides (see Box 4.11).

Hydrogen peroxide (H_2O_2) is present in normal aqueous, most of it derived from the non-enzymatic interaction between reduced ascorbate and molecular oxygen and it is reduced to H_2O by glutathione secreted by the ciliary epithelium. Most of these studies have been performed in experimental animals and it is not clear how relevant they are to the human eye. It has been suggested, for instance, that oxidized ascorbate (via the superoxide anion) is more important in degrading H_2O_2 in humans. Melatonin, a neuropeptide involved in biological circadian rhythms, is synthesized in the ciliary body from serotonin and is also an H_2O_2 scavenger. A role for xanthine oxidase has also been suggested. H_2O_2 can induce noradrenaline (norepinephrine) release from the iris/ciliary body in the aqueous and has recently been implicated in cataract formation.

Ciliary body tissue also contains a peroxiredoxin, a constituent of a widely distributed family of antioxidant enzymes, whose amino acid sequence and tissue distribution are now known. Their role is to degrade H_2O_2 and alkyl peroxides.

The choroid

Functions of the choroid

The function of the choroid is to act as the lymphovascular supply to the posterior segment of the eye.

Vascular function. The choroid is almost entirely composed of vessels embedded in a loose Type III collagen-rich matrix typical of an expansile or spongy tissue. The blood supply to the choroid has several interesting features:

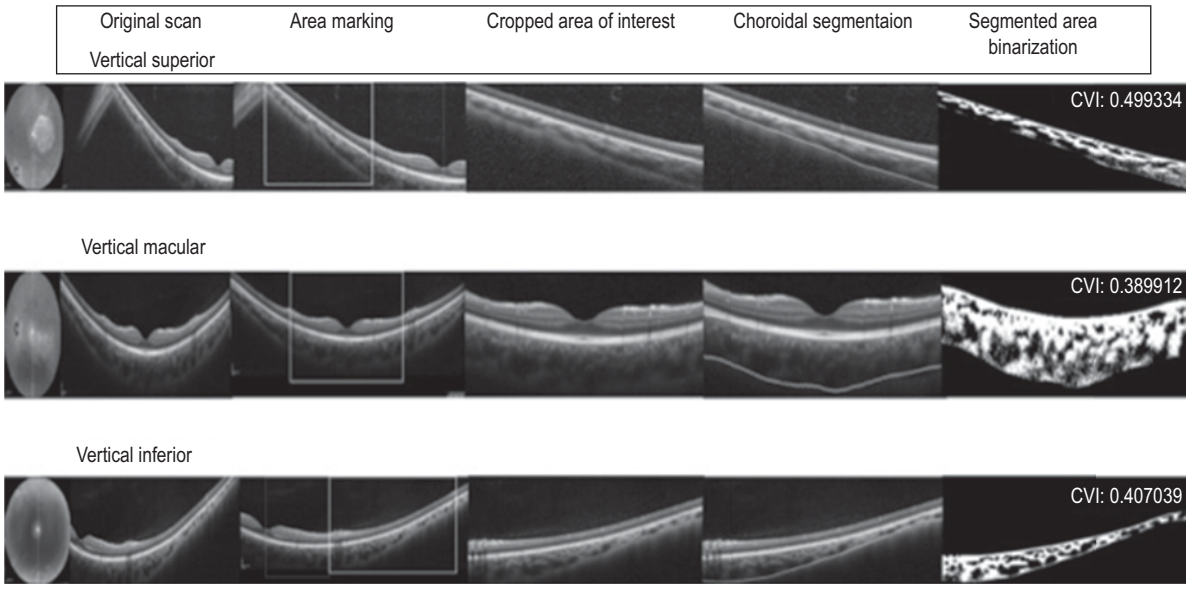


Fig. 4.57 Measurement of the thickness of the choroidal vasculature by optical coherence tomography. Quantitative image processing was performed to evaluate digitally the thickness of the choroidal vasculature in healthy young men and presented as a choroidal vasculature index (*CVI*). Significant topographical variation occurred between the macular, nasal and temporal choroid. (From Singh et al., 2018.)

- ~ 98% of the blood to the eye passes through the uveal tract, of which 85% is through the choroid.
- Blood flow occurs at a rate of 1400 mL/min per 100 g tissue, which is higher than the perfusion of blood through the kidney.
- The choriocapillaris has a lobular architecture, collecting into larger vessels and finally into four vortex veins, one in each quadrant of the globe (see Chapter 1, p. 55).
- Venous blood draining from the choroid is not desaturated; only 5%–10% oxygen is extracted during passage through the eye. The choroid supplies the outer retina, where the partial pressure of oxygen (Po_2) is highest, rapidly falling towards the retinal inner segments and then rising again, less so, towards the inner retina.
- Blood vessels in the choroid are highly fenestrated and leaky, like ciliary body vessels.
- Choroid and ciliary body blood vessels are sensitive to the Po_2 and Pco_2; in conditions of high Pco_2 the vessels expand greatly in a forward direction against the unyielding sclera, altering the forward position of the retina, causing forward displacement of the vitreous gel and lens/iris diaphragm and risking angle closure especially in small (hyperopic) eyes.
- Although previously considered to be subject to autoregulation in relation to perfusion pressure, choroidal blood flow, especially around the optic nerve, is sensitive to the effects of nitric oxide and endothelin and other as-yet unidentified vasoconstrictors.

The inner choroid (layer of the choriocapillaris) can be visualized clinically by OCT but even with enhanced digital imaging it is difficult to estimate the full thickness of the choroid (Fig. 4.57).

Non-vascular functions. The choroid has additional functions:

- It is involved in ocular temperature control by dissipating heat from the eye.
- It secretes growth factors which control scleral thickness and may have a role in emmetropization (see eFig. 4.22).
- It drains aqueous fluid from the anterior segment of the eye via uveoscleral outflow, which may be as much as 40% of aqueous outflow in humans (see p. 233).
- Higher primates, including man, have intrinsic choroidal neurones that may have a role via nitric oxide in controlling vascular diameter and blood flow to the choroid.
- In humans and other primates, non-vascular smooth muscle cells may be found in some cases forming discrete rings around blood vessels and even in the region of the fovea. The choroid also has variable numbers of melanocytes, which are very large cells spanning with extensive dendrites.

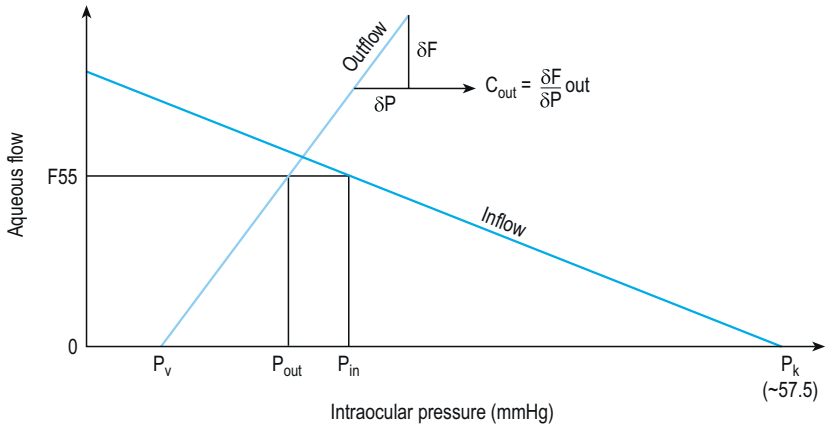


Fig. 4.58 *Aqueous* inflow declines as *intraocular pressure* rises towards P_k (about 57.5 mmHg). Goldmann estimates pseudofacility as $P_k = 0.58 \times$ brachial BP. The gradient for inflow is $(F_{55}/P_k - P_{55}) = C_{in} = C_{ps}$ where F_{55} is steady-static flow. C_{ps} is the pseudofacility. (Courtesy D. Woodhouse.)

Lymphoid function. Intraocular structures lack a recognized lymphatic system. However, through the uveoscleral outflow, drainage of intraocular fluid into conjunctival (aqueous veins) and orbital lymphatics occurs. Some reports of suprachoroidal lymphatic lacunae have been documented. The choroid contains a rich network of immune cells including mast cells, macrophages and dendritic cells (DC). DC migrate to the eye-specific draining lymph node using this route (in humans these are the pre-auricular and submandibular lymph nodes; see Chapter 1). This tissue can respond massively to intraocular inflammation which is visibly as forward swelling on OCT. (see Chapter 7, p. 461). Mast cell degranulation has been implicated in posterior segment inflammation (uveitis) as well as more recently in atrophic age related macular degeneration. In addition, choroidal and ciliary body/iris melanocytes are potential antigenic targets for autoimmune disease (see Chapter 7, p. 446).

Aqueous humour dynamics

A fundamental physiological function of the eye is to maintain an intraocular pressure (IOP) between 10 and 20 mmHg. This is achieved by the circulation of aqueous humour secreted by the ciliary body into the posterior chamber and circulated through the pupil towards the anterior chamber angle where it drains via the outflow apparatus into the episcleral veins (see Chapter 1, pp. 24-25). Factors affecting IOP include:

- circadian rhythms
- episcleral venous pressure
- rate of secretion and flow of aqueous humour
- neural (cranial nerves V and VII) and hormonal influences.

The intraocular pressure, as measured clinically, actually represents the balance between the inflow and outflow of aqueous and is altered by changes in the gradient of pressure between the posterior chamber and the anterior chamber and eventually by the episcleral pressure. The direction of uveoscleral outflow is determined by this gradient of pressure (Fig. 4.58).

Aqueous humour is secreted by the ciliary body epithelium

The rate of aqueous humour formation is about 2–3 μL/min. Aqueous humour is formed by the transport of water and electrolytes from the leaky fenestrated capillaries of the ciliary processes to the epithelial syncytium and thence across the plasma membrane of the non-pigmented epithelium (see Box 4.12).

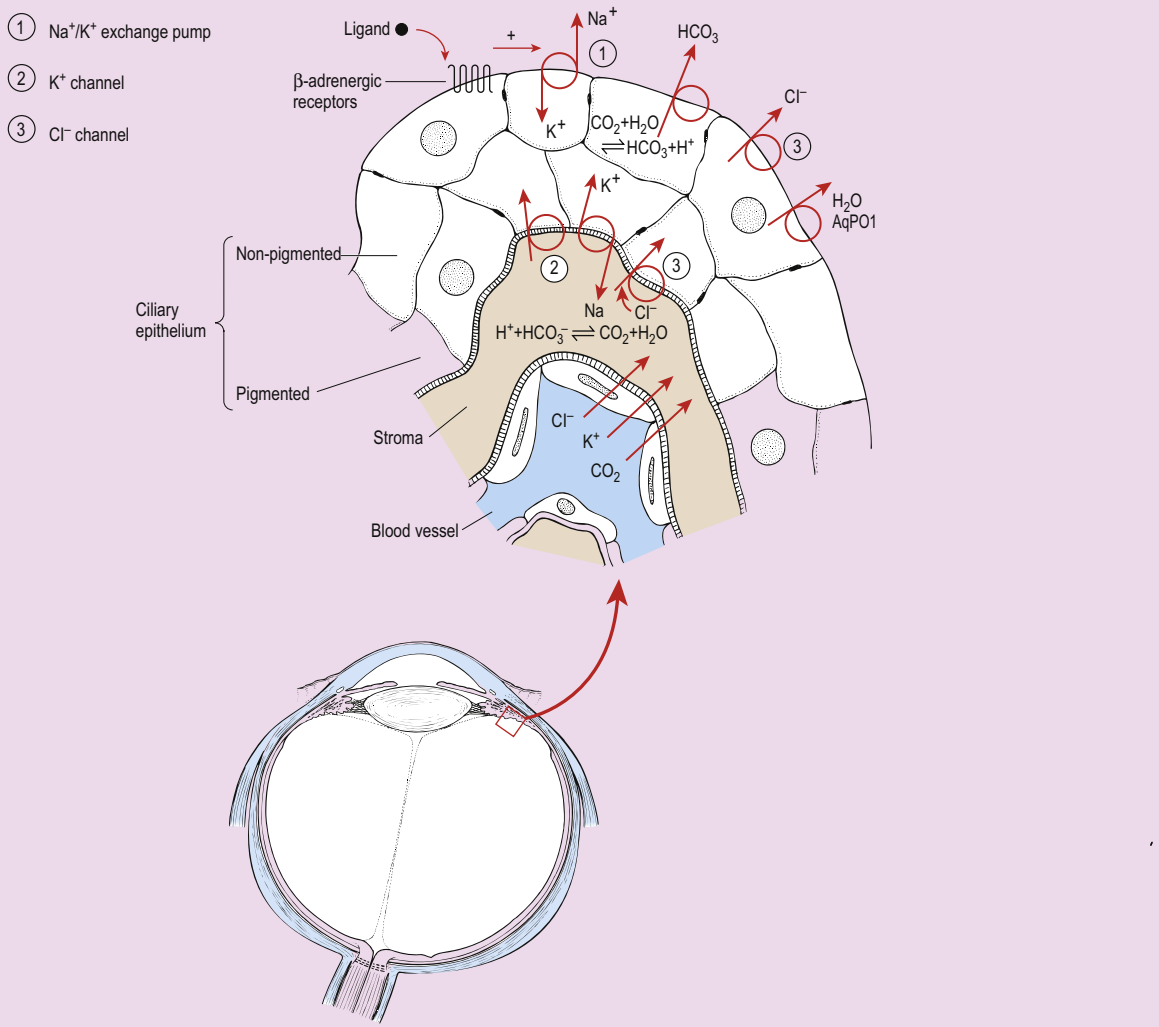
Composition of aqueous humour

The aqueous humour is composed predominantly of electrolytes and low molecular weight compounds. A low concentration of protein is present in the anterior chamber (AC) aqueous derived from the iris stromal transudate (see p. 226, Fig. 4.56) but posterior chamber (PC) aqueous is protein free and the direction flow maintains this differential. Protein is shed into the AC also from the corneal endothelium and the iris/lens cells. Several trace compounds are also present in aqueous humour, including steroid hormones, enzymes (carbonic anhydrase, lysozyme and plasminogen activator) and cytokines such as basic fibroblast growth factor (bFGF) and TGF-β (see Chapter 7). Low levels of catecholamines (adrenaline, noradrenaline and dopamine), prostaglandins and cyclic nucleotides are present in normal aqueous, but the source of these compounds is uncertain.

Box 4.12 **Secretion of aqueous humour**

Classic theory suggests that passive diffusion of water and ions from the fenestrated vessels of the ciliary body is followed by active transport of Na^+ and Cl^- across the ciliary body syncytium. This is an active secretory process involving Na^+/K^+ ATPase and carbonic anhydrase type II activity. In some respects, this has been viewed as ultrafiltration of ions and water and eventually leads to the formation of aqueous humour, which is secreted into the posterior chamber. However, the oncotic pressure of the ciliary stroma is greater than the hydrostatic pressure difference across the ciliary epithelium, so tending towards absorption of water into the ciliary body from the posterior chamber. Thus, active transport of ions in the opposite direction is the main mechanism of aqueous humour formation. This process is also under adrenergic receptor control at the level of the ciliary epithelial cells and possibly also by regulation of blood flow to the ciliary body. The possibility has been raised that H_2O transport may also be achieved by aquaporins. The ciliary epithelium has been shown to express two aquaporins, AQP1 and AQP2, but these do not seem to play a major role in aqueous humour production. In addition, some of the K^+ channels are Ca^{2+}-sensitive and can be activated by Ca^{2+} entry via Ca^{2+} channels. More recently, the role of the Cl^- ion in aqueous humour formation has assumed greater importance through its transport via the cystic fibrosis transmembrane receptor. This may play a greater role than the HCO_3^- carbonic anhydrase-mediated mechanisms and may also be facilitated by adenosine receptors.

Aquaporin water channels are also present in both the secretory machinery (AqPO4 and AqPO1) and the outflow channel (trabecular meshwork endothelium) and appear to contribute to bulk flow of water because intraocular pressure is reduced in AqPO4/1 knockout mice.



The total protein content of aqueous is very low (about 1/500 of plasma) and the major species is albumin. Immunoglobulins in normal aqueous humour are unlikely to be plasma-derived; rather, local production via iris lymphocytes or plasma cells is a likely source since local antibody production in the eye has been well recorded. Small amounts of fibronectin are also produced locally.

Aqueous humour contains detectable amounts of hyaluronic acid, derived as breakdown oligomers from vitreous hyaluronic acid during the normal process of GAG renewal (see p. 194). However, some of the aqueous hyaluronan is of higher molecular weight than vitreous, suggesting that it is produced in the anterior segment. Aqueous hyaluronan may have a role to play in regulation of IOP because perfusion of the anterior chamber with hyaluronidase leads to a marked drop in pressure. However, it is also possible that intracameral hyaluronidase affects trabecular meshwork cells and extracellular matrix, thus leading to a lowering of the IOP.

Neural/autonomic control of aqueous secretion

Adrenergic and cholinergic agonists and receptors are present in the iris and ciliary body, and autonomic innervation of this tissue occurs in muscle, vessels and epithelial cells. Adrenergic receptors are present in the ciliary epithelium and regulate IOP via the adenylate cyclase system. β-Adrenergic antagonists and α_2-selective adrenergic agonists both suppress aqueous flow. Muscarinic receptors linked to the PIP_2 second messenger system are also present in the ciliary epithelium.

Cholinergic mechanisms are not involved in control of intraocular pressure

M_3 muscarinic receptors in the ciliary epithelium cell membrane have been identified, but they do not appear to have a significant role in IOP control. Despite this, cholinergic agents such as pilocarpine are thought to have some action via reduction in aqueous secretion, although the experimental evidence is weak. Most of the action of pilocarpine appears to be mediated via its effect on outflow resistance and uveoscleral flow (see later).

Adrenergic receptors regulate IOP via adenylate cyclase

The majority of α receptors in the ciliary body are α_2, while more than 90% of β receptors are β_2. Stimulation of α_2 receptors lowers the IOP via a reduction in aqueous humour production through inhibition of adenylate cyclase. Adrenaline, a preferential α-adrenergic agonist, stimulates prostaglandin synthesis, particularly that of PGE_2 and $PGF_{2\alpha}$, the latter having potent ocular hypotensive activity. Stimulation of β receptors, particularly β_2 receptors, also leads to an increase in aqueous secretion via activation of adenylate cyclase.

The dual control of aqueous secretion through activation (β) or inhibition (α) of adenylate cyclase is mediated by their respective stimulatory and inhibitory G proteins (see Chapter 6, p. 355). Thus, α_2 agonists (e.g. clonidine) or β_2 antagonists (β-blockers, e.g. timolol) both have the ability to lower IOP. The α_2 receptors are also linked to vasoactive intestinal peptide receptors, which are co-stimulated and lead to a reduction in cAMP levels thus indirectly lowering IOP. β antagonists have no effect on aqueous flow when aqueous production is at its lowest, whereas α_2 agonists and carbonic anhydrase inhibitors do.

The mechanism whereby changes in intracellular cAMP levels alter aqueous secretion is not known but appears to involve transport of the HCO_3^- ion across the cells (Box 4.12). In addition, a number of other components appear to be involved in aqueous secretion, such as protein kinase C, which is linked to adenylate cyclase activation and thus may act as part of an intracellular signalling network connecting the two main second messenger systems (see Chapter 6, p. 355).

Nucleotides and nucleotide receptors

Low molecular weight nucleotides such as adenosine are present in aqueous humour in significant quantities (4^{-4} μM) and are released from cells by exocytosis, channel transporter or by cell damage/death. Activation of the purinergic P1 receptors (also known as the adenosine, A_3 receptors) in the ciliary epithelium increases aqueous inflow through stimulation of Cl channels (Box 4.12) while binding to A_1 receptors in the trabecular meshwork (see later) increases outflow by activation of MMPs (see Chapter 7, p. 394, Table 7.2). The wide-ranging effects of nucleotides, particularly adenosine, are also mediated via the P2 receptors (P2X and P2Y) (Sanderson et al., 2014). Topical application of some nucleotides to the eye leads to lowering of the IOP.

Interestingly, melatonin has IOP-lowering effects and the melatonin analogue 5-methoxycarbonylamino-N-acetyltryptamine (5-MCA-NAT) has similar effects. Melatonin synthesis is regulated by retinal ganglion cell photopigment, melanopsin (see p. 256), which has also been detected in the lens. A recent study using yellow filters showed that deactivating melanopsin activity increased melatonin levels in the aqueous humour and lowered IOP.

Cannabinoids are also known to reduce IOP but the effect is transient and offset by the many side effects.

Circadian regulation of aqueous humour formation

There is a diurnal variation in IOP with peak levels around 5–6 am and trough levels around 9–10 pm, possibly caused

by melatonin-based mechanisms. This is in part a result of circadian regulation of aqueous humour secretion. Secretion in humans occurs at a rate of 2.6 mL/min during the day and falls to 1.0 mL/min at night. Both β-adrenergic receptor-mediated and neuropeptide-mediated mechanisms, particularly that of vasoactive intestinal peptide, are involved. Activation of G protein-coupled adenylate cyclase leads to cAMP production, which activates protein kinase A (PkA), thus regulating the cation channels. The process is terminated by hydrolysis of cAMP by phosphodiesterase.

Is guanylate cyclase involved in intraocular pressure control?

Large amounts of brain natriuretic peptide (BNP) as well as atrial natriuretic peptide (ANP) are found in the iris/ciliary body and in the aqueous humour of rabbits and humans. Binding of the receptor for ANP activates guanylate cyclase in the ciliary body and lowers IOP. Activation of NO-guanylate cyclase pathway also leads to reduced IOP but in this case action is on the outflow pathway (see next section) (Wareham et al., 2018).

Aqueous humour outflow from the eye

Flow of aqueous humour from the eye is controlled at several different levels, including at the site of the conventional (C_t) outflow pathway via the trabecular meshwork, at the uveoscleral (C_u) outflow system, and, outside the globe, at the site of the episcleral veins (see Chapter 1, p. 24).

Control of outflow at the trabecular meshwork

Resistance to the outflow of aqueous humour occurs at the level of the endothelium covering the trabecular meshwork (TM). This cell layer is an extension of the corneal endothelium, but also extends into the matrix of the TM itself. The hydraulic conductivity at this site is around 10^{-7} cm^{-2} s^{-1} g^{-1}, which is several times greater than for other lining endothelia and is caused by water transport by transcytosis, in which membrane-bound water-containing vesicles are transported through micron-sized pores in the cell, across the meshwork. Formation of transcellular pores may be a pressure gradient-induced mechanoreceptor-based mechanism involving the actin–myosin cytoskeleton and appears less effective when the cells are stiffened, as occurs with agents that increase outflow resistance such as thrombin and sphingosine-1-PO_4.

Despite the potential contribution of the trabecular endothelial cell, the juxtacanalicular cribriform meshwork is the main site of resistance to aqueous flow into Schlemm's canal. This is due to matrix components, especially GAGs, present in this region. The trabecular beams (type I with significant amounts of type III and IV collagen) and other matrix constituents such as laminin, fibronectin and elastin, are separated by GAG-filled spaces, particularly hyaluronan, that retard the flow of fluid by virtue of its hydrophilic properties and large hydrodynamic volume (Fig. 4.59). Several GAGs occur in the trabecular meshwork matrix—namely, hyaluronan, chondroitin sulphate, dermatan, keratan and heparan sulphates—with significant variation between species and some unspecified proteoglycan material. Trace amounts of types V and VII collagen also occur in the trabecular meshwork.

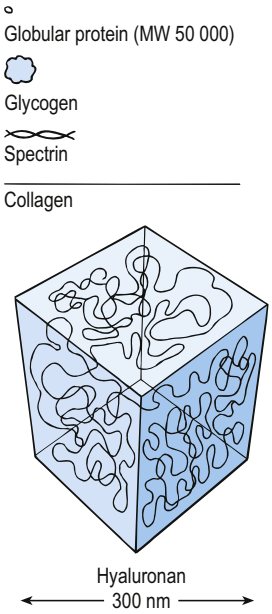


Fig. 4.59 Model of hyaluronan, showing size compared to other 'typical' molecules. Courtesy J. Alberts.

Apart from the mechanosensor or stretch receptor function of trabecular meshwork (TM) endothelial cells discussed above, TM endothelial cells have special characteristics, namely: active phagocytic properties, high levels of cytoskeletal actin (particularly sensitive to cytochalasin B) and low levels of microtubules (relatively non-responsive to colchicine). These cells also contain vimentin and desmin, thus showing some similarity to smooth muscle cells. Taken together, these findings suggest that TM endothelial cells are specialized for both endocytic transport of water and solutes, and contractility. Actin mobilization appears to be mediated via adrenergic receptors, probably of the β_2 type, which are highly responsive to adrenaline. In addition, they respond to NO via the NO-guanylate cyclase pathway and increase flow via a mechanism analogous to NO-induced vasodilation. Glucose (energy) metabolism in TM endothelial cells is predominantly glycolytic rather

than oxidative, although both enzyme systems are present and functional. Transportation of water may be achieved not merely by passive transport of H_2O packets but by activation of a water channel protein, aquaporin-1 (AQP01), found in corneal endothelial cells.

Metabolism of trabecular meshwork cells

The matrix components of the TM are thus synthesized and degraded by the endothelial cells. In addition, these cells have high levels of surface tissue plasminogen activator (tPA), higher even than in vascular endothelial cells, which is likely to play a role in maintaining patency and reducing the resistance of the outflow passages. Phagocytic activity of trabecular meshwork cells is associated with several other enzymatic activities such as GAG-degrading enzymes and acid phosphatase.

TM cells have receptors for a variety of agents including adrenaline (β_2 adrenergic receptors, decrease phagocytosis) and glucocorticoids. Both steroids and oxidative damage induce the expression of the trabecular meshwork inducible glucocorticoid response protein, myocillin. Defects in the myocillin gene are linked to glaucoma. Myocillin, present in many ocular tissues, is shed in exosomes into the aqueous humour and could potentially obstruct TM outflow if released in excess and not cleared adequately. Myocillin is found in the Golgi apparatus intracellularly and may be present in a complex with other proteins including CYP proteins (products of drug detoxifying gene, see Chapter 6, p. 353).

Other glaucoma susceptibility genes such as the optineurin and the WD repeat domain (*WDR36*) genes account for a small proportion of patients with primary open angle glaucoma and the *CYP1B1* gene in congenital glaucoma. In addition, TGF-β is involved in the glucocorticoid cell response and in the production of TM extracellular material. Steroids inhibit prostaglandin secretion by TM cells at concentrations as low as 10^{-8} mol/L. Prostaglandin synthesis by TM accounts for a significant proportion of its arachidonic acid metabolism (70% compared with less than 5% in other cells), suggesting that prostaglandins play a major role in trabecular meshwork cell physiology. In addition to substantial amounts of PGE and $PGF_{2\alpha}$, leukotriene B_4 is produced in high amounts.

TM cells contain free radical and H_2O_2 detoxifying enzyme systems present in other tissues such as the ciliary body (see p. 229). Both a catalase and a glutathione-dependent system are active in handling H_2O_2, which can reach levels as high as 25 μmol/L in the aqueous humour. Ageing is associated with increasing free radical damage and mitochondrial dysfunction and clearance of waste products by the ubiquitin–proteasome system and other lysosomal and non-lysosomal enzyme systems such as the calpains of TM cells becomes less efficient. Waste products thus accumulate in the TM itself ('the garbage catastrophe theory of ageing') as a result of the imperfect removal of oxidative damage and contribute to increased outflow resistance, ocular hypertension and glaucoma. This age-related process will also be affected by genetic susceptibility as shown by the identification of two mutations in the lysyl oxidase gene in patients with secondary glaucoma associated with pseudoexfoliation syndrome. Lysyl oxidase is required for elastin production (see p. 189) and presumably vitreous/zonule components are more readily degraded and washed through to the TM in this syndrome.

There are now many more genes and gene loci (~22, termed GLC-XX loci) linked to glaucoma with the major genes being *CYP1B1*, *myocillin*, *IL20RB*, *neurotrophin-4*, *optineurin*, *ankyrin repeat and SOCS body10 (ASB10)* and *WDR36*, but the frequency of any one gene in a glaucoma population is low while the number of patient with no known gene association is high. The physiology of IOP control is highly complex and the pathogenesis still obscure.

Uveoscleral drainage

A variable proportion of aqueous humour (up to 40%) drains directly into the anterior uvea at the ciliary body immediately posterior to the cornea and thence into the suprachoroidal space and towards the posterior pole of the eye (see Chapter 1, p. 26). The anterior uvea at this point is incompletely lined with endothelial cells. In addition, the localization of MMP-1 suggests a role for this enzyme in uveoscleral outflow. Uveoscleral drainage is possible because the pressure in the suprachoroid is 2–4 mmHg lower than in the anterior chamber; this can be reversed after trabeculectomy and can lead to choroidal effusion. This pressure differential is also less with age, leading to greater risk of choroidal effusion in such patients. Several possible mechanisms of uveoscleral flow have been proposed, including relaxation of the ciliary muscle, cell shape changes, cytoskeletal rearrangements or compaction of the trabecular meshwork matrix. Prostaglandin ($PGF_{2\alpha}$) analogues decrease the IOP by increasing the uveoscleral outflow and newer PGs combined with NO-donors may have an even greater effect on outflow (e.g. latanoprostene bunod).

Episcleral circulation

It was shown many years ago that dye-stained aqueous fluid would not drain out of the eye into the episcleral veins if the IOP was less than 15 mmHg. This was because of resistance from the combined episcleral venous pressure and the oncotic pressure in the perivenous tissues of the episcleral veins. Aqueous humour drains via Schlemm's canal into the aqueous veins, facilitated by passing through large transcellular channels and giant vacuoles on the meshwork side of the canal (see earlier). The canal has direct vascular communications on its outer wall with a network

of intrascleral collector channels that drain into the scleral veins (see Chapter 1, p. 23). A pump mechanisms may also play a part in aqueous flow generated by small changes in IOP and linked to the ocular pulse. Small valves assist pumping of aqueous from Schlemm's canal into the collecting veins and episcleral veins. Episcleral blood flow, viewed by haemoglobin video imaging (Meyer, 2018), demonstrate the pulsatile nature of the blood flow in delivery vessels, as well as the potential for reversal of flow and its peristaltic nature probably influenced by the valves. Aqueous humour is visualized as a central stream with the blood column.

The episcleral venous pressure can be measured using a non-invasive manometer (the EV-310) and is normally around 8–10 mmHg.

An intraocular pressure-independent link between aqueous production and aqueous outflow?

Bestrophin anion channels occur in the ciliary body, including the HCO_3^- channel bestrophin-2 which is selectively expressed in non-pigmented epithelial cells. Unexpectedly, bestrophin-2-deficient (best2 −/−) mice have a higher rate of aqueous inflow but a reduced IOP. This appears to be because of the very high level of a soluble adenylate cyclase in the non-pigmented epithelium, which is acutely sensitive to HCO_3^- but is absent from the outflow pathway cells. This suggests that the ciliary body may produce substances via HCO_3^--sensitive soluble adenylate cyclase, which directly influences outflow independent of pressure.

Does aqueous humour contain components that contribute to flow resistance?

Although aqueous humour has the same viscosity as isotonic saline, its passage through microporous filters *in vitro* is slower than that of saline. This effect can be abolished by proteolytic agents and detergents, but not by hyaluronidase. It has therefore been suggested that some forms of glaucoma may be caused by a build-up of a surfactant-like material with age.

The lens

The transparency of the lens is a function of the highly ordered state of its cells and extracellular matrix. In essence, the extracellular matrix of the lens is confined to its capsule, while the cells form a syncytium with interlocking cellular processes. Extensive communicating gap junctions between the cells bring about the syncytial arrangement, but the cells are not fully apposed, leaving a nanometre-sized space between cells, which permits an important and essential circulation within the lens.

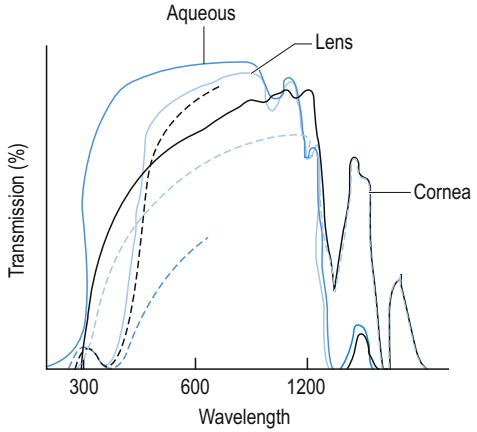


Fig. 4.60 Optical density curves for various ocular components. (With permission of Drs Jaffe and Horwitz and JB Lippincott.)

Lens transparency

Light passing through the lens is subject to scatter (stray light) in the same way as the cornea. Scatter occurs in all directions, back scatter being reflected light and forward scatter being detected by the retina. Scatter induced by the lens is minimal because of destructive interference based on the molecular architecture of the lens (see later). Light energy is therefore mostly transmitted through the lens (as well as the rest of the media).

The light transmission properties of the ocular media vary to some degree depending on the nature and the age of the tissue. The cornea, aqueous humour and lens all transmit long-wavelength light well above the limit of visible light (about 720 nm). However, the cornea absorbs short wavelengths below 300 nm but wavelengths above 300 nm are transmitted through the aqueous humour; the lens further filters the majority of the short wavelengths below 360 nm and is an absolute barrier to light below 300 nm (Fig. 4.60).

The epithelium

The single-celled epithelium of the lens and its capsule do not scatter or reflect light, essentially because the combined refractive index (RI) is the same as that of the aqueous humour (1.336). However, the epithelium is of great importance in the maintenance of fluid and electrolyte balance of the lens syncytium via ion-pump mechanisms (see later). Thus, any agent that disturbs epithelial function and/or viability (such as ionizing radiation to the lens bow region) will have significant effects on lens clarity. This applies to all aspects of lens structure and function.

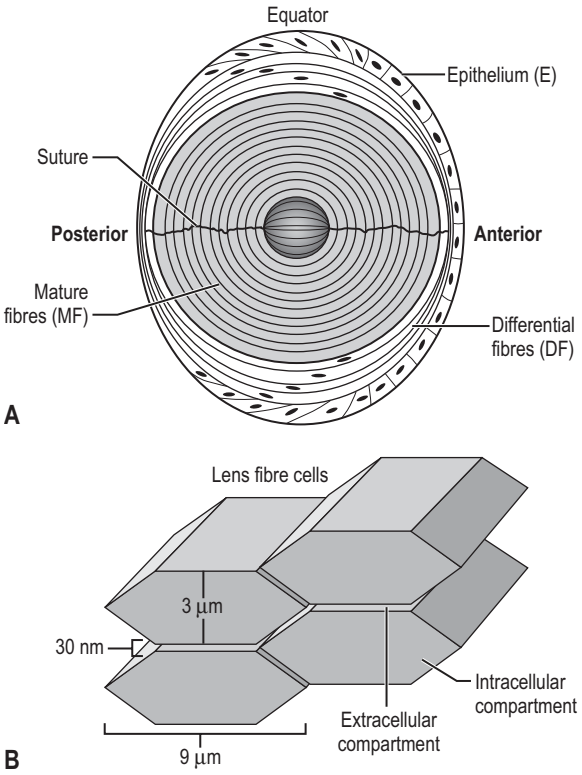


Fig. 4.61 (**A**) Three zones comprise the lens structure: the epithelium (function: transport of proteins), the anucleate mature lens fibres (function: light transmission) and the differentiating fibres (retain some organelles: function: nutrient transport link between epithelium and lens fibres). (**B**) The lens fibres appear as flattened hexagons when cut in cross-section. (From Mathias et al., 2010.)

The organization of the lens fibre cells underpins the transmission properties of the lens

It might be expected that the plasma membranes of the lens fibres would produce interference diffraction patterns that would affect the ability of the lens to transmit light. However, it has been shown that the weak diffraction rings that are produced occur in a repeating pattern with a period of the same dimension as the thickness of an individual lens fibre in the anteroposterior axis. This reduces any scatter by the plasma membranes of lens fibres and the healthy epithelium from which they arise; it has been estimated that the amount of scatter by the epithelium in the human lens is about 5% of the transmitted light.

Lens fibres differentiate from mitotic equatorial lens epithelial cells, which migrate posteriorly and centrally as new fibre cells are generated. This produces a zone of differentiating fibre cells that not only lose their nuclei but also most of their intracellular organelles as they become mature cells towards the lens centre (nucleus) (Fig. 4.61). This loss of cellular organelles as well as their anteroposterior orientation assists in allowing light transmission. The fibres are organized in a densely packed cellular arrangement, with interdigitations like pieces in a three-dimensional jigsaw puzzle (see Chapter 1, p. 33) in which extensive intercellular communication exists via lens fibre gap-junctions, generated by a set of heterodimeric proteins composed of connexins (Cx43, Cx46 and Cx50) (Fig. 4.62) as well as the gap-junction-like protein, aquaporin-0 (AQP0). Lens fibres essentially become crystallin-replete (over 90% of the total cellular protein) sacks in which crystallins are embedded within a complex cytoskeletal matrix, some components of which are also lens-specific (e.g. the beaded intermediate filament protein). The polydisperse nature of the crystallins prevents spontaneous crystallization in the packed arrangement. The high RI of the lens is caused by the crystallins; at the periphery of the lens the RI is slightly less (1.38) than at the nucleus (1.41). The water content of the lens is also greater at the periphery (75%–80%) than at the lens nucleus (68%).

As the lens fibres differentiate towards the centre (lens nucleus), this packing arrangement becomes modified and less compact (Fig. 4.63). The presence of the crystallins is in itself insufficient to explain the transparency of the lens. Transparency is predominantly the result of the packing of the crystallins in very high concentration such that they resemble a dense liquid or a glass because of the high level of 'short-range spatial order'; this means that the scatter of light from each individual molecule is related to the scatter from its immediate neighbours and that they tend to cancel each other out. At a macroscopic level, the arrangement of the crescent-like fibre cells in end-to-end concentric shells around a polar axis provides a highly ordered architecture. A series of coaxial refractive surfaces, thus created, promotes transparency of the multicellular structure. However, this general view does not apply to all species and in most there are two types of fibre cell, an S-shaped cell and a concentric cell. Overlapping tails of S-shaped cells form the lens sutures, which paradoxically lie along the visual axis and can affect optical quality by producing variability in focus.

In addition, the circulation within the lens provided for by the nanometre space between the gap junctions in fibre cells is essential for lens fibre metabolism even at its basal level. Mutations in gap junction connexin proteins account for many different types of congenital cataract.

The crystallins

Crystallins make up 90% of the water-soluble proteins of the lens. There are three types in mammals: α, β and γ, mainly on the basis of molecular weight with subtypes (e.g. αA, αB etc).

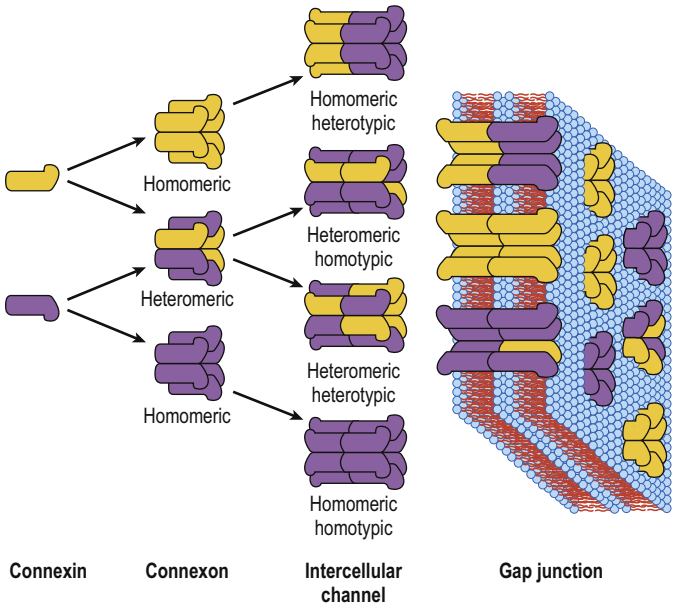


Fig. 4.62 Gap junctions are a major feature of cellular connections in the lens: they comprise many closely packed channels each formed by two hemichannels *(connexons)*, one in each of the connecting cells. (From Mathias et al., 2010.)

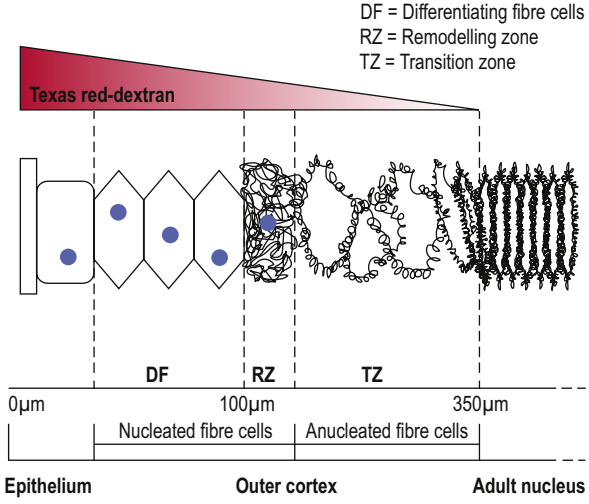


Fig. 4.63 Summary diagram of the changes in the lens as adult lens fibres differentiate from lens epithelium. (From Borchman and Yappert, 2010.)

Structural and functional information on the crystallins is available online at https://expertconsult.inkling.com

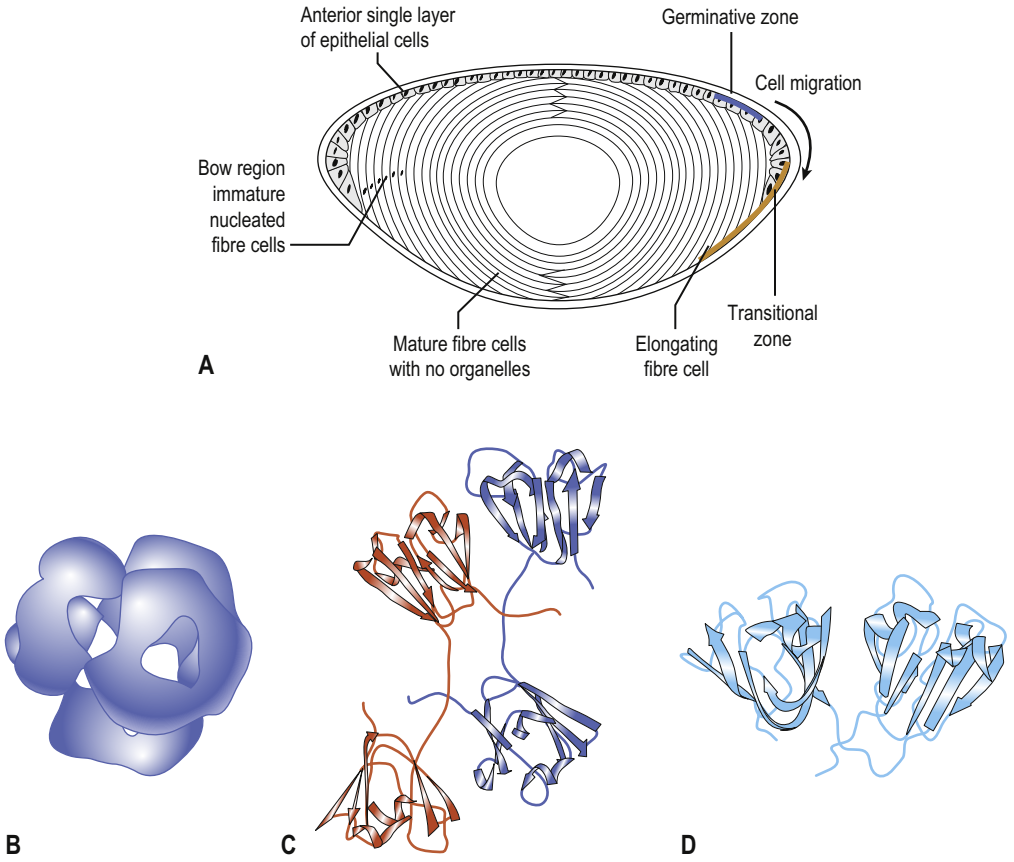
Some crystallin protein also coextracts with the urea-soluble protein, indicating that a fraction of the crystallins is strongly bound to the cytoskeletal (urea-soluble) proteins (5% total protein). The water/urea-insoluble protein represents membrane protein (2% total protein) and some crystallin is also found in this fraction when it is solubilized in detergent.

αA and αB crystallins show about 50% sequence homology. The molecules exist as polydisperse globular proteins in aggregates organized in three concentric layers or as a protein micelle. However, their true quaternary structure is unknown. αA and αB crystallins belong to the family of small heat-shock proteins and display chaperone-like activity. Studies in knockout mice have shown that αA lens-specific crystallin is required for normal lens differentiation and

The crystallins

Three types of crystallin have been identified in mammals, α, β and γ, mainly on the basis of molecular weight (eFig. 4.24). The δ crystallins have also been detected in birds. Several other 'taxon-specific' crystallins (ε, τ, ϱ, χ, μ, λ, ζ, SIII) are recognized in other species, based on the criterion that they account for at least 10% of the total water-soluble protein. More recent data suggest that in vertebrates there are only two classes of crystallins, α and γβ. Native α crystallin is of two types, αA and αB, each with a molecular weight of about 20 kDa. In the native state, however, the α crystallins form large multimeric aggregates of 300–1200 kDa (average 800 kDa), held together by non-covalent interactions.

β Crystallins range in molecular weight from 23 to 35 kDa and occur in several subtypes: βB_1, βB_2 and βB_3; βA_2, βA_3 and βB_4. Mixed aggregates of between 50 and 200 kDa occur naturally. The γ crystallins are monomeric in the native state; there are six types (*γA–E, and γS (formerly known as βS)), differentiated by charge. Not all types are present in human lenses at all ages, some such as γS and γC being present at higher concentrations in fetal than in adult lenses. The relative amounts of α:β:γ crystallin also vary greatly depending on age and other factors; in the 'typical' lens the ratio of α:β:γ is of the order 40:35:25. Protein sequence analysis has shown homology between the β and γ crystallins; in addition, all three proteins exist as β-pleated sheets.



eFig. 4.24 The structure of the human lens (**A**) and the major soluble lens crystallins (α, **B**), (β, **C**) and (γ, **D**) is shown. (From Moreau and King, 2012.)

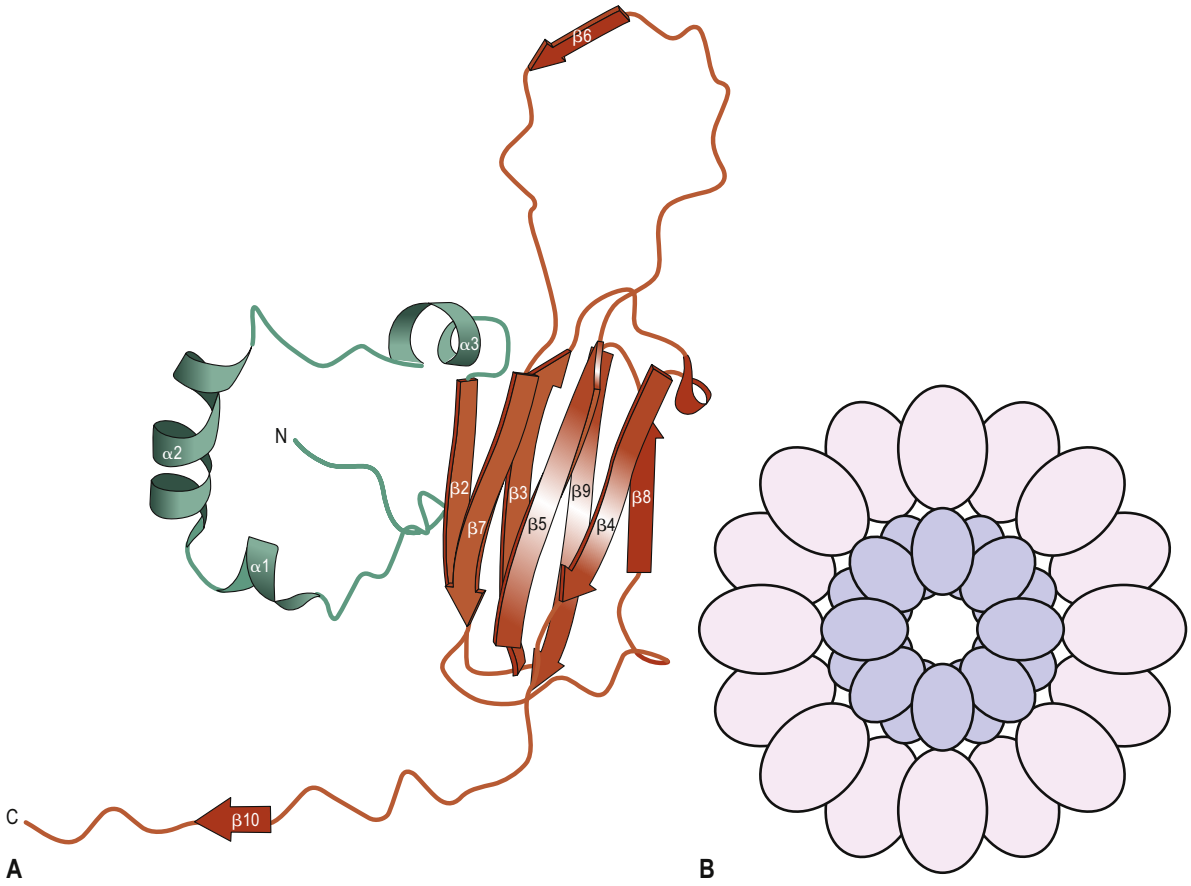


Fig. 4.64 (**A**) Secondary structure of the Hsp16.9 subunit with an ordered N-terminus. The N-terminal domain *(green)* contains three helical segments, as shown, in half of the subunits. The remaining subunits in Hsp16.9 and all subunits in Hsp16.5 have unstructured N-termini. The α crystallin domain *(brown)* consists of seven-stranded β-sandwich, an interdomain loop containing one β-strand *(*β6 *at the top)* and a C-terminal extension *(at the bottom)*, which is largely unstructured except for the short V10-strand. (**B**) A possible micelle-like structure for α crystallin. The subunits contain two domains and assemble into large aggregates through interactions between their hydrophobic N-terminal domains *(lilac)*, which are located in the centre of the aggregate. The hydrophilic C-terminal (α crystallin) domains *(*pink*)* are on the surface of the assembly. (**A** is reproduced with permission from Nature Publishing Group. **B** is from Augusteyn, 2004, with permission from the Optometrists Association of Australia.)

transparency, while αB, which is expressed in neural tissue and upregulated under conditions of stress, is not essential for lens transparency.

The crystal structure of some small invertebrate heat-shock proteins such as Hsp16.5 has allowed a model for lens αA crystallin micelles to be developed (Fig. 4.64). In this model it has been shown that neither αA nor αB is necessary to develop the molecular arrangement but that some combination of either will suffice.

Phosphorylation of the αA_2 and the αB_2 chains produces the αA_1 and the αB_1 chains. Spontaneous non-enzymatic cleavage of the molecules also occurs, as does high molecular weight aggregation, especially with age. The α crystallins, acting as molecular chaperones, 'trap' other crystallins and proteins such as intermediate filaments, which may be undergoing denaturation and unfolding. In this way they maintain lens transparency by preventing disruption of the highly ordered structure of the crystallin packing. Recombinant human crystallin has now been prepared for study of its chaperone activity. For instance, in transgenic zebrafish different human crystalline mutations have been linked to different cataract phenotypes (Wu, 2018).

The γβ crystallins are thought to have a similar structure: four repeating antiparallel β sheets in the form of 'Greek key' motifs. γ crystallin is a highly stable molecule, attributed to its extensive internal symmetry; mutations in congenital cataract and cataract of old age may be linked (Fig. 4.65).

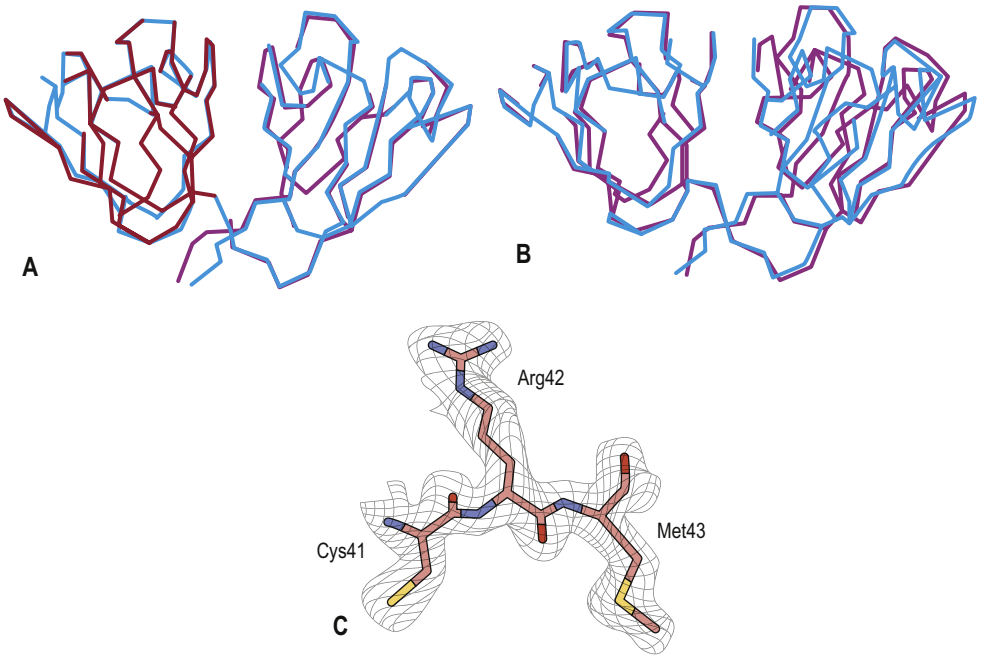


Fig. 4.65 Crystal structure of human W42R γD-crystallin, a mutant form of human γ crystallin which has only minor differences to normal γ crystallin but is protease-sensitive and is associated with certain forms of congenital cataract. (**A**) Superposition of the normal and mutated forms of γ crystallin. (**B**) Best fit superposition. (**C**) Electron density of residue Arg42 mutated from Trp42 in the normal γD crystallin. (From Ji et al., 2013.)

Multimeric complexes of β crystallins tend to form between the acidic molecules (βA_2, βA_3 and βA_4) and the basic molecules (βB_1, βB_2 and βB_3), followed by association between similar heterodimers. Homology among the various β crystallins both within and between species is quite variable and sequence analysis is still in progress for most of the human proteins. However, on the basis of γβ sequence homology a predicted structure of β crystallin has been suggested and X-ray crystallography has shown it to have some basis. In this model, the two γ-like structures are joined by a connecting peptide.

Although the molecular packing of the crystallins and their high RI contribute extensively to lens transparency, biophysical studies of molecular interactions between proteins indicate that other proteins that can adopt the correct state of phase transition and osmotic pressure and would perform this function equally well. Volume-based excluding properties of proteins are at their highest with γD crystallins, the most abundant crystallin in the lens. Multiple post-translational modifications occur over time to lens crystallins to ensure protein stability but eventually protein aggregation occurs and cataract ensues.

Cytoskeletal proteins of the lens

Cytoskeletal proteins are usually to be found in the urea-extractable fraction of lens proteins. In addition to ubiquitous microfilaments (actin, vimentin and spectrin) and intermediate filaments (see earlier section on cells and tissues), there are certain lens-specific intermediate filaments such as beaded filaments. Vimentin is the major intermediate filament in the lens cell and is present in epithelial and cortical fibre cells but not in nuclear fibre cells. A similar distribution has been found for microtubules in lens cells. Cytokeratins are not found in the adult lens. Some differences occur in relative proportions of cytoskeletal elements. Thus, talin, α-actinin and the signalling proteins are at high concentration in lens equatorial epithelium and vinculin is prominent in stable fibre cells with strong cell–cell contacts.

Beaded filaments are specific to the lens fibre. Two main species have been identified: beaded filament-specific proteins 1 and 2 (BFSP1, BFSP2, also known as filensin and CY49, respectively). In addition, there are link proteins, the plakins, which together are likely to be involved in crystallin packing and density distribution, perhaps by offering attachment sites for crystallin molecules. CP49 is also known as phakinin. Both proteins co-assemble with α crystallin (but not with vimentin) and are essential for lens transparency.

Lenses with targeted deletions of phakinin and filensin are opaque even though lens fibre morphology is normal,

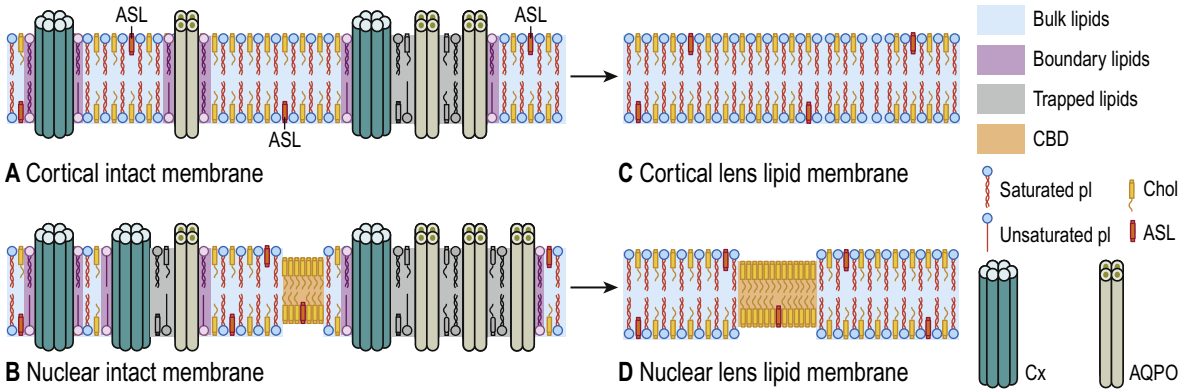


Fig. 4.66 Schematic drawing of the intact (**A** and **B**) and the lens lipid membrane (LLM) (**C** and **D**) porcine lens cortical (**A** and **C**) and nuclear (**B** and **C**) membranes. Lipid domains induced by the high cholesterol content and the presence of integral membrane proteins (mainly connexins and aquaporins) are indicated. The cholesterol analogue ASL is distributed between the cholesterol bilipid domain *(CBD)*, bulk lipids and trapped lipids. Note that cholesterol and ASL are excluded from boundary lipids. The nitroxide moiety of ASL is indicated by the *black dot*. Cortical LLMs (**C**) are formed by bulk lipids and do not contain CBDs. CBDs are not expected in cortical intact membranes (**A**). Nuclear LLMs are formed by coexisting bulk lipids and CBDs (**D**). CBDs are also expected to be present in nuclear intact membranes (**B**). (From Mainali et al., 2018.)

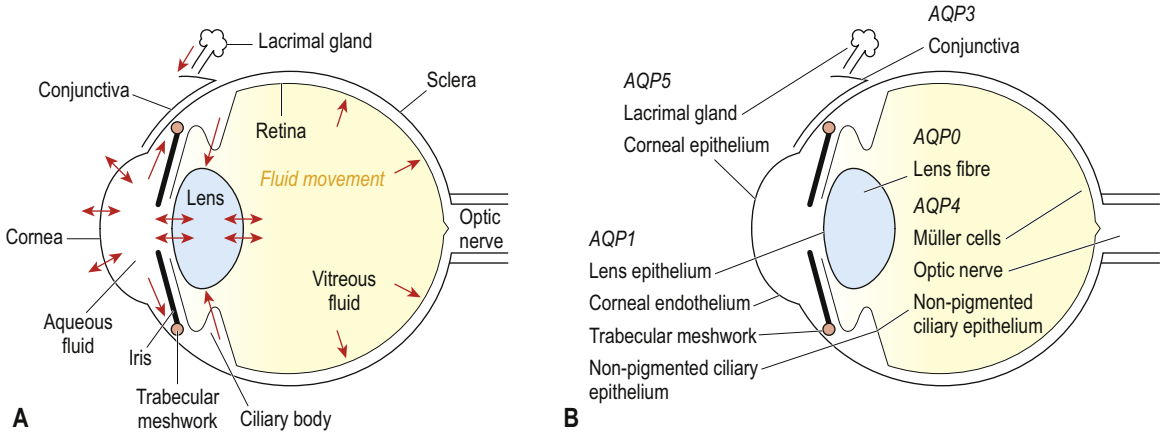


Fig. 4.67 Fluid transport and aquaporin (*AQP*) expression in the eye. (**A**) Routes of fluid movement showing secretion by lacrimal gland and ciliary body, absorption by trabecular meshwork and retinal pigment epithelium, and bidirectional movement in cornea and lens. (**B**) Sites of AQP water channel expression in ocular tissues. (From Verkman, 2003, with permission from Elsevier.)

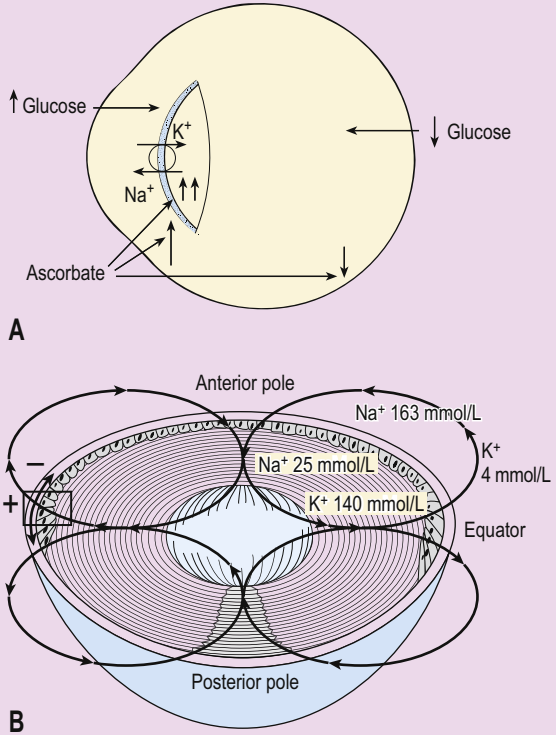
indicating that these two proteins, which co-assemble to form beaded filaments, are essential for lens transparency through lens fibre cytoskeletal organization.

Membrane lipids and proteins

Membrane lipids in lens cells are highly saturated, the main phospholipid being the extremely stable dihydrosphingomyelin. In lens cell membranes they are densely populated with proteins, thus restricting their mobility (Fig. 4.66). Cholesterol is also abundant, which in view of its exclusion from protein 'rafts' (see p. 166) leads to patches of pure cholesterol in the membrane (cholesterol bilipid domains). Lens cell membrane proteins are extractable from lens membranes in detergents such as sodium dodecyl sulphate (SDS), a major protein being lens fibre cell-specific junctional complex protein aquaporin-0 (AQP0), previously known as major lens intrinsic protein 26. In fact, this was the first aquaporin (or water channel) identified and they act as osmoreceptors or cell volume regulators. At least 100 of these genes have been described, 11 of which are present in mammalian systems and five involved in fluid transport through various ocular tissues in and out of the eye (Fig. 4.67).

Box 4.13 **Transport of molecules across the lens surface**

The lens behaves like a syncytium in which K^+ is transported into the lens and Na^+ is transported out via Na^+/K^+ ATPase present in the lens epithelium (**A**). Interestingly, the concentration of the ATPase is highest in the equatorial zone of the lens epithelium where much of the ionic transport occurs. This sets up electrical gradients with differential electrical potential differences between the lens equator and the lens poles, in part accounting for the electrical dipole that occurs in the lens (**B**). The lens also contains specific glucose transporters and transporter molecules for ascorbate and water, which ensure adequate metabolism and minimize free radical damage.



(Part B reproduced from McCaig et al., 2005.)

The *AQP0* gene is located on the cen-q14 region of the long arm of chromosome 12. The protein is not only involved in water transport but also in intercellular communication and movements of ions as part of the microcirculation of the lens. In the closed configuration it acts as an adhesive protein for the posterior lens capsule, maintaining the lens electrical dipole. In addition, it ensures fibre-to-fibre compaction via haemophilic electrostatic binding (Varadara et al., 2018). AqP0 is important for lens transparency; mutations in this gene causing cataracts. AqP0 is absent from lens epithelial cells and its relationship to gap junction proteins in other cells is unclear. However, AQP1 is present in lens epithelium.

Other membrane proteins include numerous enzymes such as ATPases and cytoskeletally attached proteins such as calpactin-1 and *N*-cadherin. Several other high molecular weight proteins also exist in the lens fibre plasma membrane such as plakoglobin, plectin, periplakin and desmocollin, members of the spectrin family with a role in maintaining cell shape as for spectrin in red cells. Lensin is a lens-specific cytoskeletal protein linking beaded filaments to cell membrane.

Extracellular matrix

The only extracellular matrix of any importance in the lens is the capsule. The capsule is constructed as for any epithelial cell basement membrane containing type IV collagen and heparan sulphate proteoglycan and acts as a diffusion barrier for the lens. Fibronectin is localized to the anterior capsule, while tenascin is present in the posterior capsule. Tenascin is one of a family of matricellular proteins that includes thrombospondin and SPARC (sialo-protein associated with rods and cones, see Fig. 4.25), the last being required for lens transparency. The $\alpha_5\beta_1$ integrin is present in the anterior lens epithelium, while the $\alpha_6\beta_1$ integrin receptor for laminin is present in equatorial and lens fibre cells, both of which are migratory.

Semipermeable membranes and physiology of the lens

As indicated previously, the lens behaves like a very large syncytium or single cell, both electrically and chemically. Active pumping mechanisms based on Na^+/K^+ ATPases exist to pump Na^+ ions out of the lens, while chloride and water are transported into the cell (see Box 4.13). The pump is located in the epithelium with the highest concentration at the lens equator, while lens fibres do not have a pump function. As a result, electrical current density across the epithelium is highest at the equator, generating a flow of current from the equator to the lens centre and this appears to be critical for intra-lens microcirculation (see Box 4.13). Current flow is supported by ion channels such as transient receptor potential vanilloid (TRPV)-4 and TRPV-1, which responds to hyposmotic stimuli (Mandal et al., 2018).

Barriers to ion and solute transport occur at the capsule and at the plasma membranes of the epithelial and fibre cells. The capsule is permeable to small molecular weight proteins (<50,000 Da), including low molecular weight crystallins, but prevents diffusion of large molecules.

At the epithelial barrier, the cells show the typical polarization of other epithelial cells but lack tight junctions at the lateral cell surface. Instead there is an extensive system of gap junctions, which permits rapid intercellular communication, thereby allowing the cells to behave as a syncytial sheet. At the junction between the epithelium and the fibre cell, the main transport mechanism is rapid endocytosis via coated vesicles, while the very extensive system of gap junctions between each lens fibre permits rapid interfibre cell movement of metabolites, amounting in effect to a microcirculation (see Figs 4.61 and 4.62). Indeed, 50% of the fibre cell plasma membrane protein consists of AQP0, while the gap junctions in lens are to some degree specialized (more efficient) than other cell types in that AQP0 appears to assist in channel formation between the cells.

The Na^+/K^+ ATPase pump in the epithelium actively exchanges Na^+ (pumped out) for K^+ (pumped in). The Na^+ passively diffuses anteriorly down a concentration gradient from the vitreous body, across the posterior lens capsule into the lens body, where it rapidly diffuses to the anterior epithelium and is pumped out into the aqueous. K^+ ions are handled in the reverse direction, eventually diffusing passively across the posterior capsule into the vitreous. Outward currents driven by the ATPase pump at the epithelium, particularly the equatorial epithelium, balance inward currents at the anterior and posterior poles into the extracellular space. Critical to this process is the presence of K^+ channels in the epithelium but not in the lens fibres, which allow an overall Na^+/K^+ exchange.

While this simple pump–leak model serves to explain ion transport across the lens body, there are several unanswered questions, such as the mechanism of Ca^{2+}/Mg^{2+} transport, which also occurs via a specific ATPase. The Ca^{2+}/Mg^{2+} ATPase is most abundant in the lens cortex. In addition, specific transporter proteins for glucose (GLUT1 in epithelium, a Na^+-dependent SGLT transporter in lens fibres) and amino acids exist in the plasma membrane of lens fibre and epithelial cells.

Transportation of water across the lens has also been shown to occur in an anteroposterior direction at a rate of about 10 μL/h, probably involving AQP0, thus participating in the overall lens microcirculation and helping to rid the lens of waste products and maintain transparency.

Lens metabolism

Carbohydrate

Glucose from the aqueous humour is the main source of energy for lens metabolism. Glucose enters the cell via an insulin-dependent glucose transporter located in the plasma membrane (GLUT1). Both the glycolytic and the pentose phosphate pathways are used and under conditions of excess glucose the sorbitol pathway is entrained (see Fig. 4.30). About 80% of glucose is consumed by the lens via glycolysis. The pentose phosphate pathway uses about 10% of the remaining glucose, providing sugar residues for nucleotide synthesis. The citric acid cycle is active in the lens epithelium, the only lens cells to possess mitochondria, mostly at the equator. The epithelium also expresses most of the aldose reductase, where excess glucose handling via the sorbitol pathway might take place. Under normal circumstances, less than 5% of glucose is used up in the sorbitol pathway. Indeed, it is unlikely that sorbitol has a significant role in the induction of complications of high ambient glucose (diabetes). Aldose reductase is apparently induced by osmoreceptors (such as aquaporins) and interferes with NADH-binding proteins (Fig. 4.68) thus disturbing regulatory free radical scavenger mechanisms (see later). In addition, aldose reductase has $PGF_{2\alpha}$-like synthase activity, is involved in the COX2 pathway and is regulated by IL-1β. Thus, it has a more direct role in inflammatory mechanisms, which underpin the complications of diabetes.

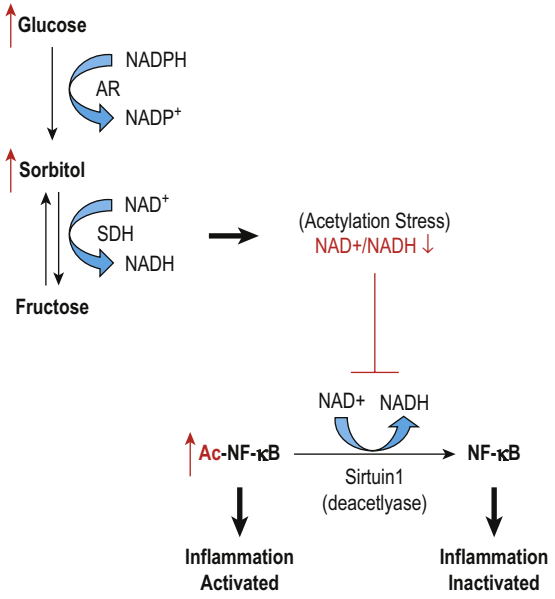


Fig. 4.68 Increased flux of polyol pathway initiates oxidative stress by reducing the ratio of NADPH/NADP+ and NAD+/NADH and inducing advanced glycation end products (AGEs). (From Chang and Petrash, 2018.)

Protein

Synthesis of new protein ceases with lens fibre cell formation and all changes that occur to lens proteins after this stage are post-translational modifications. Phosphorylation of many proteins occurs, including crystallins, cytoskeletal proteins and AQP0. Several phosphorylation

systems exist, including a cAMP-dependent protein kinase A and a phospholipid-dependent protein kinase C. Certain drugs enhance phosphorylation of intermediate filaments, including β-adrenergic compounds.

Numerous enzyme activities have been detected in lens protein extracts but the level of enzyme protein is very low. Some of the taxon-specific crystallins have apparent enzyme activity such as ε crystallin from duck lens (lactic dehydrogenase), ϱ crystallin from frog (aldose reductase) and ε crystallin from frog (lung prostaglandin F synthase). These findings are of evolutionary rather than physiological significance but because several of these enzymes are induced in cells undergoing stress it has been suggested that stress responses may be the common denominator in these homologies (see earlier, heat-shock protein and α crystallin). Stress proteins and long-lived lens crystallins may require similar properties to maintain stability and durability in anaerobic conditions. Thus, crystallins, especially the α and γβ series, are among the most conserved proteins known and interestingly are not restricted to the lens (αβ has been found in heart, lung, brain and retina). The promoter sequence of the αA gene has been shown to be lens-specific and has the capability of driving foreign genes selectively into its sequence. This might explain the interchangeability of function for apparently identical proteins from widely divergent sources. Similarly, the promoters for χ crystallins are lens-specific. During development and growth there is differential expression of the various crystallin genes in a highly regulated manner.

In addition to the evidence for enzyme activity in the crystallin proteins, the lens has several other proteolytic enzymes, including endo- and exo-peptidase activity and membrane-associated proteases. Historically, the denucleation of the lens fibre has been described as a form of attenuated apoptosis, taking place over several days rather than hours. In addition, some of the 'death' enzymes such as caspases are activated during this process, although mice deficient in caspases 3 and 6 have normal lenses. In addition, a key process in apoptosis, namely the phosphatidylserine 'inside-out' membrane flip, does not occur. Despite these caveats, it is likely that much of the signalling machinery involved in apoptosis and lens fibre differentiation overlaps in function. Interestingly, although autophagy does not play a part in this process (Atg5 –/– mice have normal lenses) mitophagy is active in the removal of organelles from lens fibres and is mediated by Bcl2 interacting protein 3-like (BNIP3L/NIX) (Brennan, et al., 2018). Bcl2 is central to apoptosis.

The neutral endopeptidases calpain I and II (cysteine Ca^{2+}-dependent enzymes) and their inhibitors, calpactins, occur in lens cells and are involved in cytoskeletal proteins and crystallins turnover. Calpain I, a possible factor in age-related cataract, is present in the epithelium and lens cortex but not in the nucleus.

Increased degradation of proteins occurs with age, particularly that of AQP0 /MIP26, linked to loss of the small 8.5 kDa ubiquitin chaperone (see p. 175.e1-e3).

Lipid

The unusually high concentrations of sphingomyelin, cholesterol and saturated fatty acids impart rigidity to the cell membrane (see earlier text), which may be important in maintaining intercellular connections. In particular, the cholesterol bilipid domains ensure a high content of cholesterol in the neighbouring regions, thus maintaining stability of the membrane even if there are significant changes in the membrane phospholipids (Fig. 4.66). The high levels of phosphatidylinositol suggest significant receptor-mediated second messenger activity in lens cells.

Redox systems in the lens microenvironment

The lens is constantly exposed to attack by oxidizing agents; the high levels of H_2O_2 and peroxidase activity in normal aqueous humour are also present in the lens. Several enzyme systems are available to minimize or to buffer the effects of oxidants, including catalase, superoxide dismutase, glutathione peroxidase and glutathione *S*-transferase. The lens contains high levels of glutathione (3.5–5.5 μmol/g wet weight), with the highest concentration in the epithelium, and detoxification via the mercapturic acid pathway is an important pathway in the lens. Glutathione is produced from the interaction between glutamate and cysteine in lens cells. Glutathione is thought to diffuse towards the centre of the lens, where lens fibre cells are essentially non-metabolizing and thus unable to generate glutathione. Flow in this direction would be counter to the proposed direction of flow in microcirculation in the lens, which is dependent on ion and water transport and underpins current flow and the lens dipole measurements (see Box 4.13). An alternative connexin-based/gap junction route for delivery of glutathione to the lens nucleus from the cortex has been proposed.

Glutathione is also important in protecting thiol groups in proteins, especially cation-transporting membrane proteins in the lens, which additionally accounts for its unusually high concentration in this tissue. More than 95% of glutathione is in the reduced state.

Catalase and low levels of superoxide dismutase have also been identified in lens epithelium, indicating that these systems are also probably important.

Ageing in the lens and cataract formation

The transmission of light decreases with age, especially for the lower wavelengths (up to a factor of 10), to the point

that at low levels of illumination an apparent tritanopia can occur (see Chapter 5, p. 316). Morphologically, the cells lose cytoskeletal organization and develop vacuolation and electron-dense bodies. An increase in sodium concentration is accompanied by a decline in the membrane potential of the lens, suggesting ion channel dysfunction. Enzymatic activities decline in the lens nucleus but not in the cortex or epithelium. In addition, the appearance of water clefts in the lens as an early sign of cataract suggests decreased function of AQP0 (MIP26) and fluid transport.

Ageing of the lens and cataract formation are not synonymous. In age-related nuclear cataract, there is extensive oxidation of cystine and methionine residues on lens proteins, while in aged lenses without cataract, oxidation is much less. Glutathione SH (oxidized glutathione) (see Box 4.6) is the key. Post-translational modification of lens proteins continues throughout life. In addition to cross-linking and degradation, which occur in any stable protein system, non-enzymatic glycation is a conspicuous event. In general, γD crystallins are synthesized in young lenses, while production of γσ and β crystallins increases with age. In addition, most of the α crystallin is lost from the water-soluble compartment to the water-insoluble compartment, as are some of the β and γ forms.

Non-enzymatic glycation of crystallins occurs at the ε amino groups of lysine, especially the high molecular weight aggregates of α crystallin. *In vitro*, this reaction produces a yellow fluorescent pigment similar to that seen in the ageing human lens. Interaction between various amino groups and aldehydes released from free radicals, especially through lipid, (per)oxidation produces fluorophores and ceroid/lipofuscin. In spite of the colour changes, the amount of protein that is glycated is less than 5% in an aged lens, which is considerably less than for other long-lived proteins such as haemoglobin and collagen. Lens crystallin glycation is more likely to be the result of its interaction with oxidized ascorbic acid than glucose on the basis of intra-lens concentrations, and it is possible that glutathione, by maintaining ascorbic acid in its reduced state, inhibits this process.

AQP0 also undergoes modification with age, losing a 5 kDa peptide to become MIP22 in increasing concentration. Cleavage occurs at both the C- and N-terminal ends of the molecule.

Reduced vision in cataract is caused by increased light scatter by lens proteins

The transmission of light by the lens is reduced when the ordered packing of the lens crystallins is disturbed. This can be induced in many ways, such as increased water accumulation within the lens, formation of high molecular weight lens protein aggregates and vacuole formation within the lens fibres with age.

Certain metabolic conditions are associated with cataract, the best known being the cataract of diabetes and a similar lens opacity in galactosaemia. In these forms of 'sugar' cataract, accumulation of water in the lens fibres was previously thought to result from the accumulation of non-degradable polyols such as sorbitol and galactitol in the lens fibre cells. High glucose/galactose concentrations in the aqueous humour lead to increased intracellular accumulation of glucose, which saturates the normal glycolytic pathways. Accordingly, aldose reductase converts excess glucose to polyols (such as sorbitol), which accumulate in the cells, thereby increasing the osmotic drag of water into the cell via activation of the osmoreceptor AQP0. However, aldose reductase also promotes $PGF_{2\alpha}$ synthesis and inflammation via activation of NFkB (see p. 247) with production of ROS, indicating that multiple pathways are likely to be involved.

Whatever the mechanism, the dysregulation in cellular metabolism, with reduction in the levels of cellular ATP and glutathione, and secondary damage to the cell, plus the increased water content causes phase separation between protein-rich and protein-poor regions of the cells and increased light scatter (cataract). As individual cells loosen their interdigitations with neighbouring cells, water clefts and vacuoles appear within the lens substance. As cells die, there is progressive increase in opacification, which in the lens cortex is seen as 'spoke-like' opacity and in the nucleus is characterized by the accumulation of insoluble protein aggregates and chromophores, causing the nucleus to change colour from yellow to red to black (cataracta brunescens/cataracta nigra). Some slight beneficial effect in protecting the macula might occur in the early phases of cataract formation as a result of the yellow filter effect blocking transmission of low wavelength light.

Cataract formation is caused by any insult to the lens

Since the lens is designed for the transmission of light, it responds to any insult that disturbs normal development or metabolism by opacification, even if this is only for a temporary period. Thus, certain congenital cataracts appear to affect only the fetal nucleus; radiation cataract may be limited if only a discrete area of the lens bow region is affected; the sunflower cataract of trauma may be the result of shearing forces momentarily separating lens fibre cells, which then restore their interconnections; and certain forms of cataract such as the 'feather' cataract after vitrectomy are thought to be caused by large volumes of fluid transfused through the vitreous cavity at too low a temperature or an incorrect electrolyte composition. Interestingly, such cataracts may be reversible, suggesting that the lens microcirculation is temporarily unable to cope with the increased fluid load but can recover when bulk fluid flow is normalized. A reversible 'osmotic' cataract also occurs in neonatal/infant acute onset

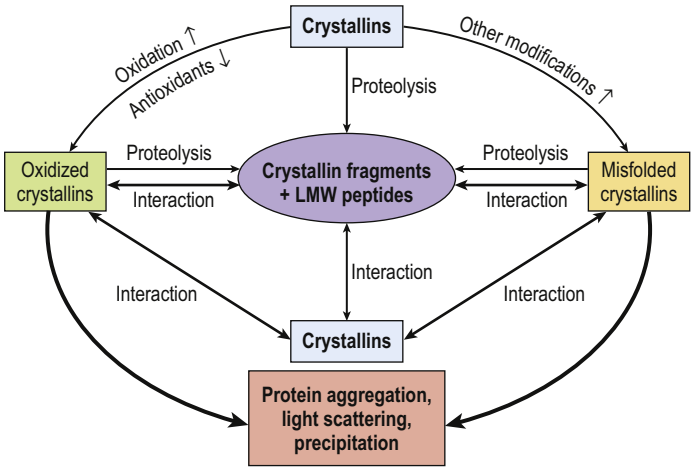


Fig. 4.69 Schematic representation of the proposed role of crystallin fragments in ageing of the lens and development of cataract. (From Sharma and Santhoshkumar, 2009.)

diabetes. Similarly, 'cold' cataract is induced in young animals and is caused by the reversible precipitation of γ crystallins by phase separation in the fetal nucleus.

Many forms of cataract have been attributed to specific mutations in lens crystallins probably related to impaired molecular packing as well as several other proteins. For instance, a mutation in the Wolfram gene (*WFS1*) has been identified as the cause of congenital nuclear cataract while another gene, the *FCO1* gene, has also been implicated in a recessive form of congenital cataract.

Age-related cataract formation is multifactorial

A wealth of studies during the past 20 years has shown many biochemical changes associated with age-related cataract. In summary, these include an increase in the insoluble components of the lens, an increase in chromophores, increased protein cross-linking and aggregation and oxidation of amino acid groups. There is a concomitant decrease in anti-oxidant enzyme systems and increased proteolytic activity. The level of glutathione is also reduced. Normal lenses contain a trypsin inhibitory activity, which may regulate age-related proteolytic activity. The major protein change in cataractous lens is the loss of αA crystallin and the selective loss of γS crystallin. In addition, there are numerous degradation peptides detectable in the water-soluble component. These mechanisms are summarized in Fig. 4.69.

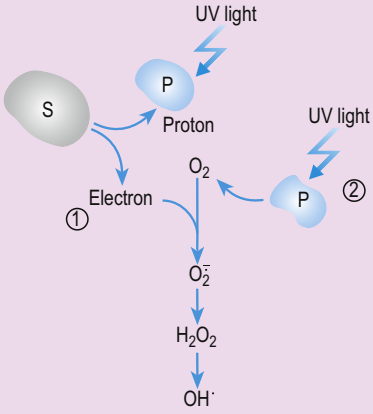
The possibility that UV light might cause or hasten some of the effects of age has been suggested by the observation that age changes appear to be more marked in the region of the visual axis than in the equatorial region of the lens. Current views are that oxidative events are the most likely mechanism of cataract formation. Near-UV light is absorbed by tryptophan, which in sunlight is converted to *N*-formyl-kynurenine, a fluorescent chromophore similar to 3-hydroxy-kynurenine, a second UV absorbent molecule in the lens. Both these compounds can act as photosensitizers and lead to the production of the free radical singlet oxygen (see Boxes 4.6 and 4.14). Free radicals downregulate the function of critical lens enzymes such as Na^+/K^+ ATPase and lead to lens swelling and opacification in the rats. Free radicals generated by near-UV light such as hydrogen peroxide impair hexokinase function, diverting more glucose to the aldose reductase pathway and further ROS production (see earlier). Oxygen increases the rate of photo-oxidation, and vitamin E, ascorbic acid and glutathione reduce the effects of light damage (effects of oxidation and protein misfolding are summarized in Fig. 4.69).

The role of UV light in human cataract is unclear, although exposure to UV-B is reportedly associated with an increase in cortical and posterior subcapsular cataract but not in nuclear cataract formation. UV-B can also produce fibril formation in crystallins *in vitro*. Interestingly, aged human lenses appear to absorb more UV-A and even visible light than young lenses.

Certain trace metals and compounds are associated with cataracts. Experimental depletion or excess of selenite leads to cataracts by a mechanism linked to Ca^{2+} homeostasis. In contrast, cyanate induces carbamylation of lens proteins and cataracts, which can be prevented experimentally with aspirin. Interestingly, aspirin usage may delay the onset of cataracts in humans. Inhibition of cholesterol synthesis also leads to cataracts in experimental animals. Metabolites and chemicals such as pyruvate and caffeine

Box 4.14 **Free radical damage and the lens**

Free radical damage in the lens may occur through oxidative metabolism but is considered mostly to be the result of UV damage by activation of endogenous photosensitizers.



Photosensitizers act by one of two mechanisms: (1) triplet sensitizer absorbs a proton from a substrate, leading to the production of free radicals (SO_2 and H_2O_2); (2) sensitizer reacts with O_2, leading to singlet O_2. Typical photosensitizers include riboflavin, tryptophan and kynurenine (all present in the lens).

have been shown to prevent effects of UV and selenium *in vitro*, respectively.

In summary, glutathione consumption with reduced glutathione reserves and unrestricted oxidation of intrinsic lens proteins, i.e. a disrupted REDOX balance underpin cataractogenesis.

Mechanism of age-related cataract formation: a failure of chaperone function

How does a disturbed REDOX balance cause cataracts? Oxidation of intrinsic lens proteins (as well as other post-translational modifications (PTMs) of βγ crystallins) lead to their progressive dysfunction. As they denature and precipitate, the α crystallins bind the unfolded proteins, but unlike true chaperones they do not have the ability to refold βγ crystallins or other lens proteins. The amino acid Lys-315 is essential in determining reversible protein folding activity in δ-crystallin but not α-crystallin (Huang et al., 2016). As a result, the chaperone capacity of the α crystallins is consumed and the complexes precipitate within the lens fibre cells, forming the insoluble protein fraction that increases with age.

The vitreous body

The vitreous body is a true connective tissue containing collagen, gags and cells

The vitreous body is 98% water, 1.0% macromolecules and the rest is solutes and low molecular weight materials.

The matrix

The vitreous transmits light by the same mechanism as the cornea, i.e. its collagen fibrils (10–20 nm) are thinner than half the wavelength of light, and the interfibrillar space is filled with GAGs (hyaluronan) at intervals that reduce the effects of diffraction in the system. Collagen imparts the gel structure to the vitreous body, is predominantly type II, similar but not identical to cartilage type II collagen (vitreous collagen has more galactosyl-glucose side chains and a higher content of alanine), and is arranged in a lattice structure in which the fibrils are suspended in a viscous hyaluronan solution. This structure is lost with age and in disease by a process known as syneresis in which the hyaluronan molecules are degraded to smaller moieties and the collagen fibrils coagment to form larger fibrils, becoming visible as 'floaters'. Some type V and type IX collagens are also present and play structural roles in gel formation as linker proteins. In addition, a hybrid molecule composed of type V/XI α chains has been detected in vitreous body in trace amounts and has been implicated in vitreous fibril formation as for cornea. In cartilage, collagen type II fibrils exist in two forms, thin (~16 nm) and thick (~40 nm) and collagen type XI is found exclusively in the thin fibrils (Fig. 4.70). Presumably collagen type XI has a similar role in the vitreous body in regulating fibril diameter.

There are few intrinsic vitreous proteins. Opticin, a member of the small leucine-rich repeat (SLR) extracellular matrix protein family, is present in the vitreous body and also in ligament, skin and retina. It also has a role in regulating fibril thickness in the vitreous body similar to collagen type IX and V/XIα proteoglycans. Importantly, opticin has anti-angiogenic properties.

The normal vitreous body in young adults has a distinct architecture (see Chapter 1, p. 37). The cortex has a higher concentration of collagen and hyaluronan than the central vitreous body and, in addition, the cortex contains other GAGs such as chondroitin sulphate (3.5% total GAGs) and heparan sulphate (0.3%), which may be important in vitreo/retinal apposition.

In the central vitreous gel, hyaluronan is essentially the sole GAG. Hyaluronan occurs as stiff, open coil disaccharide chains, which in solution become entangled at

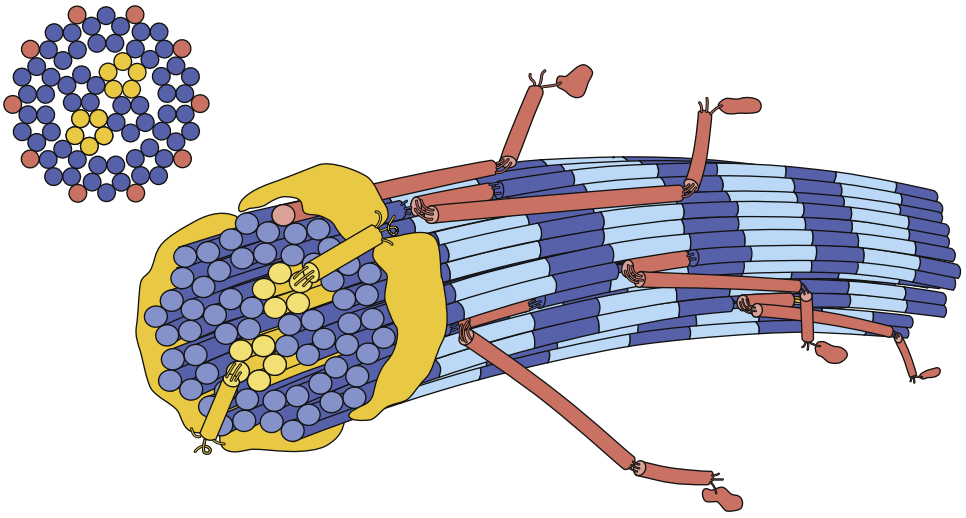


Fig. 4.70 Schematic representation of the 10 + 4 structure of a type II collagen microfibril, two pairs of collagen type XI forming a central core surrounded by 10 type II microfibrils. Vitreous collagen has a very similar arrangement of type XI/IX/II assembly with fewer cross-links. (From Kadler et al., 2008.)

concentrations above 300 μg/mL and thus add support to the gel matrix. Hyaluronan concentrations in human vitreous body vary between 100 and 400 μg/mL and the molecule binds to type IX collagen, which acts as a proteoglycan to bind the hyaluronan to the collagen fibrils. However, hyaluronan is not essential for maintaining the gel structure of the vitreous body. Therefore, some long-range interactions mediated via type IX collagen are necessary. Type IX collagen is a small non-fibrillar type of collagen that contains several non-collagenous domains; these act as the proteoglycan bridges. In addition, the chondroitin sulphate proteoglycan is present in the vitreous body at a concentration of approximately equal ratio with type IX collagen. Traces of chondroitin sulphate also comprise part of the proteoglycan versican in the vitreous body with a molecular weight of 2×10^4 to 4×10^4 Da. Hyaluronan is highly polydisperse (variable molecular sizes) and its breakdown products can activate innate immune cells (see Chapter 7, p. 385). However, as previously indicated, depolymerization or even loss of hyaluronan does not itself destroy the vitreous gel structure.

Vitreous cells

The vitreous body contains a single monolayer of cells (hyalocytes), which line the adult vitreous cortex and are responsible for production of hyaluronan in the gel. However, there is no regeneration of collagen in the vitreous body and thus there is no reconstitution of the gel after syneresis.

Hyalocytes are of two types: fibrocyte-like and macrophage-like. The latter are equivalent to the resident myeloid cells present in most non-CNS tissues. The role of hyalocytes in health and disease remains obscure. However, they do increase in number with age and have been implicated in macular hole formation (see Chapter 9).

Physicochemical properties of the vitreous gel

The viscoelasticity of the vitreous protects the retina during eye movement and deformations of the globe

The vitreous gel is non-compressible but highly viscoelastic. Thus, it responds to deformations of the globe by altering its shape to comply with external forces, but permits rapid restoration of global architecture. In this respect it behaves as a shock absorber, similar to synovial fluid, which also has a very high content of hyaluronan. These properties of the vitreous body are the result of its matrix structure, particularly its content of high molecular weight hyaluronan. This molecule has a very large hydrodynamic volume and at the concentrations present in the human vitreous body completely fills the interfibrillar spaces. Deformability of the vitreous body has mostly been measured experimentally *in vitro*. However, MRI in human volunteers revealed that even on eye movement there is considerable deformability of the vitreous gel, which may have relevance for traction events on the retina and risk of retinal detachment (Fig. 4.71). With age, this shock-absorbing property declines as syneresis takes place (see p. 192).

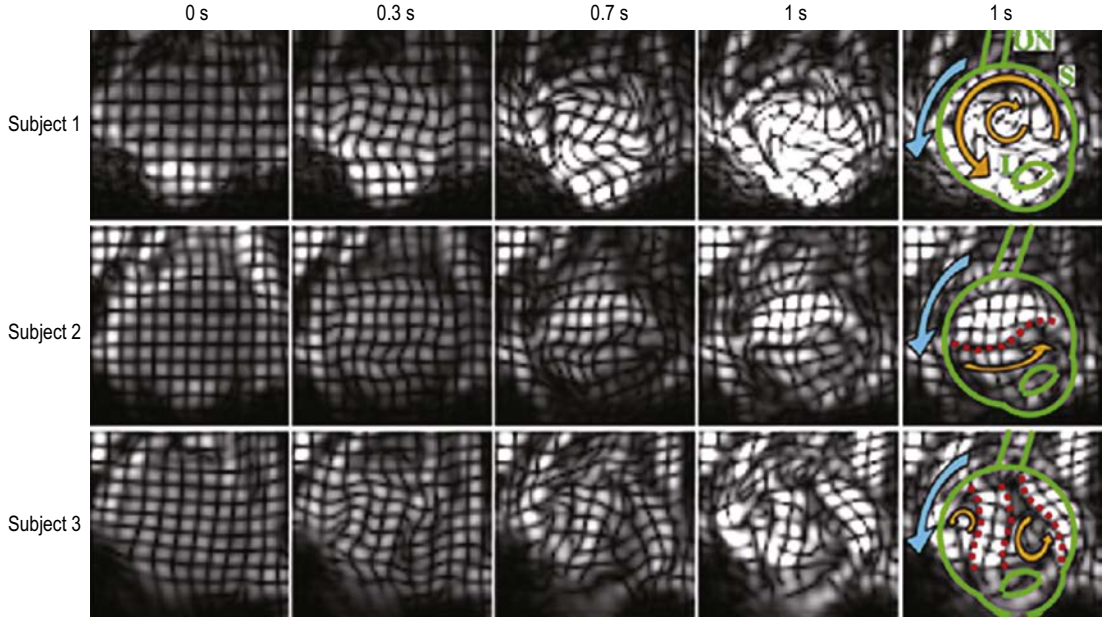


Fig. 4.71 The technique of complementary spatial modulation of magnetization was used to show the deformation of the vitreous gel during adduction movement of the eye. These changes are shown in frames 1, 5, 10 and 15 and are overlaid with sketches demonstrating the deformation in frames 5, 10 and 15. Differences in the deformations were observed for individual subjects as shown with subject 1 showing homogeneous viscoelasticity while strong 'whirling' movements were observed in subject 3. (From Piccirelli et al., 2012.)

Vitreous retards bulk flow of fluid and diffusion of small molecules

The flow of fluid through solutions of GAGs is variably retarded, depending on the nature and molecular weight of the GAG, and is greatest with hyaluronan. Flow of aqueous from the posterior chamber through the gel towards the retina is therefore slower in young eyes with formed vitreous gel than in older eyes where the vitreous gel has undergone liquefaction. Diffusion of small molecules such as glucose is also retarded by the hyaluronan in the vitreous. Electrolytes are more affected in their transvitreal transport by electrostatic interactions with hyaluronan, which is a polymeric polyelectrolyte (see p. 193). However, there is potentially more rapid percolation of fluid in the potential space between vitreous cortex and the retina depending on how firmly the cortex is attached to the retina.

The vitreous body is important for maintenance of lens clarity (sooner or later the lens develops cataracts after vitrectomy) and this has been attributed to the greater access of oxygen and thus the increased risk of oxidation events due to reactive oxygen species (Fig. 4.72). Transportation of fluid and electrolytes in a posterior direction across the retina is an important mechanism in the process of retinal apposition to the RPE (see next section).

The retina

The neural retina is highly organized in layers

The retina has two components, the neural retina and the RPE (see Chapter 1, p. 37).

Metabolic function in the retina

The retina's metabolism correlates with its blood supply: the outer retina, comprising the photoreceptors and the RPE, has high metabolic activity and receives most of its nutrients from vessels in the choroid, while the metabolism of the inner retina is serviced by the retinal circulation and is much less dependent on high energy supplies.

Glucose metabolism in the retina

Despite having the highest rate of aerobic glucose consumption of any tissue, a large proportion of the glucose utilized in the retina is converted to lactate. Lactic acid production, oxygen utilization and glucose consumption are also highest in the presence of CO_2/ bicarbonate buffering

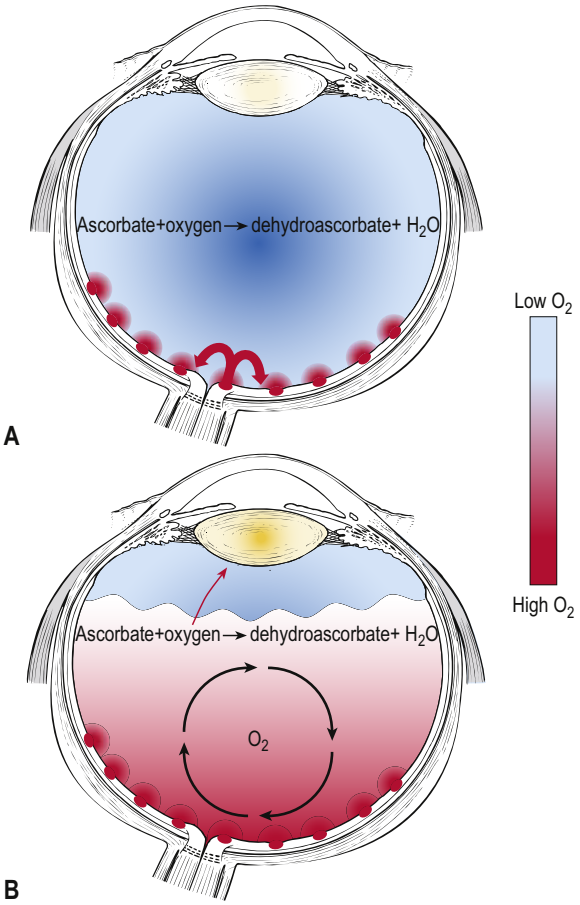


Fig. 4.72 The distribution of oxygen (**A**) in the normal eye and (**B**) after degeneration or removal of the vitreous body. Oxygen normally diffuses from the retinal vessels into the vitreous body but most of the oxygen has been consumed by retinal tissue and the concentration declines towards the central region of the gel. In degenerate vitreous body or vitrectomized eyes, more oxygen enters the fluid-filled vitreous cavity and combines with ascorbate in addition to diffusing to the lens where it is relatively toxic and can promote nuclear cataract through oxidative damage. (From Beebe et al., 2011.)

systems, suggesting a role for carbonic anhydrases in the retina. Most of the glucose utilization in the retina is taken by the photoreceptors (>80%). Sodium-coupled monocarboxylate transporters also occur in the retina, in astrocytes and Müller cells, where amongst other duties, they function in 'lactate shuttling' between glycolytic and oxidative pathways of glucose metabolism.

The retina can also metabolize other substrates for (ATP) energy stores such as glutamate, glutamic acid, malate and succinate. Retinal glucose is used to produce glutathione via the pentose phosphate pathway, which can be upregulated under conditions of oxidative stress. However, NADPH for glutathione stores is also produced by other non-pentose-dependent systems involving malate and isocitrate. Glucose, glutathione and oxygen are all required for generation of electrical activity, including the ATPase-dependent 'dark currents' (see Chapter 5, Box 5.6) in the retina.

The retina is regarded as an insulin-independent tissue, i.e. glucose is transported across the GLUT4-negative GLUT1- and 3-positive retinal endothelial for uptake by retinal cells. Glucose transport occurs by facilitated diffusion via GLUT 1 and GLUT 3 transporter proteins, similar to glucose transport into the brain. However, the retinal expresses insulin receptors and GLUT4, in addition to GLUTs 1–3, is expressed in all retinal layers, particularly the photoreceptors. Photoreceptors respond to insulin via a retina-specific insulin receptor, which is similar to brain insulin receptor in that it exists in a 'tonic' state of activity and does not change in conditions of fasting or excess glucose. The insulin receptor phosphorylates cyclic GMP-gated channels, which is central to phototransduction (see later). Photoreceptors, however, depend on GLUT1 in the RPE to furnish their supplies of glucose (Swarup et al., 2019). In addition, the pyruvate kinase isomer PK-M2, which is active in highly glycolytic processes such as tumours, is also active in photoreceptors where it promotes photoreceptor health and viability. Photoreceptor glucose metabolism influences normal retinal vascularization through the fatty acid hormone adiponectin.

Retinal neurones and photoreceptors are the most energy-demanding cells in the body mainly because of the need to maintain neuronal polarization (essential for generation of dark currents, see later) by ionic pumps requiring large amounts of ATP and use oxidative phosphorylation more than glycolytic pathways to generate ATP. However, in addition to glucose, retinal neurones utilize glycolysis-generated lactate and pyruvate from Müller cells as substrates from ATP production. Photoreceptors in turn release glutamate, which is taken up and metabolized by Müller cells ('lactate shuttling'). Müller cells act as energy stores, as well as 'managing' lactate production, since they contain high levels of glycogen, especially in species that lack a retinal blood supply. Müller cells also respond to metabolic stress by invoking the unfolded protein response (UPR) (see p. 197) involving HIF1a and GLUTs 1 and 2. Dysregulation of the UPR impacts on retinal VEGF production and neovascularization, for instance in diabetes.

Protein metabolism in the retina

Many of the neurotransmitters required for normal retinal cell function (see p. 271) are generated as free amino acids during glucose metabolism in the citric acid cycle; the non-neurotransmitter taurine, the most abundant amino acid

in the retina, is an agonist for the neurotransmitter GABA/glycine receptors and may have a neuroprotective role in the retina (Hadj-Said, 2017). Taurine (and GABA) also enter the retina through the same transporter (TAUT). Interestingly, the bile acid tauroursodeoxycholic acid (TUDCA), a conjugate of ursodeoxycholic acid (UDCA) and chenodeoxycholic acid, may prevent retinal degeneration. Entry of some amino acids to the retina is linked to ion (Na+) transport but some such as glutamate are toxic and require conversion to glutamine by glutamine transferase (synthase) in Müller cells. Ready uptake of amino acids is essential not only for the supply of neurotransmitters but also for amino acids involved in other physiological processes such as arginine, which is necessary for the synthesis of NO, the major regulator of endothelial cell function. Protein synthesis, as studied by methods such as leucine incorporation, is most active in the photoreceptors during such processes as photoreceptor renewal (see Box 4.15).

In addition to retina-specific proteins (see later), several proteins common to many tissues are present. Müller cells secrete the ECM that is the internal limiting membrane, a typical collagen/laminin-based extracellular matrix. The interphotoreceptor matrix (see p. 259) contains fibronectin and proteoglycans, which are also widely distributed throughout the retina. Tenascin-C and Tenascin-R are present in the neuro-glial extracellular matrix and are expressed strongly during retinal development and repair where they may have an immunomodulatory role. Tn-C and Tn-R have opposing effects in retinal synapse remodelling.

Several growth factors are present in the retina, such as insulin-like growth factor 1 (IGF-1) and acidic and basic fibroblast growth factor (bFGF). Basic FGF is present in vascular basement membranes and in the interphotoreceptor matrix layer, where it may play a trophic role in outer segment renewal. Similarly, VEGF can be induced in Müller cells and has a neurotrophic effect.

The retina contains a high content of lipid (20%)

The predominant lipids in the retina are the phospholipids, phosphatidylcholine and phosphatidyl-ethanolamine (comprising 80% of total lipid). There is also a high content of polyunsaturated fatty acids in the retina, especially in the outer segments, some containing more than six double bonds (known as 'supraenes'). This renders the retina particularly susceptible to oxidative damage (see p. 200). Disc membranes of photoreceptors are rich in phosphatidylethanolamine and the plasma membrane has higher levels of cholesterol, which inhibits rhodopsin activity. Thus, the older disc membranes at the tip of the photoreceptor (see Box 4.15) have lower levels of cholesterol than the fresh, newly formed discs, allowing easier activation of rhodopsin at the photoreceptor tip.

Lipid metabolism is varied and complex in the retina; thus, in addition to synthetic activities in microsomes, exchange of bases between different lipid species occurs, in addition to frequent acylation–deacylation reactions. Retinal lipids are continually undergoing degradation via phospholipases and modifications including decarboxylation and methylation. There is also considerable uptake of circulating lipids, particular PUFAs by retinal cells including RPE cells and photoreceptors, which express LDL receptors as well as major facilitator superfamily domain containing protein-2 (MFSD2A) transporter, an intrinsic protein of the blood–retinal and blood–brain barrier.

Blood flow in the retina

A balance between the perfusion pressure and the resistance in the blood vessels determines blood flow in a tissue. Blood flow in the retina is autoregulated, i.e. the retinal vessel calibre varies with the cardiac output thus maintaining a constant, which also meets the nutritional demands of the tissue. In the adult, the blood flow in the retina is maintained constant over a range of ocular perfusion pressures from 45 to 145 mmHg, which is achieved by changes in blood vessel diameter, as shown in mice deficient in the eNOS. Retinal blood flow comprises about 5% of the total flow to the eye, the majority passing through the choroid. Autoregulation of retinal vessels at the optic nerve head, for instance in response to changes in intraocular pressure, can be imaged using laser doppler flowmetry (LDF). More definitively, using such as laser speckle flowgraphy (LSF) has confirmed autoregulation in the retinal circulation but not in the choroid (Kiyota et al., 2018).

The blood–retinal barrier regulates the passage of molecules into the retina

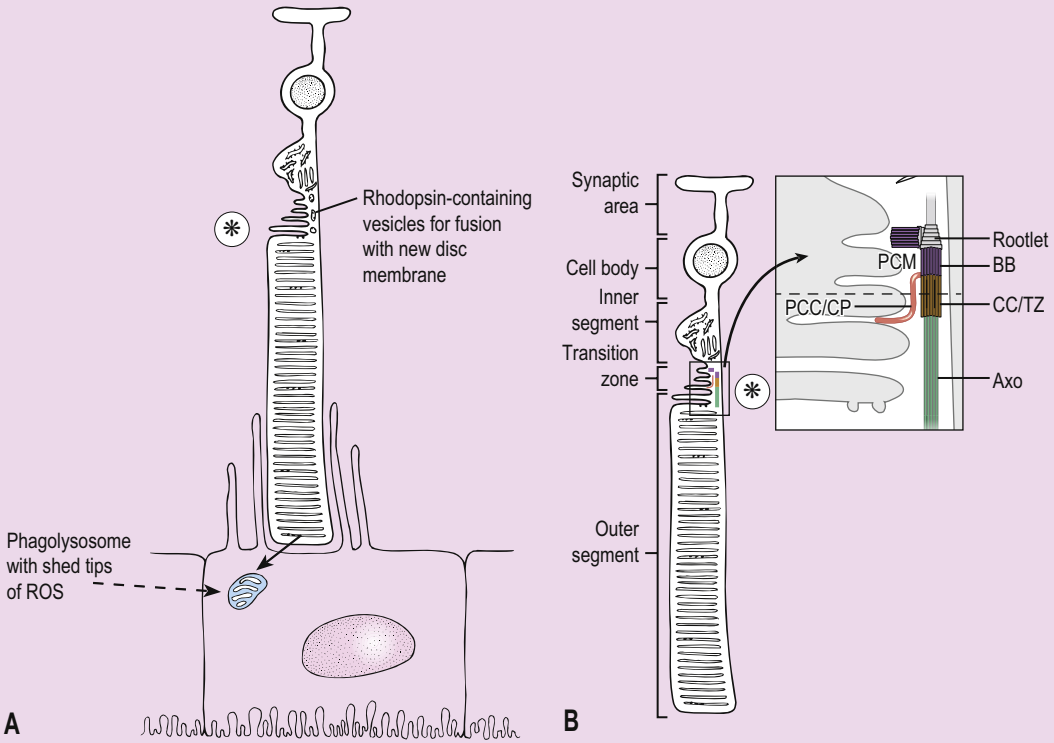
Tight junctions maintain the blood–retinal barrier between the endothelial cells of the retinal vessels and similar tight junctions in the RPE. Thus, the retinal vessels and the RPE are impermeable to the passage of molecules greater than 20,000–30,000 Da, and small molecules such as glucose, amino acids and ascorbate are transported by facilitated diffusion (mostly GLUT1 at both sites, but also GLUT3 in the retinal vessels). Similar transporters are present for amino acids including arginine and both classes of transporters are present in smooth muscle cells and pericytes, ensuring rapid transport and uptake of these moieties into the retinal tissue.

High-affinity receptors for insulin (IR), pro-insulin, IGF-1 and IGF-2 occur in the retinal vessels and in the photoreceptors (Lofqvist). Since the IR-dependent GLUT4 is absent from retinal vessels, insulin and IGF-1/2 do not play a role in glucose transport via retinal vessels. Rather these hormones play a major role in regulating protein synthesis and cell survival in retinal cells.

Box 4.15 **Photoreceptor renewal**

Photoreceptor renewal occurs by synthesis of new protein-rich membrane at the outer limiting membrane, with shedding and phagocytosis of the outer segment tips by the RPE (**A**). Discs are formed by the evagination of the plasma membrane ⊛ at the junction between the inner and outer segments, while rhodopsin and other proteins synthesized in the ER and Golgi apparatus are transported in vesicles for fusion to the newly formed plasma membrane ⊛. Photoreceptor renewal is similar in rods and cones; phagocytosis of the receptor tips occurs in a diurnal manner. The mechanism of phagocytosis is unclear but involves a membrane glycoprotein CD36, which is also involved in the uptake of apoptotic neutrophils and oxidized low-density lipoprotein by haematopoietic cells such as macrophages. In addition, rod outer segment (ROS) phagocytosis is associated with induction of cyclo-oxygenase-2 (COX-2), an enzyme involved in prostaglandin synthesis. Recent studies have also shown the importance of the cilia in the functioning of the photoreceptor, since genetic mutations in a number of genes controlling cilia proteins produce forms of retinal degeneration now known as ciliopathies (**B**).

There are four distinct compartments in photoreceptor primary cilia, indicating known proteins that define their respective extent: (1) distal cilium or axoneme (Axo; *green*); (2) connecting cilium/transition zone (CC/TZ; *orange*); (3) basal body (BB; *purple*); and (4) periciliary complex or ciliary pocket (PCC/CP; *red*). These compartments act as discrete functions in the cilium (see inset in figure).



(Part B reproduced from Rachel and Swaroop, 2012.)

Retinal blood flow may be partly under autonomic control

Retinal vessels possess all four types of high-affinity adrenergic receptors (α1, α2, β1, β2), which when activated *in vitro* cause vasodilatation. In addition, despite well-recognized mechanisms of autoregulation (metabolic coupling), some degree of autonomic control exists in humans through innervation of the central retinal artery at the optic nerve head, which also supplies choroidal vessels by hitching a ride along the short and long posterior ciliary vessels. The retinal vascular bed may be one of the few systems to lack perivascular mast cells, an important response element for catecholamines.

Retinal blood flow is responsive to hyperoxia (vasoconstriction) and hypercapnia (vasodilatation), the latter via the prostaglandins PGD_2 and PGE_2. Other mediators of changes in vessel diameter include the eicosanoid, PGI_2, endothelin and nitric oxide. PGI_2 and endothelin have been detected in retinal vessels and presumably are released under appropriate conditions. Nitric oxide is released via endothelial nitric oxide synthase and provides basal levels of vascular dilatation. Retinal illumination induces release of NO but NO does not significantly affect autoregulation in retinal vessels. Contractile activity in the retinal vessels is attributed to the pericytes, whose role therefore may be to regulate blood flow. However, they also play a role in the immunological functions of the blood retinal barrier. Early loss of pericytes in diabetes may account for the increase in blood flow that occurs in the retina in diabetes and may contribute to the development of retinopathy.

Photoreceptors

Photoreceptors are specialized for reception of visual stimuli and have unique characteristics (see Chapter 1, p. 40).

Metabolism and turnover

Photoreceptors are some of the most highly metabolic cells in the body, utilizing glucose both aerobically and anaerobically. Photoreceptor outer segments lie in apposition to the RPE in the interphotoreceptor matrix between the apical microvilli of the RPE cell. Extensive protein and lipid synthesis ensures a continuous turnover of new outer segment membrane at the junction with the inner segment; the tips of the outer segment (OS) containing the oldest discs are phagocytosed as small packets of about 200 discs by the RPE cell, a process that occurs in a diurnal manner just after light onset (Fig. 4.73). Three RPE cell surface receptors have been identified that are involved in the phagocytosis of outer segments, namely CD36, the tyrosine kinase MerTK and the integrin $\alpha_v\beta_5$. Loss of these proteins may lead to retinal degenerations (MerTK: human retinitis pigmentosa and rat RCS disease; $\alpha_v\beta_5$ integrin: RPE lipofuscin accumulation). Diurnal regulation of phagocytosis is under the control of the secreted glycoprotein milk fat globulin-EGF8 (MFG-E8). MFG-E8 is important in immune tolerance, being preferentially secreted by dendritic cells (DC) (see Chapter 7) and acts as a linker between DC/macrophage $\alpha_v\beta_5$ integrin and apoptotic cells, the same role it plays in RPE phagocytosis of apoptotic OS tips and olfactory ensheathing cells in clearing neuronal cell debris. Activation of MerTK is under autocrine regulation of its natural ligand, Gas6 (growth arrest specific gene 6), which is also expressed in RPE cells or via 'tubby' (Tub) proteins and tubby-like proteins (Tulps 1–4), which are expressed

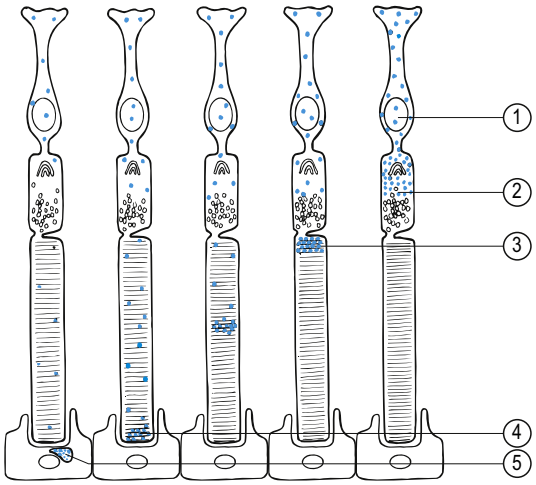


Fig. 4.73 Photoreceptor renewal and points where damage may occur: during *(1)* transcription; *(2)* post-translational modification; *(3)* incorporation into disc membrane; *(4)* disc shedding; *(5)* phagocytosis by retinal pigmented epithelial.

in photoreceptors and generate 'eat me' signals similar to those that occur in apoptotic cells.

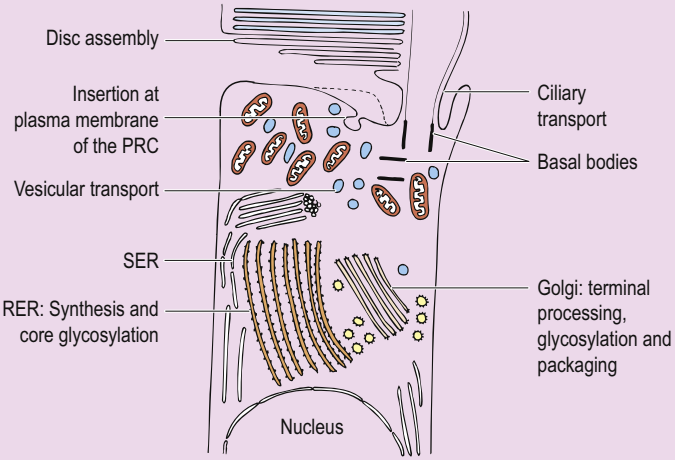
Complete renewal of the rod outer segment takes about 9–10 days. In contrast, although cone outer segments are phagocytosed in a similar manner, the process appears to be more random and occurs when darkness occurs; cone membranes and their integral proteins are much more stable and long-lasting. Rod and cone renewal is critically dependent on protocadherin (Usher syndrome type I protein, USH1), which links the calyceal processes of the photoreceptor to the outer segments. *In vivo* there are several cones (2–4) and even more rods in contact with the RPE cell surface with increasing cone density towards the fovea (up to 20/ RPE cells)(Liu et al., 2016). Cone shedding has been visualized in healthy humans using adaptive optics linked to OCT (AO-OCT).

Insertion of rhodopsin into the disc plasma membrane follows a well-defined pathway from the inner segment RER to the outer segment plasma membrane infolding (see Box 4.16). Glycosylation of rhodopsin takes place through combined co-translational and post-translational events in a classic lipid-carrier mechanism using Dol-P-P-GlcNAc (Fig. 4.74), which can be inhibited by tunicamycin. Acylation of rhodopsin also occurs in the membrane via the saturated fatty acid, palmitic acid.

Lipids in photoreceptors are replaced both by membrane turnover and by molecular replacement. ATP synthesizes the abundant stores of phosphatidylcholine from large intracellular pools of free choline and phosphorylation; the activated choline then reacts with 1,2-diacylglycerol to form phosphatidylcholine. Similar mechanisms

Box 4.16 **Rhodopsin synthesis**

Insertion of rhodopsin into the outer segment plasma membrane is facilitated by the lack of a signal peptide, which permits integration of opsin into the lipid bilayer by co-translational coupling of glycosylation and asymmetric insertion via specific insertion sequences in the protein.



operate for the synthesis of phosphatidylserine and phosphatidylethanolamine; all three phospholipids are synthesized in the RER but are transported to the newly forming outer segment membrane by different mechanisms. The omega-3 essential fatty acid, docosahexaenoic acid (DHA), is preferentially taken up in photoreceptors (and brain cells), retained under control of adiponectin receptor 1 and metabolized to docosanoids and elovanoids, which protect the RPE and CNS neurones from damage (Bazan, 2018). DHA is required for maintaining normal disc morphology (Fig. 4.75).

Several chemical reactions are associated with disc shedding, although the specific stimulus and its site of origin (i.e. the photoreceptor or the RPE cell) are not known. Circadian rhythms and light–dark cycles are under the control of specific genes (CLOCK Per) regulated through melanopsin-positive intrinsically photosensitive cells in the retina and the suprachiasmatic nucleus in the brain (Fig. 4.76), which controls the release of melatonin into the circulation from the pineal gland. However, circadian light–dark rhythm of disc shedding is under the control of locally produced melatonin (and the 5-methoxytryptophol-synthesizing enzymes) in photoreceptors (Fig. 4.77) acting on melatonin receptors (MT1–3 eFig 5.2) in the photoreceptors and in inner retinal amacrine cells. Dopamine counter-regulates retinal circadian rhythms by suppressing melatonin synthesis through the D2/4 receptor. Dopamine may also affect melanopsin, a photopigment present in inner retinal ganglion cells, which acts on the MT2 receptor to regulate melatonin secretion (see Fig. 4.76) (Ostrin, 2019).

The mammalian retina melatonin-responsive circadian oscillator acts in phase with both rhodopsin and cone opsin synthesis and disc shedding. In addition, several other compounds affect disc shedding, such as the excitatory amino acids, glutamine and aspartate, as do certain essential divalent ions (Ca^{2+}, Mn^{2+}). The D4 dopamine receptor also counter-regulates several of these non-melatonin disc-shedding regulators such as a light-sensitive pool of cAMP.

Phosphoinositide metabolism is considerably greater than phosphatidylcholine or phosphatidylethanolamine metabolism in the photoreceptor, but its precise role in phototransduction is unclear (see p. 266). The PIP2 phosphatase appears to be essential for function at the photoreceptor synapse. Cytidine triphosphate is also a product of light transduction and is linked to phosphatidylinositol formation.

Photoreceptor cell-specific proteins

The highly differentiated visual cell contains many unique proteins, including integral membrane proteins,

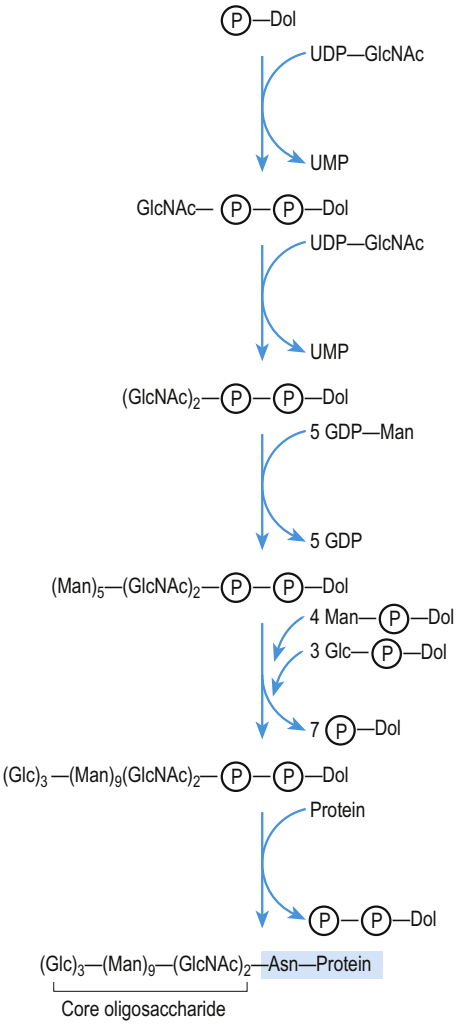


Fig. 4.74 Glycosylation and acetylation of rhodopsin.

membrane-associated proteins and cytosolic proteins, of which more than 30 alone are involved in the phototransduction cascade. Some of these are shown in Table 4.8. Most of these have to pass through the connecting cilium to reach to outer segments, which has the structure of an axoneme composed of a basal body and a stiff microtubule-rich transition zone. Mutations in ciliary proteins cause retinal degenerations known as ciliopathies, including rod-cone dystrophies and Leber's congenital amaurosis (Fig. 4.78). However, there are many more proteins, and many of them have been evolutionarily conserved. Three categories of proteins are described: photoreceptor cell type-specific, process-specific (e.g. phototransduction itself, the retinoid cycle and developmental regulation) and a specific functional role (e.g. opsin, G protein, etc.).

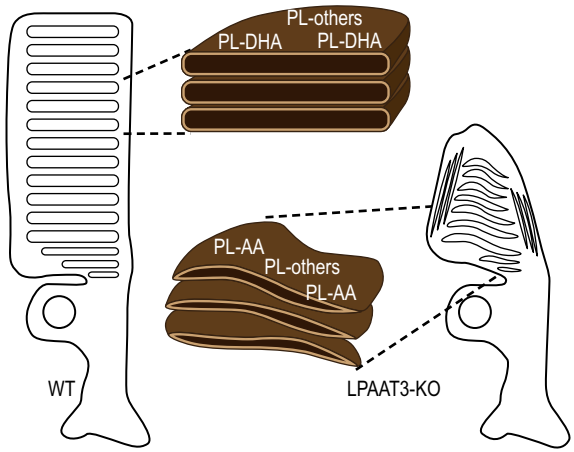


Fig. 4.75 Model figures for the role of docosahexaenoic acid in disc renewal. Low levels of PL-DHA in LPAAT3-KO-induced abnormal disc morphology/organization, which was not rescued by increased PL-AA levels. (From Shindou et. al., 2017.)

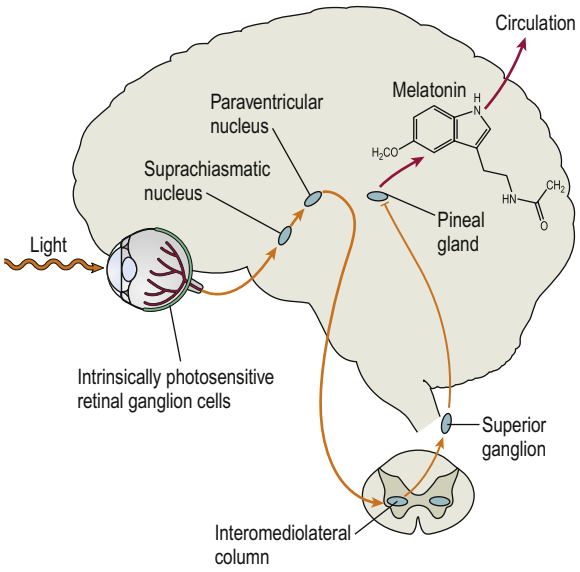


Fig. 4.76 Pathway from light activation of the melanopsin positive, intrinsically photosensitive retinal ganglion cells to pineal gland melatonin suppression. (From Ostrin, 2019.)

The protein composition of the disc membrane is predominantly rhodopsin (90%), whereas the plasma membrane has a wider range of cell-specific proteins and less rhodopsin (50%). Rhodopsin is the visual receptor protein; proteins such as peripherin/RDS (Fig. 4.79) and the spectrin-like protein Rom-1 have structural functions in the maintenance of the photoreceptor shape, similar to those of

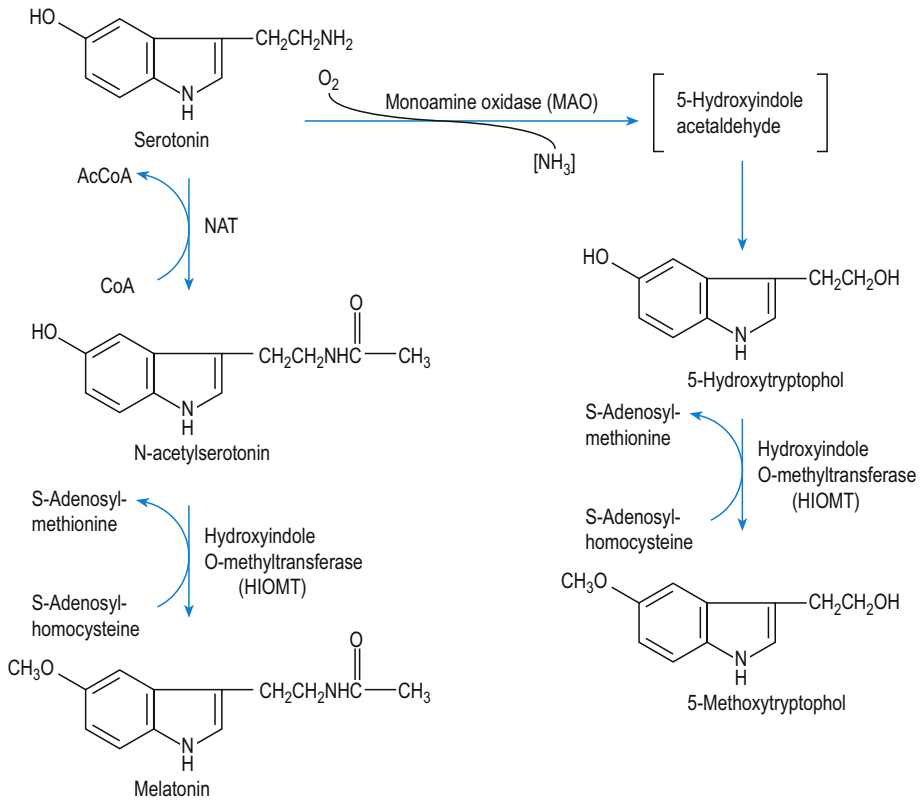


Fig. 4.77 Regulation of disc shedding. Disk shedding may be under the control of melatonin-synthesizing enzymes such as serotonin N-acetyltransferase (*NAT*), which is present in the pineal gland and retina.

Table 4.8 **Photoreceptor proteins**

Integral membrane proteins	Peripheral/cytosolic proteins
Rhodopsin	Arrestin (48 kDa protein, S antigen)
cGMP channel	Transducin
Na^+/Ca^{2+}-K^+ exchanger	Phosphodiesterase
Glucose transporter	Phosducin
Guanylate cyclase	Rhodopsin kinase
Peripherin/RDS	Guanylate cyclase
Rom-1	
ABCR/rim protein	
Retinal dehydrogenase	

spectrin/ankyrin-like proteins in red cells. Both are members of the tetraspanin family of proteins and may also have a role in membrane fusion important for rod disc generation.

Remarkably different mutations in the RDS protein (Fig. 4.80) produce different clinical types of retinal degeneration such as autosomal dominant retinitis pigmentosa and macular dystrophy (see Chapter 9, p. 519). The Na^+/Ca^{2+} exchanger facilitates ionic transport during phototransduction and maintains Ca^{2+} homeostasis. In addition, the spectrin-like protein Rom-1 is linked to a second ion-channel protein of 67 kDa. Stacking of rod and cone discs is a central organizational requirement for ordered photoreceptor function and a set of glutamic acid-rich proteins (GARPs) interact with RDS/Rom to structure stacking of discs and to participate in ion-channel functions. GARPs tend to be relatively unstable as a result of their unfolded native state and may accelerate retinal degeneration.

In contrast to the integral membrane proteins, many of the peripheral membrane and cytosolic proteins are not exclusive to the photoreceptor and are intimately involved in the light amplification cascade itself (see later). Transducin is a member of the family of G proteins and is composed of three chains (α, β, γ), which dissociate during the light response.

Certain other proteins are located in the photoreceptors as well as in other areas of the retina; these include the insulin receptor, GLUTs 1 and 4, IGF-1 receptor, IGF-1-binding protein, and FGF all with implications for diabetic retinopathy (see Chapter 9).

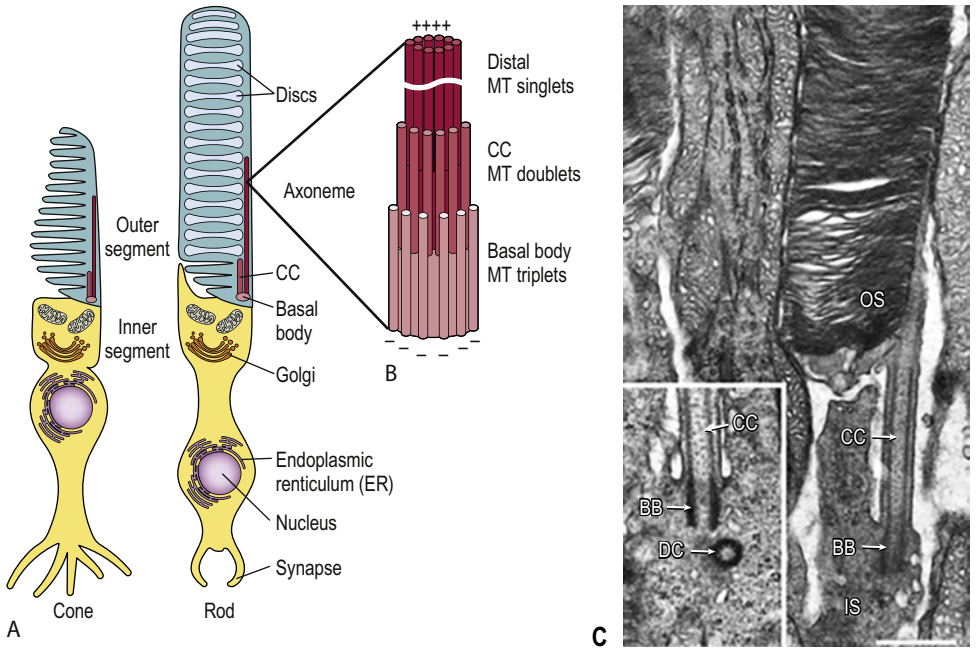


Fig. 4.78 Basal body/ axoneme cytoskeleton. **(A)** Schematic of cone and rod photoreceptors depicting the outer segment where phototransduction occurs, the inner segment containing the endoplasmic reticulum and Golgi apparatus, the nuclear region and the synaptic terminal. **(B)** enlarged detail of the axoneme cytoskeleton consisting of basal body (= mother centriole) proximally, transition zone and axoneme distally. C, electron micrograph of mouse rod, partial view, revealing the basal body, microtubule stabilised CC and outer segment (OS) stacked with membrane discs. Note the daughter centriole is out-of-plane and not visible; inset shows a different BB/DC pair. Scale 0.5µm.

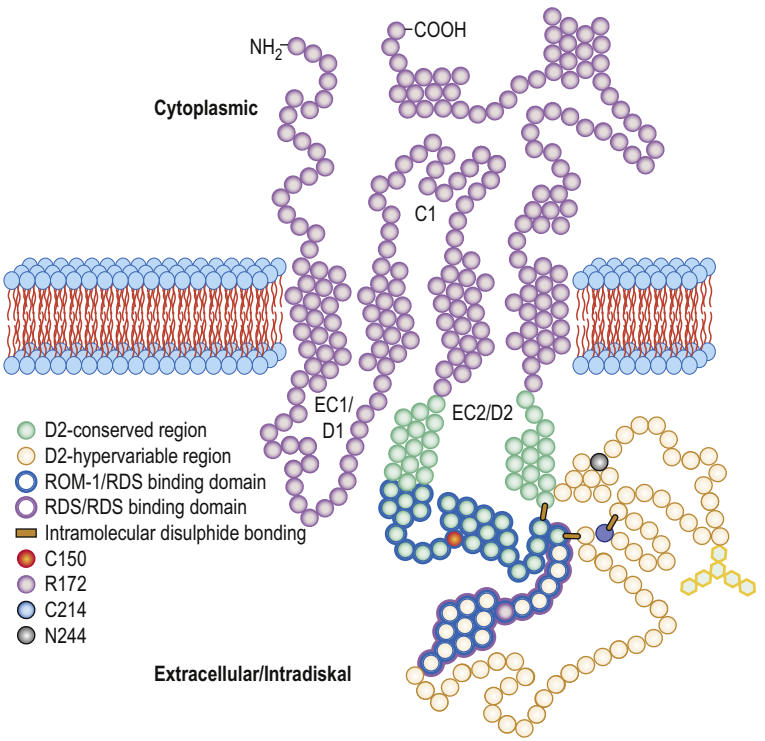


Fig. 4.79 Structure of the RDS/peripherin protein. Structural features of the RDS protein. RDS contains four transmembrane domains, two ransducing loops *(D1 and D2)*, cytoplasmic N- and C-termini, and a small cytoplasmic loop *(C1)*. (From Chakraborty et al., 2013.)

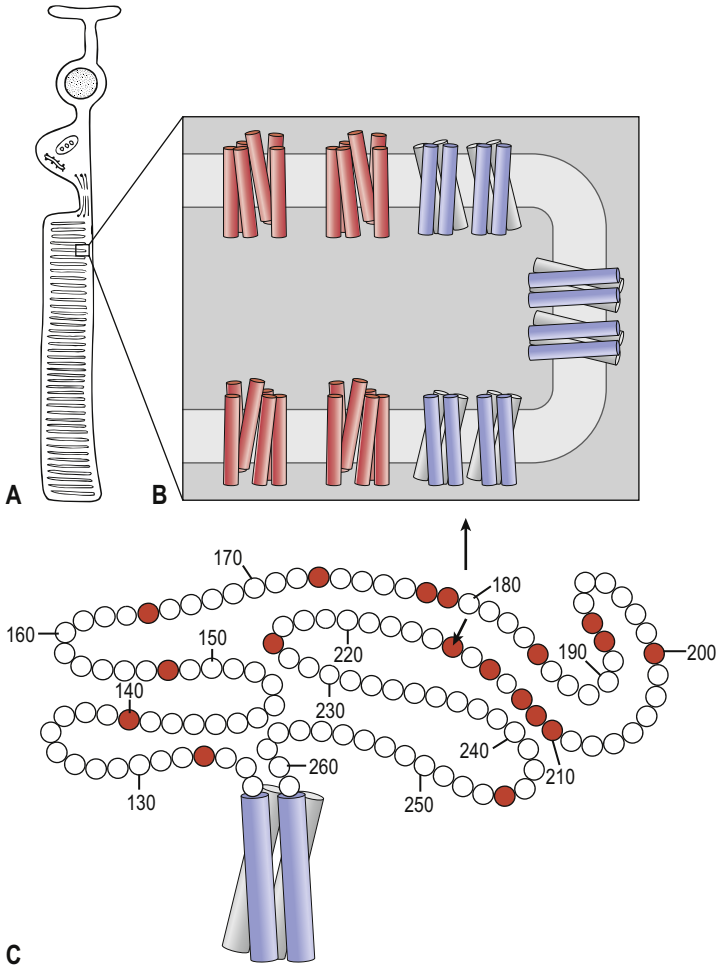


Fig. 4.80 The rim region of photoreceptor disc shown in (**A**) are expanded in (**B**) to reveal the 4-span (tetraspanin) arrangement of the RDS/peripherin molecule with the intradiscal portion of the molecule shown in (**C**). Specific amino acids locating mutations responsible for forms of retinitis pigmentosa are identified in *red*. (From Vos et al., 2010.)

Melanopsin and neuropsin both entrain circadian rhythms

Melanopsin is one of a family of opsins now renamed OPN1-5. Melanopsin (OPN4) is present in the retina, lens, ciliary body and cornea, as well as in non-ocular structures. In the retina, five different types of specialized ganglion cells (intrinsically photosensitive retinal ganglion cells, ipRGCs) with large receptive fields (see Chapter 5, Box 5.8, p. 296) express OPN4 and are located in the peripheral retina. OPN4 cells entrain the circadian clock, sleep and pupillary responses to light. Remarkably, the light (melanopsin) SCN axis (see Fig. 76) also leads to sympathetic stimulation (norepinephrine release and hedgehog signalling) and can increase skin pigmentation and activate hair follicle stem cells leading to hair growth. A second opsin (neuropsin, OPN5) also regulates circadian rhythms through one of the ipRGC and even entrains the expression of corneal circadian clock (Fig. 4.81).

Photoreceptors are easily damaged but readily regenerate

The extensive metabolism and rapid turnover of photoreceptor outer segments render them highly susceptible to damage. Damage may occur at any level from synthesis of new membrane to phagocytosis of the outer segment tips (see Fig. 4.82). Photoreceptors also degenerate when they are separated from the RPE, as in retinal detachment, or when there is a subretinal collection of fluid; photoreceptors are lost in inflammatory and metabolic retinal diseases and are probably highly

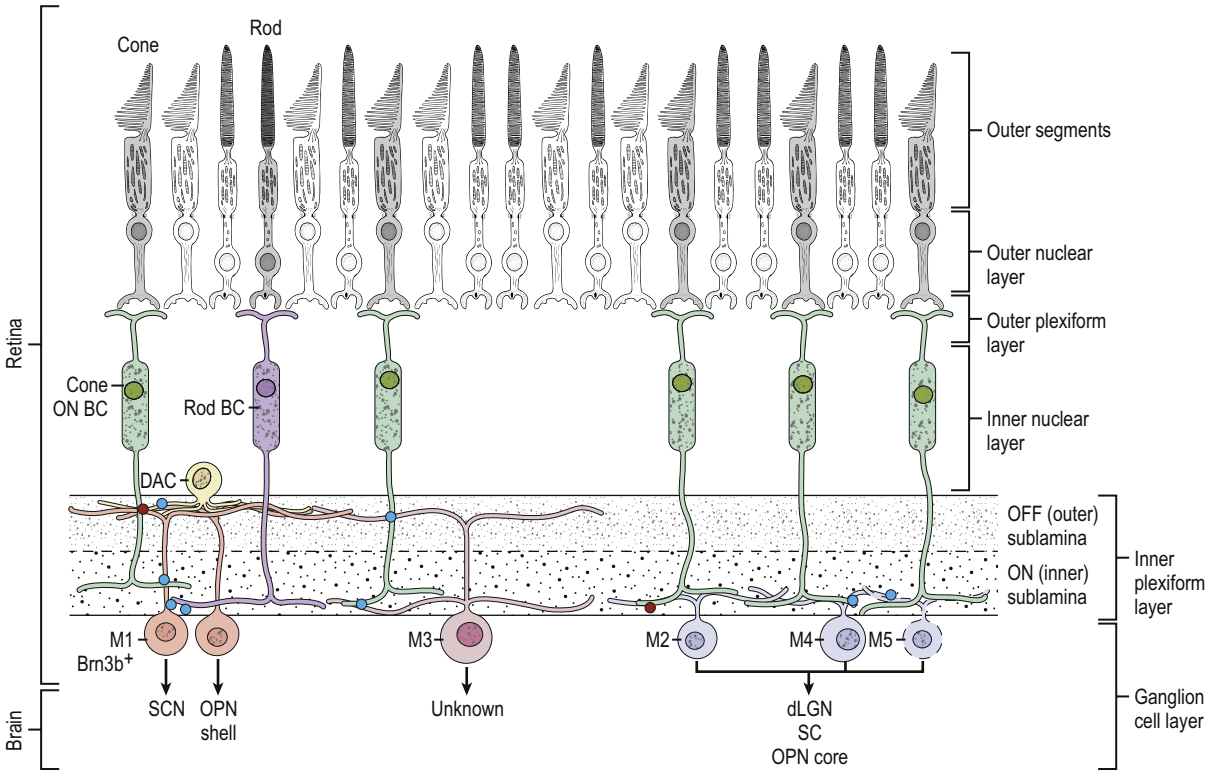


Fig. 4.81 Five distinct morphological types of melanopsin-expressing ipRGC exist, as shown here, and project to different parts of the brain such as the lateral geniculate nucleus *(LGN)*, the suprachiasmatic nucleus *(SCN)* and the olivary pretectal nucleus *(OPN)* (see Chapter 1, p. 45 and Chapter 5, p.314). (From Schmidt et al., 2011.)

susceptible to free radical damage. Specific photoreceptor loss occurs in various forms of retinal degeneration: inherited forms of these disorders are collectively known as retinitis pigmentosa often linked to mutations in retinal specific proteins (see p. 254 and Chapter 3, p. 160). Ophthalmic genetic diseases are currently being catalogued and over 180,000 mutations have been attributed to a range of ocular diseases, many of which include the retinal degenerations.

Damage to the retina also occurs as part of normal physiology via both light- and oxygen-induced mechanisms. Around one-third of the mutations associated with retinal degenerations occur as ciliopathies (see earlier), in which the Wnt signalling pathway is implicated.

Further information on Wnt signalling and cell functions is available online at https:// expertconsult.inkling.com

The *Wnt/Drosophila wingless* family of genes expresses highly conserved secreted glycoproteins that interact with the products of two sets of genes (*Frizzled* and genes for low-density lipoprotein-related proteins, LRPs). In the fruit fly, *Drosophila*, a species that is frequently used to study molecular genetics, Wnt signalling is central to normal photoreceptor development as well as to brain development, through regulation of pro-survival factors (anti-apoptotic factors) such as Dickkopf (*Dkk3*) in cells. In fact, this mechanism highlights the fact that photoreceptor damage in retinal degenerations and other conditions is a process involving accelerated apoptosis, which occurs by caspase-dependent and caspase-independent mechanisms (see p. 174).

Light-induced retinal damage, thought to be involved over time in AMD, varies with the intensity, wavelength, duration, cyclical nature and previous antioxidant status. Cyclical light is less injurious to the retina than constant illumination of equivalent power, an effect associated with higher levels of ascorbic acid, vitamin E and glutathione, and lower levels of 22:6(n-3) fatty acids in the tissue. RPE65, the vitamin A/retinoid cycle enzyme is an intrinsic RPE protein with a critical amino acid residue at position 450, which regulates the rate of rhodopsin synthesis through the action of two proteins, elongation of very long fatty acid (VLFA)-like 1 (ELOVL1) and fatty acid transport protein 4 (FTAP4), both of which have VLFA synthase activity. This demonstrates the complex biochemical machinery in place to protect photoreceptors from light damage. The

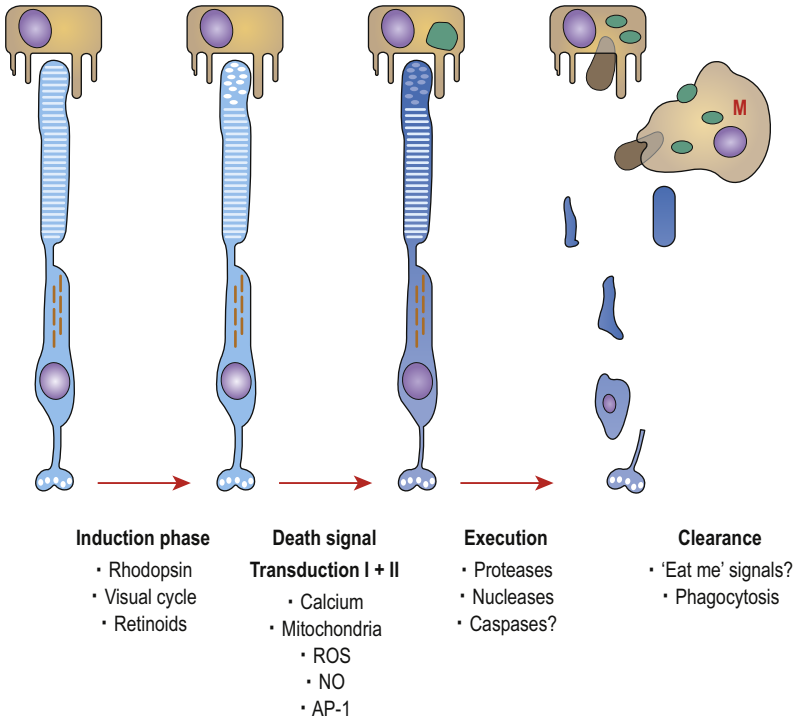


Fig. 4.82 Schematic drawing of major components that contribute to light-induced photoreceptor apoptosis. *Induction phase:* rhodopsin is essential in this process. *Death signal transduction:* requires Ca^{2+} and transcription factor AP-1. *Execution:* mechanism is not clear and might require caspases. *Clearance:* signals inducing phagocytosis by RPE and, in acute light damage, macrophages, are not known. (From Wenzel et al., 2005, with permission from Elsevier.)

death signal for photoreceptors involves induction of the pro-apoptotic transcription factor AP-1, such that inhibition of regeneration of rhodopsin or suppression of AP-1 can prevent light-induced damage to photoreceptors (Fig. 4.82). ABCA4, a flippase that translocates retinaldehyde from the disc to the plasma membrane (see later text), is also present in the RPE cells in the endolysosomes and together with MerTK promotes phagocytosis of outer segments. TULP 1–4 'tubby' proteins send 'eat me' signals to RPE cells for ROS tip phagocytosis. Mutations in ABCA4 cause Stargardt disease.

Extensive antioxidant enzyme systems in the retina to deal with light damage: glutathione peroxidase and glutathione *S*-transferase levels become markedly raised while glutathione peroxidase4 (Gpx4) is essential for normal photoreceptor development. Superoxide dismutase (SOD) (the Cu^{2+}/Zn^{2+} form) is also present in significant amounts in several layers of the retina, but levels of catalase are low. Despite this, treatment with SOD and catalase in addition to other antioxidants (e.g. *N*-acetylcysteine and thioredoxin) can prevent retinal degeneration in experimental models. Melatonin also has important antioxidant activity. In addition, vitamin E is of major importance in reducing photoreceptor damage, mainly by its role in inhibiting lipid peroxidation at different stages in the process of free radical damage (see Box 4.15).

Iron (Fe^{2+} ions) interacts with hydrogen peroxide to produce high levels of hydroxyl radicals (see Box 4.6, p. 200), which cause retinal injury. Local release of iron may occur in diseases involving retinal or vitreous haemorrhage and when intraocular foreign bodies are present. Ceruloplasmin, which contains 95% of the copper in the serum, mediates iron metabolism via ferroxidase, which converts ferrous to ferric iron before it is delivered to serum transferrin after efflux from the cell. Patients with the recessive disorder leading to altered ceruloplasmin blood levels have severe progressive retinal degeneration. Iron homeostasis is essential to normal retinal function and the five main iron metabolism regulatory proteins are present in the RPE cell (Fig. 4.83).

The macular region of the retina is particularly susceptible to light damage. The yellow xanthophyll (carotenoid) pigments lutein, zeaxanthin, α and β carotene, lycopene and meso-zeaxanthin, which are obtained from dietary plants and transported in the blood, act as a filter to block short-wavelength light. Chemically they are tetra-terpenoids with a 40 carbon isoprene backbone (Fig. 4.84) and they can be measured *in vivo* by imaging using Raman spectroscopy. They

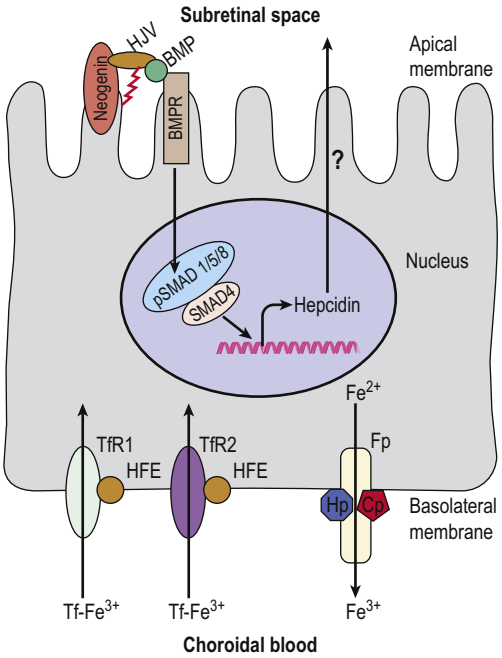


Fig. 4.83 Expression pattern of haemochromatosis genes in retinal pigment epithelium (RPE). *Cp*, Ceruloplasmin; *Fp*, ferroprotein; *HFE*, HLA-like protein involved in iron homeostasis; *Hp*, hephastin; *TfR*, transferrin receptor. (From Gnana-Prakasam et al., 2010.)

belong to a larger family of around 600 carotenoids that have an antioxidant effect, particularly at low levels of oxygenation. Sustained levels of these pigments are considered protective against macular degeneration and to promote longevity generally. They and other moieties such as the amino acid taurine have been labelled 'longevity' vitamins, in line with the concept of categorizing proteins generally as either survival proteins or longevity proteins (Ames, 2018).

Finally, melanin in the RPE layer is a very effective free radical scavenger, particularly in the reduced form, where it may act as a trap for stray light-induced free radicals emanating from the photoreceptors. However, the role of melanin is more likely to be related to protection of the RPE cells themselves because these are terminally differentiated cells. Melanin may act as an antioxidant because of its ability to bind metal ions such as Fe^{2+}, a process well known to occur in fungi. Melanin granules in RPE cells appear to be connected to the lysosomal enzyme system in RPE cells and loss of melanin is associated with AMD.

The interphotoreceptor matrix is the biological glue for retinal adhesion

The interphotoreceptor matrix extends from the outer limiting membrane of the retina (see Chapter 1, p. 44) to the surface of the RPE cell. It is extremely narrow, almost a potential space, but it contains some unique and physiologically important molecules. These include interphotoreceptor retinol-binding protein (IRBP; accounts for at least 70% of the interphotoreceptor matrix protein), which transports retinoids between the

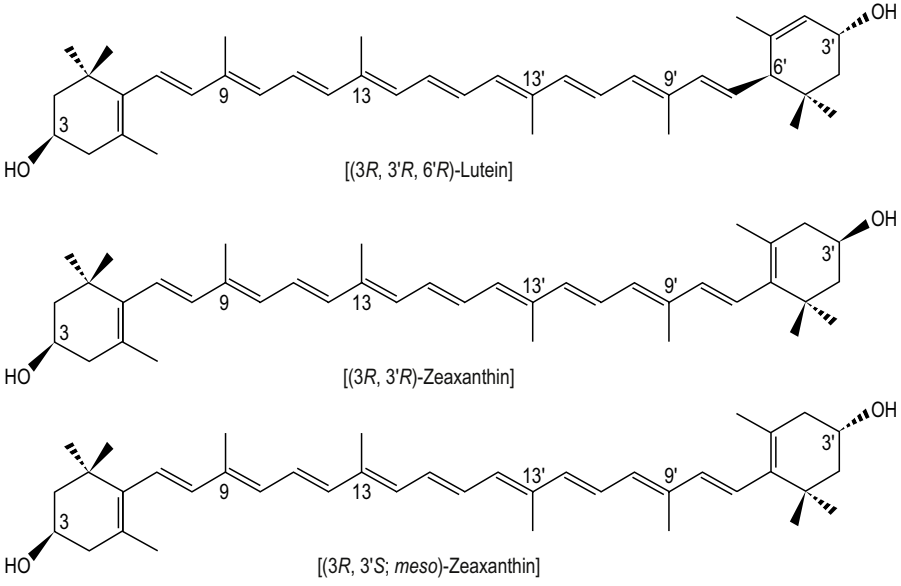


Fig. 4.84 The chemical structure of macular pigments. (From Abdel-Aal et al., 2013.)

RPE cell and the photoreceptor, and several species of proteoglycan, which provide a coating for the photoreceptor outer segment. Interestingly, the 6-sulphated (DeltaDi6S) chondroitin species appears to be more prominent around cones than around rods. IRBP appears to be essential for photoreceptor survival but is not required for the visual cycle. In addition, there are several cell surface proteins that form part of the glycocalyx of the cell but may also play a part in maintaining retinal apposition, which is essentially maintained by the bulk flow of fluid across the retina. These include fibronectin, intercellular adhesion molecule 1 (ICAM-1) (see Chapter 7 p. 398) and the CD44 antigen, also known as the hyaluronan receptor. There is also a second hyaluronan-binding protein, SPARC (sialo-protein associated with rods and cones), which may also be important in providing a scaffold to the matrix.

The interphotoreceptor matrix of the cone is compartmentally separated from that of the rod by an insoluble matrix sheath containing its own specific proteoglycans, which presumably play a role in the regulation of their different forms of visual excitation. Chondroitin-6-sulphate appears to be the major proteoglycan–GAG in the cone matrix sheath, while a similar but less well-defined sheath around rod outer segments appears to be composed of sialyl conjugates. It has been suggested that these proteoglycan sheaths around both rods and cones are important in retinal–RPE adhesion, for instance via hyaluronan–CD44 interactions on the apical microvilli as well as via SPARC and SPARCAN through hyaluronan-mediated motility receptor (RHAMM) domains.

The interphotoreceptor matrix contains other GAGs, including non-sulphated chondroitin and hyaluronan (about 14% of total). These are present as proteoglycans and are probably mostly synthesized by the RPE cell. Chondroitin sulphate binds interphotoreceptor matrix proteins (IMPG1/2) while hyaluronan binds to the proteins brevican and versican (members of the hyalectin family of proteins), which are also important in maintaining retinal apposition to the RPE. The surface tension adhesive force of 24 mN/m of the retina is a result of the IPM and its proteoglycan content (Gonzalez-Fernandez, et al., 2018). There are also a number of proteolytic enzymes and their inhibitors, including matrix metalloproteinases and tissue inhibitor of matrix metalloproteinases (TIMPs) mostly as RPE cell surface bound molecules to mannose-6-phosphate receptors. In addition, αβ crystallins have been identified in the IPM derived from the RPE in an exosome-mediated mechanism.

The retinal pigment epithelium

The RPE is a pluripotent cell

The retinal pigment epithelium (RPE) is a multifunctional pluripotent cell, which befits its embryological origins. It possesses several cytokeratins characteristic of epithelial cells, yet also contains vimentin, which is a mesenchymal cell protein. One group of microfilaments, namely actin, myosin, α actinin and vimentin, is organized as a ring (or belt) around the cell and inserts into typical zonula occludens; a second type of microfilament bundle, containing actin, myosin, fodrin and vimentin, occurs in the apical processes and is involved in microvilli polarity and photoreceptor renewal. The RPE also contains a set of microtubule-associated proteins: light chain-1 (MAP1 LC3A, B and C) of which LC3B is important in lipid homeostasis and whose defects lead to retinal degeneration.

Under conditions of stress, RPE cells may express proteins more typical of macrophages and other myeloid cells. These include receptors for the Fc portion of immunoglobulin, the CD68 molecule and inducible nitric oxide synthase (iNOS) (see Chapter 7, p. 397). RPE cells also express the leucocyte marker CD36, which is involved in ROS phagocytosis as well as $\alpha_v\beta5$ integrin and the tyrosine kinase MerTK. In addition, RPE cells contain high quantities of phosphatidylcholine and phosphatidylinositol with high levels of saturated fatty acids and arachidonic acid. This may explain to some extent their ready ability to generate prostaglandins with immunosuppressive properties. The high content of cholesterol in RPE membranes indicates a low plasma membrane fluidity compared, for instance, with rod outer segment membranes, whose cholesterol content is low.

Turnover in normal RPE is similar to that in endothelial cells (i.e. very slow or nil) since it is a post-mitotic cell, but under certain circumstances RPE cells can proliferate and contribute to pathological processes as in retinal detachment (see Chapter 9). When appropriately stimulated, RPE cells can synthesize and secrete growth factors such as FGF, IGF-1 and interleukin-1, which most likely have a role in the normal physiology of the retina. In addition, RPE cells are an important inducible source of vascular endothelial growth factor (VEGF). The production of growth factors by the RPE is not fully understood teleologically and it may play a part in regulating other tissues, such as the choroid and sclera, and thus indirectly have a role in the development of myopia.

A brief introduction to the optics of the eye is available online at https://expertconsult.inkling.com

However, in pathological situations this may contribute to conditions such as diabetic retinopathy and subretinal neovascularization. To offset this risk, RPE cells constitutionally secrete a specific anti-angiogenic protein, pigment epithelium-derived factor (PEDF), as well as secreting thrombospondin-1 into the pericellular matrix.

The role of the RPE as the second site of the blood–retinal barrier (see Chapter 1, p. 40) is based on normal tight junctions containing the tight junction-specific proteins claudin-3 and -19 and is completed with a ring of cytoskeletal actin–myosin (see previous text); since the RPE also has a major role in pumping fluid across the retina from the subretinal space to the choroid, the tight junctions are

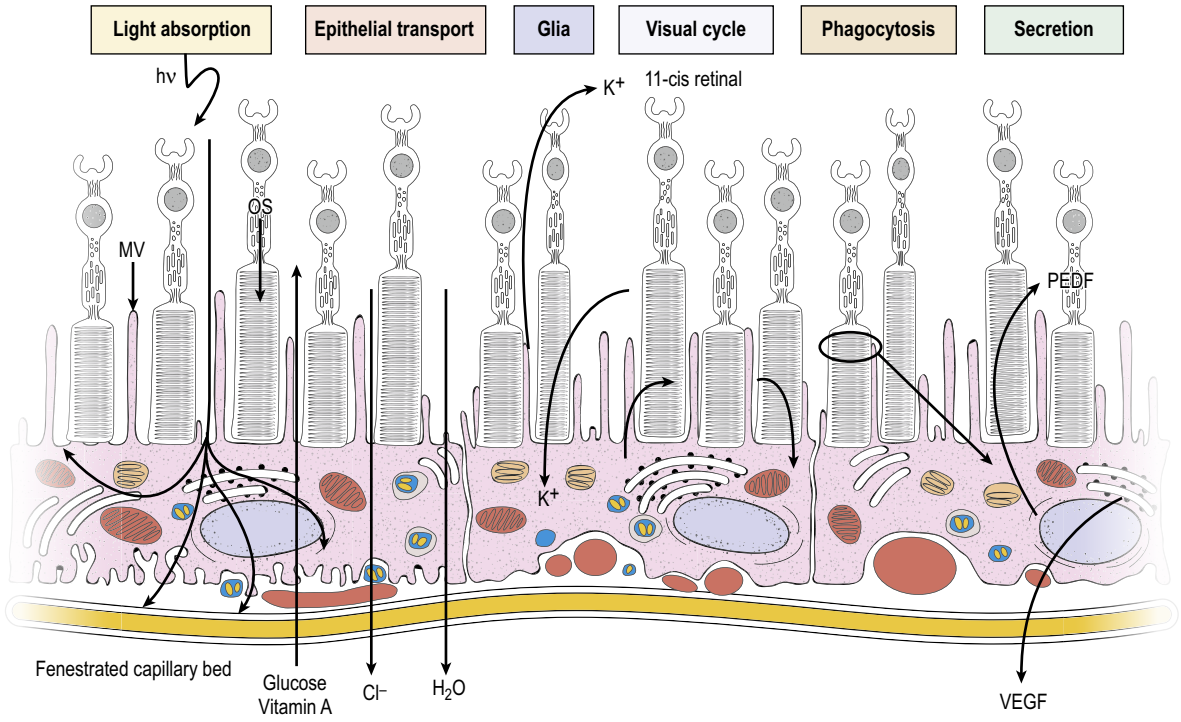


Fig. 4.85 Summary of retinal pigment epithelium (RPE) functions. *MV*, microvillus; *OS*, outer segment; *PEDF*, pigment epithelium-derived growth factor; *VEGF*, vascular epithelium growth factor. (From Strauss, 2005.)

integrated with a series of membrane transporters including Cl^- and HCO_3^- channels, under control of ATPases and carbonic anhydrase, as well as GLUTs 1 and 3. Thus bidirectional transport of various metabolites occurs, but the main bulk flow of fluid from the retinal side to the choriocapillaris is mostly by the cystic fibrosis transmembrane conductance regulator (CFTR), the Ca^{2+}-dependent Cl^- channel, which is one of the family of ABC transporters (mutations in which are responsible for certain forms of retinal degeneration) rather than by aquaporin 1. Aquaporin 1 has not been detected in human RPE *in vivo* except overlying drusen in AMD. Bestrophin 1, the gene responsible for Best's disease, may also be involved in either Cl^- or HCO_3^- transport in the RPE. The RPE cell is therefore a highly polarized epithelium, organized for transport, with its apical microvilli in apposition to the photoreceptor cell and its basal infoldings towards the choroid (Fig. 4.85).

RPE cells also express proteins on their apical surface that would be basolaterally expressed in other epithelia. These include N-CAM (a cell adhesion molecule) and EMMPRIN (CD147, extracellular matrix metalloproteinase inducer, also known as basigin), which are involved in photoreceptor adhesion and phagocytosis, respectively. EMMPRIN positions monocarboxylate transporters (for lactate and pyruvate) in the membrane. CD147 activates T cells. In addition, the phagocytosis receptor $\alpha_v\beta_5$ integrin is expressed on the apical surface together with the Na^+-K^+ ATPases involved in transport. Other apical membrane-associated proteins include ezrin, which is associated with long apical microvilli, radixin and moesin/yurt, some of which are essential to apical basal polarization and are dependent on a range of signalling molecules such as a lipid phosphatase (PTEN), a serine-threonine kinase, and a transmembrane protein (Crumbs). Reversed apical polarization occurs postnatally and is the result of suppressed decoding of specific basolateral signals.

The RPE sits on a prominent basement membrane, Bruch's membrane, composed of five discrete layers, part of which incorporates the basement membranes of the RPE and the choriocapillaris, respectively. In addition to basement membrane proteoglycans and matrix proteins, Bruch's membrane contains hyaluronan and chondroitin sulphate plus types I, III, VI and VII collagen, and elastin.

Photoreceptor function is critically dependent on a healthy RPE layer

The RPE has multiple functions (Table 4.9) but its essential role is to maintain the physiology of the photoreceptors. Thus, it removes 'spent' photoreceptor tips in the diurnal

Table 4.9 Functions of the retinal pigmented epithelial

Photoreceptor renewal
Retinal attachment
Interphotoreceptor matrix production
Transport of water and metabolites
Retinoid metabolism
Blood–retinal barrier
Immunoregulation
Free radical scavenging

process of receptor renewal and participates in 11-*cis* retinol recycling. Indeed, it has been known for over 100 years that contact with the RPE was necessary for the bleached retina to regain its 'visual purple'.

ROS removal is faster than cone tip removal. Phagocytosed outer segment tips are digested in the extensive RPE phagolysosomal system, a process that continues throughout life. Solubilized waste materials are transported across the extensive basal infoldings of the cell into the choriocapillaris. With age, there is accumulation of autophagic bodies and lipofuscin pigment, which may reflect the declining ability of the RPE cell to handle large amounts of relatively indigestible material. This may be linked to reduced melanin production with age. The partly digested rod outer segment 'retinoid' material, some of which is known as A2E (see later), has been implicated in the development of age-related macular degeneration.

Retinoid material in the phagocytosed outer segment tip is either fully degraded or retained as lipofuschin. It is not recycled for regeneration of bleached rhodopsin. Instead, this occurs in the cytosol of the RPE cell. The conversion of 11-*cis* retinal to the all-*trans* retinal during phototransduction is accompanied by release of the chromophore into the interphotoreceptor matrix (Fig. 4.86; and see later). Regeneration of the 11-*cis* retinal from the all-*trans* retinal takes place only in the RPE, since retinol isomerase (RPE 65) is unique to this cell and Müller cells in the retina (see later text). Therefore, the all-*trans* retinal must somehow enter the RPE cell by a means other than outer segment tip phagocytosis. A slightly different process occurs for cone chromophore regeneration. In cones, after phototransduction, 11-*cis* retinal is reduced to all-*trans* retinal, which is then transported to Müller cells, where it is converted to 11-*cis* retinol and transported back to cones, where it is reconverted back to 11-*cis* retinal (Fig. 4.86). This last step can only occur in cones.

How the transport of the chromophores in the interphotoreceptor matrix is achieved is not yet clear, but it is probably more subtle than the proposed mechanism of retinoid shuttling in which IRBP 'transports' the retinoids between the two cells; rather, it is dependent upon the 'buffering' effect of IRBP, in which low-affinity binding of retinoid to IRBP permits its

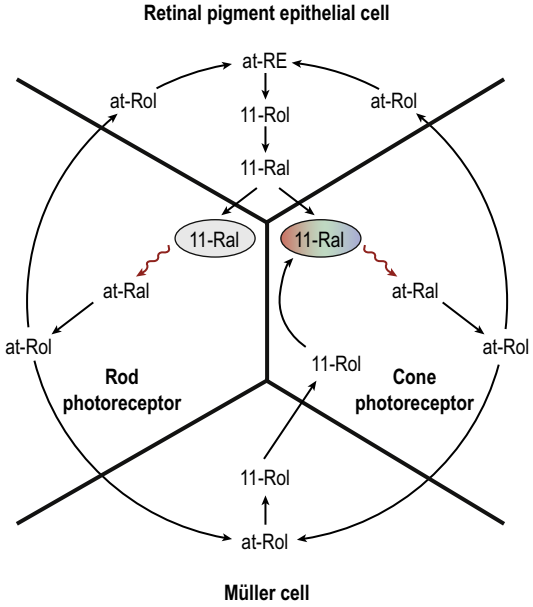


Fig. 4.86 Reactions of rod and cone visual cycles. The areas separated by *solid lines* represent cellular compartments of a retinal pigment epithelial cell, rod and cone photoreceptor cells, and Müller cells. The ovals surrounding 11-Ral represent rod *(grey fill)* and cone *(tricolour fill)* visual pigments. Photoisomerization reactions are shown in *red*. All other chemical reactions are catalysed by enzymes. Retinoids are chaperoned by retinoid-binding proteins (not shown) during intercellular and intracellular diffusion. Diffusion of at-Rol from rod photoreceptor cells to Müller cells has not been demonstrated. *At-RE*, All-*trans* retinyl esters; *at-Ral*, all-*trans* retinal; *at-Rol*, all-*trans* retinol; *11-Ral*, 11-*cis* retinal; *11-Rol*, 11-*cis* retinol. (From Saari, 2012.)

release at the appropriate site depending on the local concentration and amount of chromophore bleach. This notion is indirectly supported by evidence from the IRBP knockout mouse; this mouse has a fully functional neurochemical visual cycle but an accelerated dark adaptation response. In contrast, the exclusive localization of the high-affinity binding protein cellular retinol-binding protein (CRBP) to the RPE cell will help to drive the 'flow' of all-*trans* retinol in this direction and assist in the development of concentration gradients. Docosahexaenoic acid (DHA, see p. 253), an essential photoreceptor fatty acid, in contradistinction to palmitic acid, induces a rapid and specific release of 11-*cis* retinal from one of the two retinoid binding sites on IRBP.

For rods, most of this activity takes place in the dark. However, IRBP also functions in transporting 11-*cis* retinol to cones, which can take place in light conditions. There are two binding sites for the chromophore on the IRBP molecules, which may also have functional significance in terms of vitamin A transport.

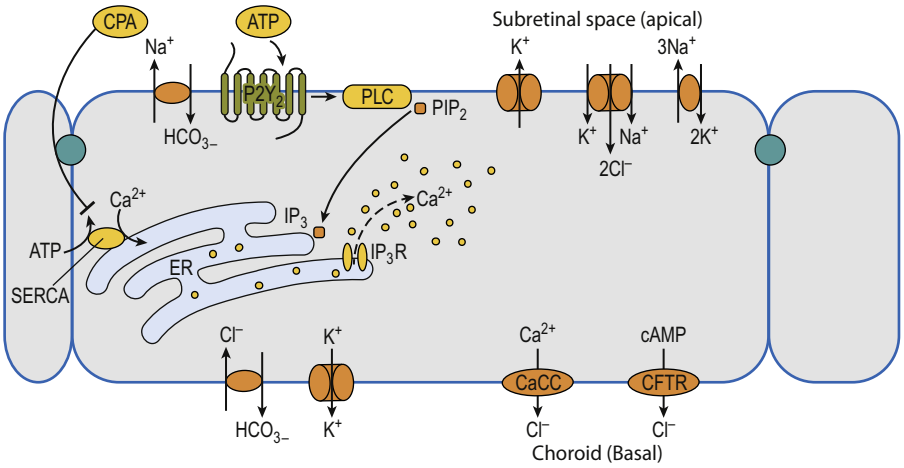


Fig. 4.87 ATP-induced Ca2+ increase in human iPSC-RPE indicates intact Ca2+ signalling. Binding of ATP to purinergic receptors P2Y2 on the RPE apical surface leads to an increase in cytosolic IP3 and subsequent release of Ca2+ from the ER *(yellow)*. Inhibition of ER Ca2+-ATPase with the compound CPA prevents refilling of ER Ca2+ stores, thus blocking the ATP-evoked response disrupting the polarized distribution of receptors, ion channels and transporters (e.g. apical membrane Na+/K+-ATPase, Na+/K+/2Cl- cotransporters, K+ channels, and Na+/HCO3- exchangers, as well as basolateral membrane CaCC, CFTR Cl- channels, K+ channels, and Cl-/HCO3- exchangers). (From Miyagishima et al., 2016.)

The RPE is polarized with tight junctions, but transport is bidirectional

The transport function of the RPE is bidirectional in that, as shown previously, retinoid transport and glucose transport are achieved through polarized distribution of specific receptors (Fig. 4.87). In contrast, the bulk of transport occurs from the retina to the choriocapillaris, particularly for digested outer segment material. There is also a significant bulk flow of fluid from the retina to the choroid, probably via 'solute drag' through the Na^+/K^+ ATPase pump located in the apical plasma membrane of the RPE and by transport of non-ionic solutes such as amino acids and glucose. Interestingly, the healthy RPE cells do not express aquaporins (see later) unlike Müller cells and astrocytes but have active HCO_3^- and Cl-transporters (CFTR, see earlier). HCO_3^- is the major ion linked to fluid transport mediated by a carbonic anhydrase-regulated system (see p. 261). There are six isoforms of carbonic anhydrase in the RPE cell. The exchange of ions (and bulk water) establishes an electrical potential across the RPE monolayer, positive on the retinal side, of 5–15 mV, and also maintains steady pH regulation in the face of high lactic acid production by the photoreceptor cells. However, as indicated above, the Ca^{2+}-activated Cl^- channel, CFTR, and bestrophin membrane proteins play a significant role in trans-RPE fluid flow. Calcium release from the RPE ER is dependent on binding of ATP to purinergic receptors on the RPE. A summary of RPE fluid transport mechanics is shown in Fig. 4.83.

The net rate of fluid transport across the RPE is about 4–6 μL/cm^2 per hour. Transretinal fluid flow has been proposed as a major mechanism for maintaining retinal apposition; indeed, it has been suggested that clinical RPE detachments may result from a breakdown of the transport mechanisms for fluid across the RPE as a result of focal damage.

It is clear, therefore, that any process that impairs retinol transport into the RPE from central stores in the liver will affect vision; because vitamin A requirements are supplied through the diet, any condition affecting this, such as protein–calorie malnutrition or one of the malabsorption syndromes, may produce visual symptoms. One of the earliest symptoms of severe malnutrition, endemic in developing countries, is night blindness caused by lack of vitamin A.

The RPE also has enzymes of the cytochrome P_{450} drug-metabolizing system and the associated ArH pathway (Heo, et.al. 2011). In addition, RPE melanin is effective in drug detoxification but its content decreases with age.

The chemistry of the visual response

The main function of the retina is to convert light energy into an interpretable signal for cortical cells in the brain. This process begins with photochemical events in the photoreceptor.

Photochemical reactions in the retina

Rhodopsin, vitamin A and photoreceptor turnover

Rhodopsin, a 348 amino acid long major integral cell membrane protein in rod photoreceptor outer segments, is synthesized in the ER and Golgi apparatus of the inner segments and transported in protein-rich vesicles to the outer segment, where fusion occurs with the newly formed disc membranes in the periciliary ridge complex (see ciliopathies, p. 257). It has three glycosylation sites containing various branched combinations of *N*-acetylglucosamine and mannose residues. Opsin is a seven-turn (α helix) membrane-spanning protein containing a serine- and threonine-rich cytosol-exposed C-terminus, which is variably phosphorylated, and an intradisc N-terminus (Fig. 4.13). This structure follows the general pattern for G protein types of membrane receptor, which induce cell signalling on binding their specific ligand by activating adenyl cyclase to raise intracellular concentrations of the second messenger cAMP (see p. 164 and Chapter 6, p. 355). The ligand for rhodopsin, 11-*cis* retinal, is already bound in the dark and dissociates when activated by a photon of light.

Rhodopsin behaves like a genuine receptor, but with differences: in the resting state (i.e. in the dark), Na^+ channels in the outer segment plasma membrane are held 'open' by cGMP, synthesized by guanylate cyclase. This causes a relative depolarization of the outer segment compared with other cells (−57 versus −78 mV). Light activates rhodopsin causing transdisk membrane signalling via sequential activation of other membrane-bound proteins, transducin and phosphodiesterase, thereby lowering cytosolic cGMP, i.e. it acts like a second messenger in reverse. This has the effect of closing the leaky Na^+ channels in the plasma membrane and causing a relative hyperpolarization (to −87 mV), thereby generating the electrical response (see later).

Activation of rhodopsin is achieved via isomerization of retinol, a vitamin A compound that lies 'nested' between the first and last transmembrane loops of the rhodopsin molecule, with its long axis in the plane of the membrane (Fig. 4.13). Modelling studies show the relationship between the chromophore pocket and the opsin molecule (Figs 4.13 and 4.88). In this reaction, the tail of 11-*cis* retinal on conversion to all-*trans* retinal becomes elongated and more perpendicularly disposed to the isoprene retinal ring structure; this has the effect of generating greater interaction between the retinal and its binding sites to the side-chains of amino acids Lys206 and Lys296, thereby heightening the energy state of the rhodopsin molecule. All-*trans* retinal becomes converted to all-*trans* retinol, which does not fit within the rhodopsin transmembrane loops and it is during this stage that rhodopsin converts through its various intermediaries to opsin, the changes being demonstrable by colour bleaching (Fig. 4.88). The chromophore detaches from the bleached opsin and diffuses away into the interphotoreceptor matrix to be taken up by the RPE cell; in the dark it undergoes isomerization within the RPE cell (see Figs 4.84 and 4.86) to 11-*cis* retinol, which binds to cellular retinal binding protein (CRALBP) and becomes converted to 11-*cis* retinal-CRALBP, which is transported to the cell membrane and is transferred to interphotoreceptor retinal binding protein (IRBP) on which it is shuttled back to the photoreceptor to reattach to rhodopsin and recommence the cycle.

The process of 11-*cis* retinal binding to opsin occurs by the formation of a Schiff base to Lys296, releasing a single molecule of water. Interestingly, 'ordered' water molecules are also included in the overall conformation of rhodopsin as tightly bound 'prosthetic' structural moieties giving cohesion to the overall structure of the molecule (see Fig. 4.13). It is possible that these water molecules are held to the structure through electrochemical bonds that become disrupted when light converts 11-*cis* retinal to all-*trans* retinal, thus creating space for the chromophore to detach.

There are three cone opsins in humans that have around 50% homology to rhodopsin and to each other apart from the red and green opsins, which are 95% homologous to each other (Fig. 4.89). The release of all-*trans* retinal from cone pigments is much faster than from rhodopsin and even binding of *cis*-retinal is less stable, possibly related to the amount of water binding.

The conversion of 11-*cis* retinal to all-*trans* retinal and then to all-*trans* retinol is the fundamental chemical reaction to take place during the visual impulse, and the chemical reaction has been known for many years. Laser flash photolysis studies have chemically identified the intermediates that occur in the breakdown of vitamin A (see Figs 4.86 and 4.88). Normal vision depends on a plentiful supply of vitamin A, which must be provided exogenously because humans cannot synthesize it. Dietary vitamin A comes from β-carotene in plants, while animal sources provide the chromophore as a retinyl ester linked to fatty acids. Thus, visual deterioration, especially at low illumination (night blindness), is an early sign of malnutrition and malabsorption syndromes, but sensitive electrophysiological tests may be required to reveal its full extent. Retinoids are stored in the RPE as non-toxic retinyl esters; retinoic acid, which can also be synthesized during retinoid metabolism, has many functions in cell biology and is known to have a role in immune tolerance, but free retinoic acid is toxic to cells and thus has to be generated from retinyl esters, similar to visual retinoids. In the retina, all-*trans* retinol to 11-*cis* retinol isomerization is regulated by palmitoylation under the control of the RPE-specific protein RPE-65 (retinoid isomerohydrolase, retinol isomerase) (see Fig. 4.86) whose role serves to switch off the visual cycle in the dark.

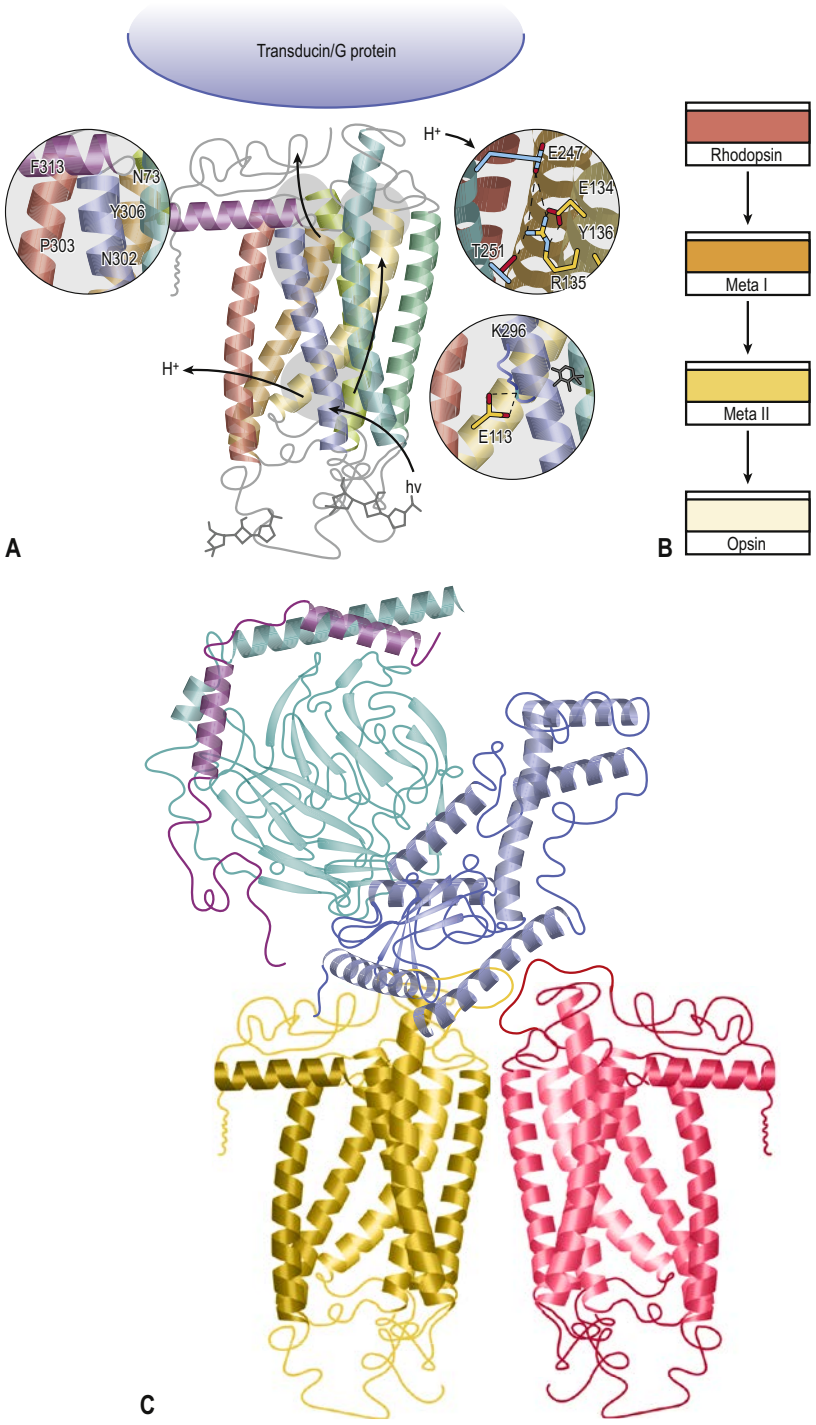


Fig. 4.88 Activation of rhodopsin by light. (**A**) Rhodopsin absorbs a photon of light, which leads to isomerization of 11-cis retinal to all-trans retinal. Transmission of this signal to the cytoplasmic surface triggers nucleotide exchange on the heterotrimeric G protein ransducing. Transducin dissociates and activates the downstream signalling events. (**B**) Spectral changes in rhodopsin upon activation as detected by photography. Once rhodopsin in its dark 11-cis retinal-bound state (**A**) is exposed to light, it immediately goes through a series of photointermediate states, including metarhodopsin I (Meta I) (**B**), and eventually progressing to the Rho+ (metarhodopsin II (Meta II)) activated state (**C**). Upon treatment with hydroxylamine, the chromophore is hydrolysed, resulting in a largely colourless solution. (**C**) Shows a model of the G protein rhodopsin complex. (From Palczewski, 2012.)

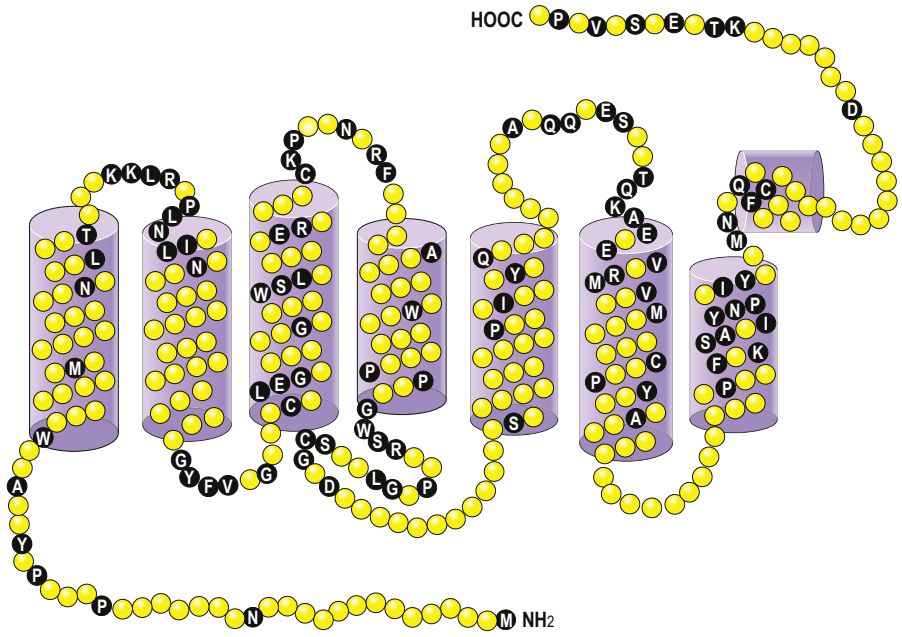


Fig. 4.89 Sequence identity among the human rod and cone opsins. Opsin secondary structure is represented in two dimensions based on alignment with the known structure of bovine rhodopsin. Residues that are identical between the human rod and human red, green and blue cone opsins are denoted as *black circles* with the conserved amino acid residues identified. (From Tang et al., 2013.)

Phototransduction: the conversion of light energy to an electrochemical response

The conversion of the energy stored in a single photon of light to an electrical response is possible because of the extensive amplification of the molecular cascade involved in closure of the Na^+ channels.

Further information on phototransduction and the Na^+ channel is available online at https://expertconsult.inkling.com

These channels are kept open by cGMP (termed cyclic nucleotide gated channels, CNGP), which acts as a second messenger in this system.

In the dark, open Na^+ channels allow Na^+ to exit and Ca^{2+} to enter and to maintain a relative depolarization. This is linked to a Na^+/Ca^{2+} exchanger, which maintains a steady intracellular concentration of Ca^{2+}. In the light, the Na^+ channels are closed, leading to a relative hyperpolarization and a decline in the intracellular Ca^{2+} concentration. Closure of the Na^+ channels is achieved by hydrolysis of cGMP by phosphodiesterase (PDE), one of the responses that is greatly amplified in the cascade. Signal amplification is achieved by activation of many molecules of transducin by one molecule of metarhodopsin (R*), probably in the region of 300× (Fig. 4.90).

Further information on phototransduction and the Na^+ channel is available online at https://expertconsult.inkling.com

In this sense, activated rhodopsin acts like an enzyme since it is available to activate further molecules of transducin until the light is switched off by the combined action of recoverin (rhodopsin kinase) and arrestin (also known as S antigen). Similarly, around 600 molecules of cGMP are hydrolysed by a single molecule of PDE and, indeed, the limitation of its activity is determined by the availability of cGMP. There is a further three-fold amplification of channel opening by one cGMP molecule.

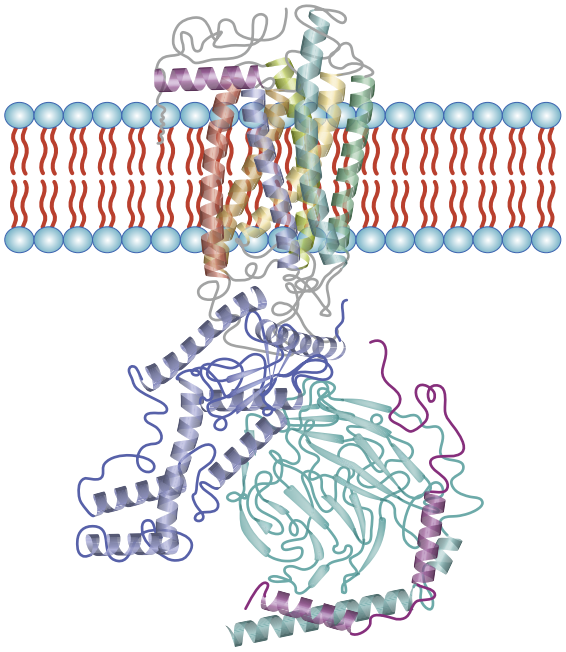
Phototransduction is a biological cascade

The mechanism of cGMP hydrolysis in phototransduction has been extensively investigated; the sequence of events in this response is outlined in Fig. 4.91.

Certain enzymes play important roles in this process, including guanylate cyclase and cGMP phosphodiesterase. Guanylate cyclase activation restores cGMP levels with the help of recoverin while activated rhodopsin (R*) is deactivated to the phosphorylated form R–P and bound by arrestin (S antigen). Recoverin (Rcv) has considerable effect on re-activation and phosphorylation and is dependent itself

The phototransduction cascade and channel opening

Activation of a single molecule of rhodopsin generates an amplification cascade that leads to the opening of many channels and the induction of a change in the resting potential of the photoreceptor (eFig. 4.25).



eFig. 4.25 Type 2 rhodopsin (Rh) *(rainbow coloured)* embedded in a lipid bilayer (heads *red* and tails *blue*) with transducin (GDP) below. $G_{t\alpha}$ is coloured *red*, $G_{t\beta}$ *blue* and $G_{t\gamma}$ *yellow*. There is a bound GDP molecule in the $G_{t\alpha}$-subunit and a bound retinal *(black)* in the rhodopsin. The N-terminus of rhodopsin is *red* and the C-terminus *blue*. Anchoring of transducin to the membrane has been drawn in *black*. *PDE*, Phosphodiesterase.

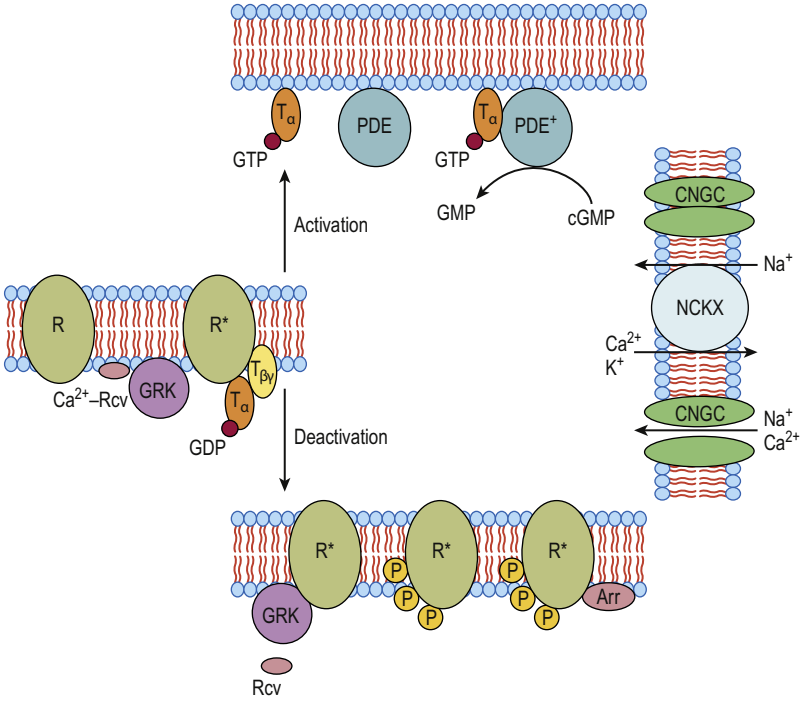


Fig. 4.90 Schematic representation of the vertebrate phototransduction cascade and the deactivation process for activated rhodopsin (R^+) regulated by recoverin *(Rcv)*. Photon absorption activates R. R^* then triggers the exchange of *GTP* for *GDP* on the ransducing alpha subunit *(Tα)*. Tα-GTP binds to cyclic nucleotide phosphodiesterase *(PDE)*. Stimulated PDE hydrolyzes free cyclic guanosine monophosphate *(cGMP)*. In darkness, cyclic nucleotide-gated ion channels *(CNGC)* allow an influx of Na^+ and $Ca2^{+,}$ while during illumination CNGC is shut off by cGMP decrease. The sodium-potassium exchanger (NCKX) is not affected by light, which results in a light-induced [$Ca2^+$]© decline. Rcv modulates phosphorylation of R^+ via GRK in a $Ca2^+$-dependent manner. Phosphorylated R then is fully deactivated by binding of arrestin *(Arr)*. *P,* Phosphorylation. (From Zang, 2018.)

on the amount of bound Ca++ ions, regulated by Na/Ca exchangers and Na/Ca,K exchangers (NCKX), the last of which operates independently of the light response. During these exchanges there is considerable redistribution of photoreceptor molecules in the shift from light to dark (Fig. 4.92). Amplification is therefore a function of the time interval at each of the stages of processing.

Does phosphoinositide metabolism have a role in phototransduction?

The importance of Ca++ ions in regulating phototransduction is now recognized and is activated via receptor-mediated activation of phospholipase C (PLC) and G (PLG) proteins. This results in the hydrolysis of phosphatidylinositol 4,5-bisphosphate (PIP_2) to form inositol triphosphate (IP_3) and diacylglycerol (DAG). IP_3 induces mobilization of calcium stores, while DAG activates protein kinase C. Two major calcium-binding proteins, calmodulin and GCAP1 and 2 (calcium-dependent modulator of guanylate cyclase), are involved in phototransduction. While Ca++ ions regulate recoverin function, PKC and PLC regulate the translocation of arrestin from the inner to the outer segments in the dark (Fig. 4.92).

Phototransduction in rods and cones is essentially the same in rods and cones. Depolarization of the plasma membrane (by inflow of Na+ and Ca++) (see Fig. 4.90) in the dark is maintained by modified cGMP gated ion channels (CNG-A1/B1 in rods and CNG-A3/B3 in cones), which remain open to ion exchange under control of varying levels of disc membrane cGMP produced by retinal guanylyl cyclases (GC1/2 in rods, GC1 in cones). Activation of opsin (R in rods, cone opsins in cones) and then transducin (Gt1 in rods, Gt2 in cones) followed by phosphodiesterase (PDE6A/B in rods, PDE6C in cones) results in hydrolysis of cGMP and closure of the CNG channels with consequent hyperpolarization. Ca++ ion binding to GC –activating proteins (GCAPs) stops GC activation and is itself regulated by rod and cone specific NCKXs, which respond to the overall concentration of Ca++ in the photoreceptor.

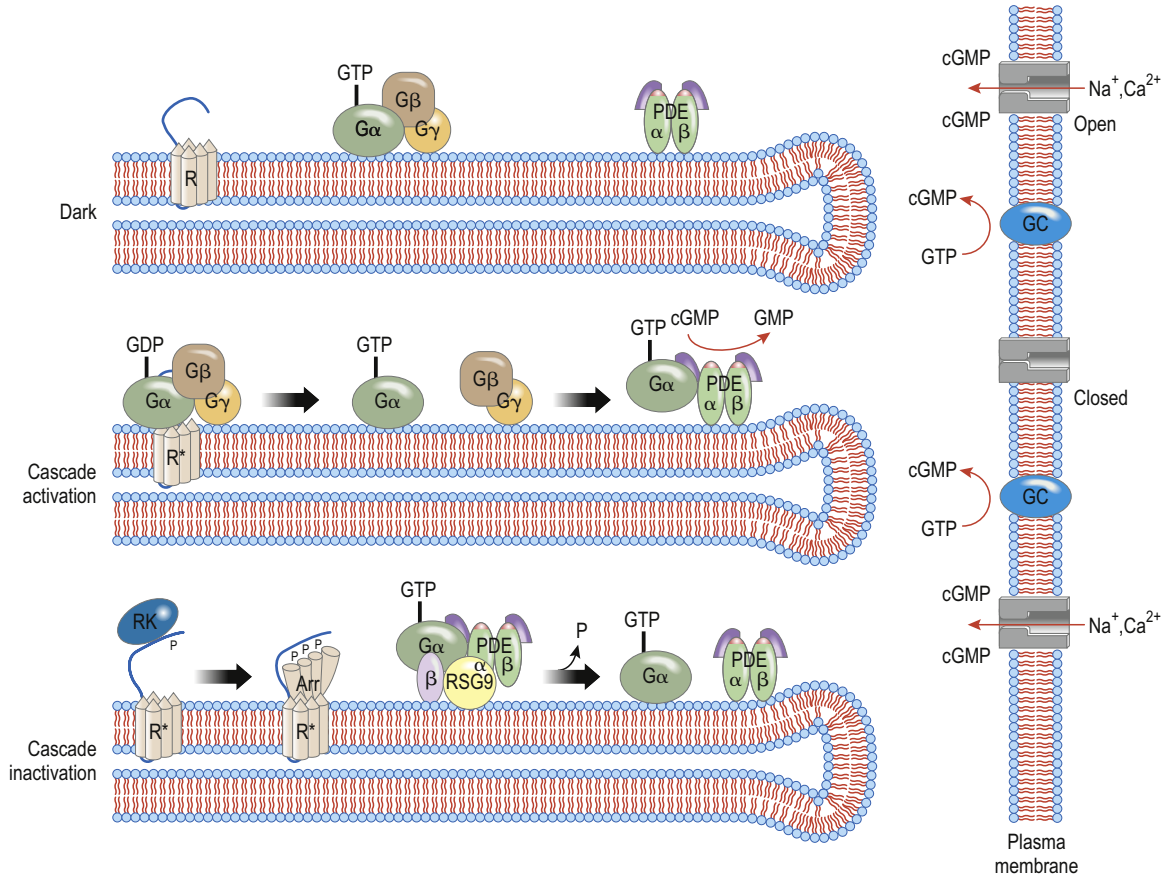


Fig. 4.91 Schematic of phototransduction cascade activation and inactivation. The *upper disk* illustrates inactive rhodopsin ©, transducing (*Gα, Gβ and Gγ* subunits), and PDE (*α, β and γ* subunits), in the dark. The reactions in the *middle disk* illustrate light-induced ransducing and PDE activation. The reactions in the *lower disk* represent *R** inactivation via phosphorylation by rhodopsin kinase *(RK)* followed by arrestin *(Arr)* binding and ransducing/PDE inactivation by RGS9–Gαβ-PDEαβ complex. (From Burns et al., 2005, with permission from Elsevier.)

Renewal of photoreceptors is associated with accumulation of lipofuscin pigment in the RPE cells

Estimates suggest that RPE cells phagocytose about 2000–4000 discs per day. To cope with this phagocytic load, the RPE cell has an extensive lysosomal enzyme system that can digest about 50% of its load within 1–2 hours (important in terms of cyclic rod outer segment tip shedding). Uptake of shed rod outer segment tips is receptor mediated, through the $\alpha_v\beta_5$ integrin receptor signalling, the receptor tyrosine kinase Mer (Mertk) and CD36 (see p. 260). CD36 is expressed on many cells and has many ligands including fatty acids. CD36 recognizes oxidized phosphatidylcholine, a product of aged outer segment tips, and may be involved in the engulfment process rather than initial binding. Interestingly, the $\alpha_v\beta_5$ receptor is also involved in retinal adhesion to the RPE layer (see p. 260), functioning in a different diurnal cycle from phagocytosis.

With age, accumulation of retinoid material and lipofuscin occurs in the RPE. Free all-*trans* retinal, produced after the photoactivation of rhodopsin by light, is considered to bind to phosphatidylethanolamine (PE), forming N-retinylidene-PE (NRPE) from which N-retinylidene-N-retinylphosphatidylethanolamine (A2PE) is generated by the binding of a second molecule of all-*trans* retinal. Phospholipase D then cleaves PE from A2PE to generate A2E. A2E compounds are piridinium retinoid salts, including the retinoid 11-cis-retinal. They are chemically reactive and have been identified as a potential culprits in lipofuscin responsible for AMD.

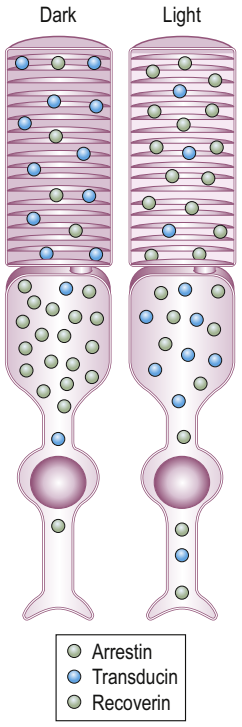


Fig. 4.92 Schematic representation of ransducing, arrestin and recoverin subcellular distribution in the dark-adapted and light-adapted rod. (From Burns et al., 2005, with permission from Elsevier.)

Their damaging effects are exacerbated in the presence of intracellular iron.

Other compounds include malondialdehyde and 4-hydroxy nonenal as well as carboxyethyl-pyrroles, derivatives of docosahexaenoic acid (DHA) a major fatty acid in photoreceptors (see p. 253), which is implicated in the pathogenesis of AMD.

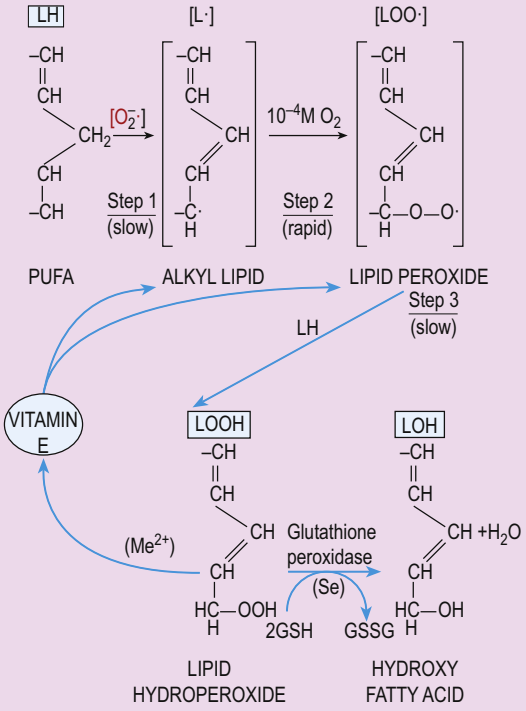
Photoreceptors are easily damaged, particularly by excess light radiation, and several free radical scavenging mechanisms are in place to minimize the damage. In addition, vitamin E has a significant role as a free radical scavenger (see Box 4.17).

Autophagy and mitophagy process waste material in the RPE for removal

Clearance of cellular waste material is mediated by autophagy (see p. 260), in which the autophagosome is generated by fusion with lysosomes. Engulfed photoreceptor outer segments are targeted by the phagophore for removal under the control of the protein p62/sequestrone, which links autophagy with the ubiquitin-proteasome pathway.

Box 4.17 **Vitamin E as a free radical scavenger**

Vitamin E limits free radical damage at several stages in lipid peroxidation.



Regeneration of vitamin E under light stress requires ascorbic acid. In experimental vitamin E deficiency, there is a marked accumulation of lipofuscin in the RPE, which may be relevant to a similar accumulation of lipofuscin in age-related macular degeneration. *PUFA*, Polyunsaturated fatty acid.

Lipofuscin granules are considered partially lysosome-digested photoreceptor waste material, which may also fuse with melanosomes to form melanolipofuscin bodies. With age, and experimentally in vitamin E deficiency, there is a greater shift towards fusion with lipofuscin granules. In addition, a process termed mitophagy, which may fail with age, also removes ageing or damaged mitochondria. Eventually these bodies are degraded, solubilized and their products are transported out of the cell often as microvesicle or exosomes. The accumulation of vesicular material between the basement membrane and the RPE cell has been attributed to a breakdown in the capacity of the cell to deal with this load. These basal liner deposits are an early pathological sign of AMD.

Table 4.10 **Retinal neurotransmitters and neuromodulatory peptides**

Agent	Site of action
Neurotransmitters	
Glutamate	Photoreceptor
γ-aminobutyric acid	Horizontal amacrine cells
Glycine	Amacrine cell
Taurine	
Tyrosine derivatives	
Dopamine	Horizontal amacrine cells
Noradrenaline	
Adrenaline	
Tryptophan derivatives	
Serotonin	Photoreceptor
Aspartate	Photoreceptor
β-alanine	
Histidine derivative	
Histamine	
Neuromodulators	
Peptides	
Vasoactive intestinal peptide	Amacrine, ganglion cells
Angiotensin I and II	Amacrine, ganglion cells
Substance P	Amacrine, ganglion cells
Luteinizing hormone releasing hormone and thyrotrophin releasing hormone	Amacrine, ganglion cells
Leu- and metenkephalin	Amacrine cell
β-endorphin	Amacrine, ganglion cells
Somatostatin	Amacrine, ganglion cells
Neurotensin	Amacrine cell
Glucagon	Amacrine cell

Synaptic events between photoreceptors and cells of the inner nuclear layer

Bipolar cells of the inner nuclear layer (INL) synapse either with rod or with cone photoreceptors but not with both. In addition, they synapse with other cells in the INL such as horizontal cells. There are around 10 different types of cone bipolar cells but only one type of rod bipolar cell. Several types of neurotransmitter may be involved (Table 4.10). In the resting state, the dark currents produced by the open cation channels in the outer segment (see Chapter 5, Box 5.6, p. 294) are accompanied by high levels of neurotransmitter release at the synaptic junction with the bipolar cells. In the light, there is a decrease in transmitter release, which alters the transmission of electric potentials in the bipolar cells. However, this simple description hides a deal of complexity.

Two types of bipolar cell

Synaptic transmission at neuromuscular junctions is mediated by acetylcholine, released from the nerve ending when the action potential wave arrives. Binding of acetylcholine to its receptor on the muscle initiates a second message, which causes a depolarization in the cell owing to opening of the voltage-gated Na^+ channels.

At the synaptic junction between the photoreceptor cell and the bipolar cell, a different type of reaction occurs. First, the photoreceptor does not transmit an action potential wave, but presents a graded hyperpolarization response, which depends on the intensity of the light stimulus. In the dark, while the photoreceptor is in a relatively depolarized state, the neurotransmitter glutamate (see Table 4.10) is released from the presynaptic photoreceptor terminal and binds to a retina-specific metabotropic receptor (metabolic function-based receptor) on one type of bipolar cell (the ON-bipolar), which produces hyperpolarization and keeps the bipolar Na^+ channel closed. When light activates the photoreceptor, glutamate release ceases and the intracellular concentration of cGMP rises, thus opening the Na^+ channels and causing depolarization of the cell membrane, allowing the ON-bipolar cell to transmit a signal.

In contrast, glutamate binding to the ionotropic receptor (ion channel-based receptor) on a second type of bipolar cell (the OFF-bipolar; see Chapter 5 p. 296) has the reverse effect of inducing hyperpolarization. To summarize, therefore, when light activates the photoreceptor and transmits the message to the bipolar cell, ON-bipolar cells are depolarized by reduction in glutamate release and OFF-bipolar cells are hyperpolarized by an increase in glutamate release. Additionally, glutamate may bind to horizontal cells, where it also produces hyperpolarization.

This type of graded response is achieved by specialized photoreceptor structures termed ribbon synapses, which are exquisitely organized to transmit information. Several membrane proteins are involved in the cytoskeletal structure and function of the synapse, such as dystrophin and retinoschisin, which have clinical relevance (mutations in retinoschisin cause retinoschisis, a retinal degeneration involving splitting of the retinal layers) and the intersynaptic cleft, a space containing other essential matrix components (Fig. 4.93). In both ON- and OFF-bipolar cells, the action of glutamate is considered excitatory, since there is a defined response. Glutamate is unlike acetylcholine in that there is no enzyme that rapidly degrades the neurotransmitter and shuts down the response; instead, the glutamate merely diffuses away or is taken up by nearby glial cells and inactivated.

The differential binding of glutamate to ON- and OFF-bipolar cells via specific receptors is a direct chemical correlate of the electrophysiological and psychophysical responses that can be obtained from these cells and is a

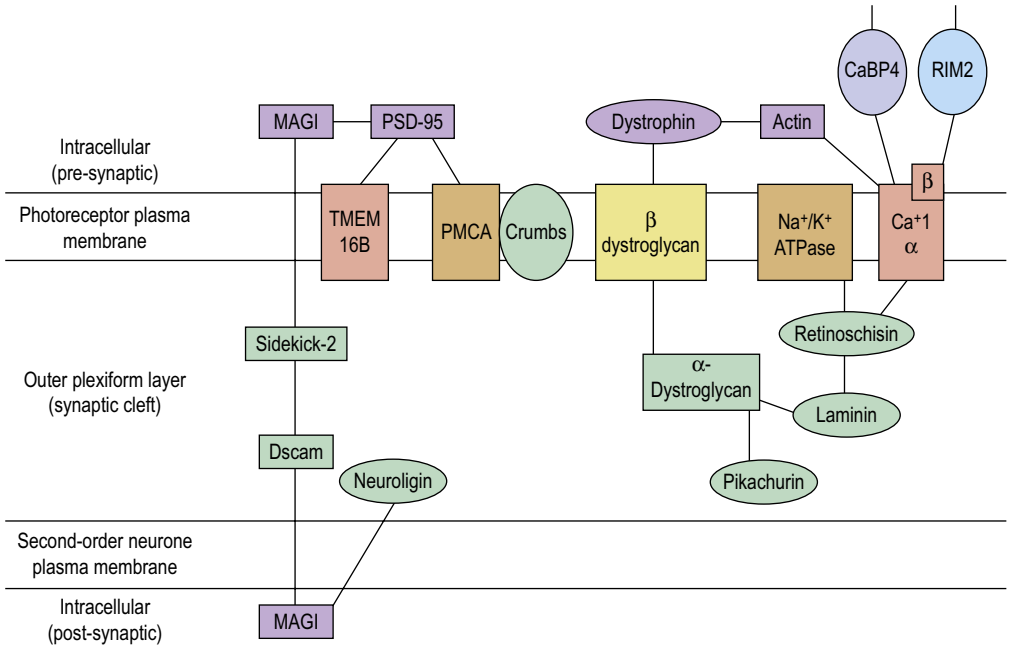


Fig. 4.93 Proteins central to the organization of the outer plexiform layer (OPL). Two major nodes of protein–protein interactions occur in the extracellular matrix (ECM) of the OPL ECM within the OPL. The sidekick-2 trans-synaptic scaffold attaches MAGI proteins in apposed cells, with putative lateral presynaptic interactions to PSD-95. PSD-95 also appears to anchor the Cl(Ca) channel. TMEM16B and PMCA extrusion pump at non-ribbon sites in the terminal. Scaffolding interactions between a transmembrane dystroglycan complex and the retina-specific ECM proteins pikachurin and retinoschisin, linked through synaptic laminin and cytoplasmic actin, while also attaching to the extracellular face of transmembrane CaV channels and the Na^+/K^+-ATPase. (From Mercer and Thoreson, 2011.)

good demonstration of nature's use of a binary (ON–OFF) system to process complex sensory phenomena. The segregation of responses at the bipolar cell level is retained to the level of the visual cortex.

Neurotransmitters and neuromodulators

In contrast to glutamate, other neurotransmitters such as glycine and GABA have a lateral inhibitory action at the synaptic junction and are considered to be neuromodulators, i.e. they modulate the effect of glutamate; neuromodulators have a much longer action, often mediated through second messenger systems in the cell. There are many different types of neurotransmitter, most of which are amino acids or amino acid derivatives, and the majority are active in transmission of electric impulses within the mammalian retina. As indicated above, typical excitatory transmitters include acetylcholine and glutamate, while γ-aminobutyric acid (GABA) and glycine are inhibitory to their function. Several other neuromodulatory peptides exist that have direct or indirect effects on neural responses. Dopamine is probably the best known of these (see later). Neuronal nitric oxide synthase (nNOS) is present and active in amacrine cells but recently type 9 bipolar cells (see later) have been shown to express nNOS also. Nitric oxide may therefore modulate retinal signalling.

Glutamate is the major neurotransmitter in the retina

As indicated previously, glutamate (Fig. 4.94) is the major mediator of synaptic transmission between the photoreceptor and the bipolar and horizontal cells. The functional organization of the retina is centred around the specificity of glutamate for four discrete types of glutamate receptor on different types of cells (ON- and OFF-bipolar cells and two types of horizontal cell). Glutamate receptors are classified as ionotropic (iGluR) and metabotropic (mGluR). iGluR are by far the more ubiquitous and are further grouped as:

- *N*-methyl-D-aspartate (NMDA) selective receptors
- quisqualate receptors
- AMPA/kainate receptors on hyperpolarizing bipolar cells
- APB (sign-inverting glutamate receptor agonist) receptors.

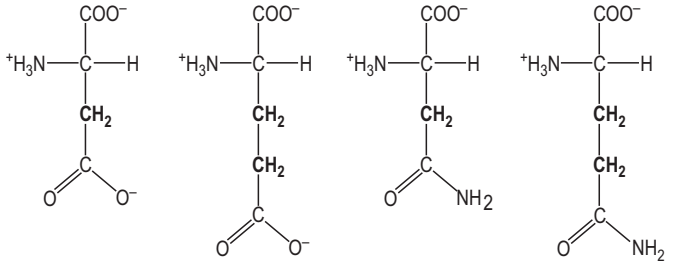


Fig. 4.94 Chemical structure of the acidic amino acids aspartate, glutamate, asparagine and glutamine.

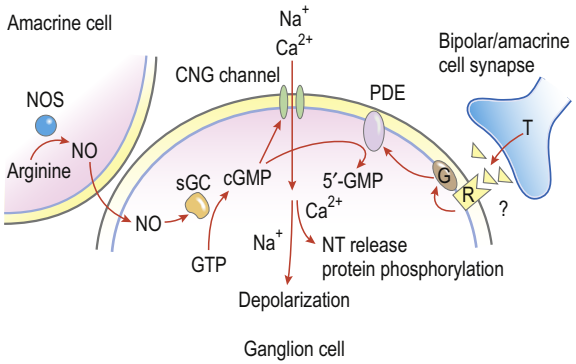


Fig. 4.95 Schematic representation of pathways that might regulate the activity of the cGMP-gated channel in retinal ganglion cells. The cGMP synthesis could be stimulated by nitric oxide adjacent amacrine cells. Activation of cGMP-gated channels will increase Ca^{2+} influx and enhance Ca^{2+}-driven processes. At the same time, G protein-coupled receptors activated by neurotransmitters from bipolar and amacrine cell terminals may regulate the activity of one or more PDEs which control hydrolysis of cGMP and thus the activity of the cGMP-gated channels. (From Barnstable et al., 2004; modified from Ahmad et al., 1994.)

Each of these receptors has specific characteristics related to whether it is ionotropic (directly affecting ion channels, e.g. NMDA) or metabotropic (primarily affecting aspects of neuronal cell metabolism, such as cGMP levels, and secondarily altering ion-channel permeability, e.g. APB receptors) (see Fig. 5.12, p. 298). Synaptic transmission in ganglion cells may also use glutamate. However, small cyclic nucleotides may confer gating activity (Fig. 4.95) and cAMP and cGMP have opposing gating effects in their interactions with horizontal/amacrine cells and ganglion cells in the regulation of Ca^{2+} influx. In addition, they appear to be modulated by nNOS derived from neighbouring amacrine cells. NO in the retina may be important in visual adaptation under normal ambient demands (Walter et. al., 2014).

Horizontal and amacrine cells

Horizontal and amacrine cells directly modify the rate of electrical firing in bipolar cells by release of excitatory (horizontal cells) and inhibitory (horizontal and amacrine cells) neurotransmitters. Glutamate is the major excitatory transmitter of horizontal cells and GABA, a derivative of glutamate distributed widely throughout the retina, functions particularly in the horizontal and amacrine cells as an inhibitory mediator. Some horizontal cell negative feedback to cones can be mediated via glutamate-gated Ca^{2+} channels. Glycine is also an inhibitory neurotransmitter in amacrine cells, which have a significant role in determining the size of receptive fields of individual ganglion cells (see Chapter 5 p. 297).

Acetylcholine is a major excitatory neurotransmitter of amacrine cells and often co-localizes in GABAergic starburst cells, suggesting that these cells can act in a complex excitatory/inhibitory manner in relation to ganglion cell receptive fields. Ganglion cells use glutamate as a major transmitter.

Amacrine cells also receive input from melanopsin-type (OPN4) in the retina. Intrinsically photosensitive retinal ganglion cells (ipRGCs, see p. 256) excite a subset of amacrine cells via dopamine receptors and are coupled to a further subset of GABA-responsive amacrine cells via gap junctions. In this way, information regarding irradiance, circadian rhythms and pupillary responses is integrated with primary photoreceptor responses. Neuropsin (OPN5)-sensitive cells also signal to amacrine cells but there is minimal excitatory light response in mammals. OPN5 may have a role in avian UV light response.

Dopamine exerts a neuromodulatory effect on retinal function

Dopamine occurs in amacrine cells and inner plexiform layer cells in the inner retinal layers and clearly has a role to play in visual function because the reduced and delayed b-wave electroretinogram responses (see Chapter 5 p. 305)

seen in patients with Parkinson's disease are normalized on treatment with L-DOPA. Dopamine receptors (there are two broad classes and several subclasses) are present on all retinal neurones and therefore dopamine probably has many as-yet undetermined effects, one of which is to uncouple photoreceptor-driven horizontal cell gap junctions. Horizontal cells are probably the source of activity in the surround region in bipolar cell centre-surround receptive fields, suggesting that regulation of spatial contrast sensitivity may be under dopaminergic control.

Dopamine may have many other effects including regulation of A11-type amacrine cells, which couple rod-associated bipolar cells and ganglion cells, and it may even have a paracrine role in altering the level of photoreceptor cytoplasmic cAMP. Dopamine has also been implicated in development in retinal neurones because it has been shown that apomorphine inhibits form-deprivation myopia in experimental animals by a dopamine D_2-receptor mechanism acting within either the retina or the RPE.

Dopamine is inactivated by cellular re-uptake, a process mediated by specific Na^+-dependent membrane transporter proteins, which have been cloned. Dopamine is antagonized by melatonin, which is synthesized and released by photoreceptors, accounting for the photopic to scotopic transition in rod responsiveness. It is also counteracted by NMDA, which may induce impaired dopamine synthesis. Interestingly, melatonin has neuromodulatory effects on dopamine-sensitive cells as well as on photoreceptors (Fig. 4.91). The role of melatonin, dopamine and circadian rhythms has been addressed above (see Chapter 5, p 283).

It can therefore be seen that, while much remains to be determined concerning the mediators of neural function in the retina and visual cortex, a view is already emerging of differential stimulation and inhibition of function induced by a variety of interacting neuromodulators and transmitters, which correlates with observed psychophysical events.

Conclusion

While the eye has many 'eye-specific' biochemical and cell biological functions and roles, many of these are founded on general principles that apply to other systems and, indeed, comparison of ocular and non-ocular processes offers much insight into basic fundamental principles. Where differences exist, they assist in highlighting the properties and functions of molecules and reactions, while introducing us to the specific mechanisms in the eye. In this way, a greater understanding of the general principles is achieved.

A list of further reading and references is available online at https://expertconsult.inkling.com

References

Ahmed, I., Akram, Z., Iqbal, H.M.N., Munn, A.L., 2019. The regulation of Endosomal Sorting Complex Required for Transport and accessory proteins in multivesicular body sorting and enveloped viral budding – an overview. Int. J. Biol. Macromol. https://doi.org/10.1016/j.ijbiomac.2019.01.015.

Alberts, B., et al., 2015. Molecular Biology of the Cell, Chapter 3, sixth ed. Garland Science, New York, p. 159.

Baehr, W., et al., 2019. Insights into photoreceptor ciliogenesis revealed by animal models. Prog. Retin. Eye Res 71, 26–56. https://doi.org/10.1016/j.preteyeres.2018.12.004.

Braun, R.J., Driscoll, T.A., Begley, C.G., King-Smith, P.E., Siddique, J.I., 2018. On tear film breakup (TBU): dynamics and imaging. Math. Med. Biol 35, 145–180. https://doi.org/10.1093/imammb/dqw023.

Chang, K.C., Petrash, J.M., 2018. Aldo-keto reductases: multifunctional proteins as therapeutic targets in diabetes and inflammatory disease. In: Vasiliou, V., Zakhari, S., Mishra, L., Seitz, H. (Eds.), Alcohol and Cancer. Advances in Experimental Medicine and Biology, vol 1032. Springer.

Cheah, M., Andrews, M.R., 2018. Integrin activation: implications for axon regeneration. Cells 7. https://doi.org/10.3390/cells7030020.

Cohen, S.A., Hatt, H., Kubanek, J., McCarty, N.A., 2008. Reconstitution of a chemical defense signaling pathway in a heterologous system. J. Exp. Biol. 211, 599–605. https://doi.org/10.1242/jeb.009225.

Cordes, T., Michelucci, A., Hiller, K., 2015. Itaconic acid: the surprising role of an industrial compound as a mammalian antimicrobial metabolite. Annu. Rev. Nutr. 35, 451–473. https://doi.org/10.1146/annurev-nutr-071714-034243.

Crabtree, M., Knott, T., 2018. From making ribosomes to protecting the integrity of the genome, these membraneless compartments play important roles in the cell. Their behavior is rooted in basic physics. The Scientist (Issue), 2018.

Crane, G.M., Jeffery, E., Morrison, S.J., 2017. Adult haematopoietic stem cell niches. Nat. Rev. Immunol. 17, 573–590. https://doi.org/10.1038/nri.2017.53.

Csordas, G., Weaver, D., Hajnoczky, G., 2018. Endoplasmic reticulum-mitochondrial contactology structure and signaling functions. Trends Cell. Biol. 28, 523-540, doi:10.1016/j.tcb.2018.02.009.

Dua, H.S., 2019. Dua's Layer (Personal communications).

Elder, M., Srinivasan, S. Meibomography 2018. https://www.opticianonline.net/cet-archive/4737. February 2018.

Ferrero, D.M., et al., 2013. A juvenile mouse pheromone inhibits sexual behaviour through the vomeronasal system. Nature 502, 368–371. https://doi.org/10.1038/nature12579.

Gogola, A., Jan, N.J., Lathrop, K.L., Sigal, I.A., 2018. Radial and circumferential collagen fibers are a feature of the peripapillary sclera of Human, Monkey, Pig, Cow, Goat, and Sheep. Invest. Ophthalmol. Vis. Sci 59, 4763–4774. https://doi.org/10.1167/iovs.18-25025.

Goldmann, W.H., 2018. Intermediate filaments and cellular mechanics. Cell. Biol. Int. 42, 132–138. https://doi.org/10.1002/cbin.10879.
Han, M.K.L., de Rooij, J., 2016. Converging and unique mechanisms of mechanotransduction at adhesion sites. Trends Cell. Biol. 26, 612–623. https://doi.org/10.1016/j.tcb.2016.03.005.
Jan, N.J., et al., 2018. Crimp around the globe; patterns of collagen crimp across the corneoscleral shell. Exp. Eye Res. 172, 159–170. https://doi.org/10.1016/j.exer.2018.04.003.
Lancaster, M.A., Knoblich, J.A., 2014. Organogenesis in a dish: modeling development and disease using organoid technologies. Science 345, 1247125. https://doi.org/10.1126/science.1247125.
Mainali, L., O'Brien, W.J., Subczynski, W.K., 2018. Detection of cholesterol bilayer domains in intact biological membranes: methodology development and its application to studies of eye lens fiber cell plasma membranes. Exp. Eye Res. 178, 72–81. https://doi.org/10.1016/j.exer.2018.09.020.
Miyagishima, K.J., et al., 2016. In pursuit of authenticity: induced pluripotent stem cell-derived retinal pigment epithelium for clinical applications. Stem. Cells. Transl. Med. 5, 1562–1574. https://doi.org/10.5966/sctm.2016-0037.
Naslavsky, N., Caplan, S., 2018. The enigmatic endosome – sorting the ins and outs of endocytic trafficking. J. Cell. Sci. 131. https://doi.org/10.1242/jcs.216499.
Notara, M., et al., 2018. UV light-blocking contact lenses protect against short-term UVB-induced limbal stem cell niche damage and inflammation. Sci. Rep 8, 12564. https://doi.org/10.1038/s41598-018-30021-8.
Ostrin, L.A., 2019. Ocular and systemic melatonin and the influence of light exposure. Clin. Exp. Optom. 102, 99–108. https://doi.org/10.1111/cxo.12824.
Rowan, S., Bejarano, E., Taylor, A., 2018. Mechanistic targeting of advanced glycation end-products in age-related diseases. Biochim. Biophys. Acta Mol Basis Dis. 1864, 3631–3643. https://doi.org/10.1016/j.bbadis.2018.08.036.
Shindou, H., et al., 2017. Docosahexaenoic acid preserves visual function by maintaining correct disc morphology in retinal photoreceptor cells. J. Biol. Chem. 292, 12054–12064. https://doi.org/10.1074/jbc.M117.790568.
Singh, S.R., et al., 2018. Wide-field choroidal vascularity in healthy eyes. Am. J. Ophthalmol. 193, 100–105. https://doi.org/10.1016/j.ajo.2018.06.016.
Sneddon, M.W., Emonet, T., 2012. Modeling cellular signaling: taking space into the computation. Nat. Methods 9, 239–242. https://doi.org/10.1038/nmeth.1900.
Swafford, A.L., 2015. Cell Membrane: Functions, Role & Structure – Video & Lesson. https://study.com/academy/lesson/cell-membrane-functions-role-structure.html.
Torricelli, A.A., Santhanam, A., Wu, J., Singh, V., Wilson, S.E., 2016. The corneal fibrosis response to epithelial-stromal injury. Exp. Eye Res 142, 110–118. https://doi.org/10.1016/j.exer.2014.09.012.
Tseng, R.C., Chen, C.C., Hsu, S.M., Chuang, H.S., 2018. Contact-Lens Biosensors. Sensors 18. https://doi.org/10.3390/s18082651.
Turei, D., Korcsmaros, T., Saez-Rodriguez, J., 2016. OmniPath: guidelines and gateway for literature-curated signaling pathway resources. Nat. Methods 13, 966–967. https://doi.org/10.1038/nmeth.4077.
Zang, J., Neuhauss, S.C.F., 2018. The binding properties and physiological functions of recoverin. Front. Mol. Neurosci. 11, 473. https://doi.org/10.3389/fnmol.2018.00473.

Further reading

Abdel-Aal el, S.M., Akhtar, H., Zaheer, K., Ali, R., 2013. Dietary sources of lutein and zeaxanthin carotenoids and their role in eye health. Nutrients 5, 1169–1185.

Ahmed, I., Akram, Z., Iqbal, H.M.N., Munn, A.L., 2019. The regulation of Endosomal Sorting Complex Required for Transport and accessory proteins in multivesicular body sorting and enveloped viral budding – an overview. Int. J. Biol. Macromol. https://doi.org/10.1016/j.ijbiomac.2019.01.015.

Alberts, B. et al., 2015. Molecular Biology of the Cell, Chapter 3, sixth ed. Garland Science, New York. page 159.

Ambache, N., 1957. Properties of irin, a physiological constituent of the rabbit's iris. J. Physiol. 135, 114.

Ames, B. N. Prolonging healthy aging: Longevity vitamins and proteins. Proc Natl Acad Sci U S A 115, 10836–10844, https://doi:10.1073/pnas.1809045115 (2018).

Andersen, J.L., Kornbluth, S., 2013. The tangled circuitry of metabolism and apoptosis. Mol. Cell 49, 399–410.

Appenzeller-Herzog, C., Hall, M.N., 2012. Bidirectional crosstalk between endoplasmic reticulum stress and mTOR signaling. Trends Cell. Biol. 22, 274–282.

Asnacios, A., Hamant, O., 2012. The mechanics behind cell polarity. Trends Cell. Biol. 22, 584–591.

Baehr, W., et al., 2018. Insights into photoreceptor ciliogenesis revealed by animal models. Prog. Retin. Eye Res. 2018. https://doi.org/10.1016/j.preteyeres.2018.12.004.

Barabino, S., Chen, Y., Chauhan, S., Dana, R., 2012. Ocular surface immunity: homeostatic mechanisms and their disruption in dry eye disease. Prog. Retin. Eye Res. 31, 271–285.

Bazan, N. G. Docosanoids and elovanoids from omega-3 fatty acids are pro-homeostatic modulators of inflammatory responses, cell damage and neuroprotection. Mol Aspects Med. 64, 18–33, https://doi:10.1016/j.mam.2018.09.003 (2018).

Beebe, D.C., Holekamp, N.M., Siegfried, C., Shui, Y.B., 2011. Vitreoretinal influences on lens function and cataract. Phil. Trans. R. Soc. London B. Biol. Sci. 366, 1293–1300.

Bishop, P.N., 2000. Structural macromolecules and supramolecular organisation of the vitreous gel. Prog. Retin. Eye Res. 19, 323–344.

Bloomfield, S.A., Volgyi, B., 2009. The diverse functional roles and regulation of neuronal gap junctions in the retina. Nat. Rev. Neurosci. 10, 495–506.

Bojic, S., et al., 2019. CD200 expression marks a population of quiescent limbal epithelial stem cells with holoclone forming ability. Stem. Cell 36, 1723–1735. https://doi.org/10.1002/stem.2903.

Bonanno, J.A., 2012. Molecular mechanisms underlying the corneal endothelial pump. Exp. Eye Res. 95, 2–7.

Borchman, D., Yappert, M.C., 2010. Lipids and the ocular lens. J. Lipid Res. 51, 2473–2488.

Bovolenta, P., Esteve, P., Ruiz, J.M., Cisneros, E., Lopez-Rios, J., 2008. Beyond Wnt inhibition: new functions of secreted Frizzled-related proteins in development and disease. J. Cell Sci. 121, 737–746.

Bowne-Anderson, H., Zanic, M., Kauer, M., Howard, J., 2013. Microtubule dynamic instability: a new model with coupled GTP hydrolysis and multistep catastrophe. Bioessays 35, 452–461.

Boyer, N.P., Higbee, D., Currin, M.B., Blakeley, L.R., Chen, C., Ablonczy, Z., et al., 2012. Lipofuscin and N-retinylidene-N-retinylethanolamine (A2E) accumulate in retinal pigment epithelium in absence of light exposure: their origin is 11-cis-retinal. J. Biol. Chem. 287, 22276–22286.

Braun, R.J., Driscoll, T.A., Begley, C.G., King-Smith, P.E., Siddique, J.I., 2018. On tear film breakup (TBU): dynamics and imaging. Math. Med. Biol. 35, 145–180. https://doi.org/10.1093/imammb/dqw023.

Brennan, L.A., et al., 2018. BNIP3L/NIX is required for elimination of mitochondria, endoplasmic reticulum and Golgi apparatus during eye lens organelle-free zone formation. Exp. Eye Res 174, 173–184. https://doi.org/10.1016/j.exer.2018.06.003.

Bresson, E., Lacroix-Pepin, N., Boucher-Kovalik, S., Chapdelaine, P., Fortier, M.A., 2012. The prostaglandin F synthase activity of the human aldose reductase AKR1B1 brings new lenses to look at pathologic conditions. Front. Pharmacol. 3, 98.

Brieher, W.M., Yap, A.S., 2013. Cadherin junctions and their cytoskeleton(s). Curr. Opin. Cell Biol. 25, 39–46.

Brizzi, M.F., Tarone, G., Defilippi, P., 2012. Extracellular matrix, integrins, and growth factors as tailors of the stem cell niche. Curr. Opin. Cell Biol. 24, 645–651.

Brockerhoff, S.E., 2011. Phosphoinositides and photoreceptors. Mol. Neurobiol. 44, 420–425.

Bron, A.J., Yokoi, N., Gaffney, E.A., Tiffany, J.M., 2011. A solute gradient in the tear meniscus. I. A hypothesis to explain Marx's line. Ocul. Surf. 9, 70–91.

Brooke, M.A., Nitoiu, D., Kelsell, D.P., 2012. Cell–cell connectivity: desmosomes and disease. J. Pathol. 226, 158–171.

Bulow, H.E., Hobert, O., 2006. The molecular diversity of glycosaminoglycans shapes animal development. Annu. Rev. Cell Dev. Biol. 22, 375–407.

Burke, J.M., 2008. Epithelial phenotype and the RPE: is the answer blowing in the Wnt? Prog. Retin. Eye Res. 27, 579–595.

Burridge, K., Wittchen, E.S., 2013. The tension mounts: stress fibers as force-generating mechanotransducers. J. Cell Biol. 200, 9–19.

Butovich, I.A., 2011. Lipidomics of human meibomian gland secretions: chemistry, biophysics, and physiological role of meibomian lipids. Prog. Lipid Res. 50, 278–301.

Candia, O.A., Alvarez, L.J., 2008. Fluid transport phenomena in ocular epithelia. Prog. Retin. Eye Res. 27, 197–212.

Cejka, C., et al., 2017. Therapeutic effect of molecular hydrogen in corneal UVB-induced oxidative stress and corneal photodamage. Sci. Rep 7, 18017. https://doi.org/10.1038/s41598-017-18334-6.

Chakraborty, D., Rodgers, K.K., Conley, S.M., Naash, M.I., 2013. Structural characterization of the second intra-discal loop of the photoreceptor tetraspanin RDS. FEBS J. 280, 127–138.

Chang, K.C., Petrash, J.M., 2018. Aldo-keto reductases: multifunctional proteins as therapeutic targets in diabetes and inflammatory disease. In: Vasiliou, V., Zakhari, S., Mishra, L., Seitz, H. (Eds.), Alcohol and Cancer. Advances in Experimental Medicine and Biology, vol 1032. Springer.

Cheah, M., Andrews, M.R., 2018. Integrin Activation: Implications for Axon Regeneration. Cells 7. https://doi.org/10.3390/cells7030020.

Cheema, A.S., Mozayan, A., Channa, P., 2012. Corneal collagen crosslinking in refractive surgery. Curr. Opin. Ophthalmol. 23, 251–256.

Cipriani, V., Leung, H.T., Plagnol, V., Bunce, C., Khan, J.C.,

Shahid, H., et al., 2012. Genome-wide association study of age-related macular degeneration identifies associated variants in the TNXB-FKBPL-NOTCH4 region of chromosome 6p21.3. Human Mol. Gen. 21, 4138–4150.
Claessen, J.H., Kundrat, L., Ploegh, H.L., 2012. Protein quality control in the ER: balancing the ubiquitin checkbook. Trends Cell. Biol. 22, 22–32.
Couchman, J.R., 2010. Transmembrane signaling proteoglycans. Annu. Rev. Cell Dev. Biol. 26, 89–114.
Cohen, S.A., Hatt, H., Kubanek, J., McCarty, N.A., 2008. Reconstitution of a chemical defense signaling pathway in a heterologous system. J. Exp. Biol. 211, 599–605. https://doi.org/10.1242/jeb.009225.
Cordes, T., Michelucci, A., Hiller, K., 2015. Itaconic acid: the surprising role of an industrial compound as a mammalian antimicrobial metabolite. Annu. Rev. Nutr. 35, 451–473. https://doi.org/10.1146/annurev-nutr-071714-034243.
Crabtree, M., Knott, T., 2018. From making ribosomes to protecting the integrity of the genome, these membraneless compartments play important roles in the cell. Their behavior is rooted in basic physics. The Scientist 2018 Dec 1 Issue.
Crane, G.M., Jeffery, E., Morrison, S.J., 2017. Adult haematopoietic stem cell niches. Nat. Rev. Immunol. 17, 573–590. https://doi.org/10.1038/nri.2017.53.
Csordas, G., Weaver, D., Hajnoczky, G. Endoplasmic reticulum-mitochondrial contactology (2018) structure and signaling functions. Trends Cell. Biol. 28, 523-540. doi:10.1016/j.tcb.2018.02.009.
Dagenais, M., Skeldon, A., Saleh, M., 2012. The inflammasome: in memory of Dr. Jurg Tschopp. Cell. Death Differ. 19, 5–12.
Dartt, D.A., 2009. Neural regulation of lacrimal gland secretory processes: relevance in dry eye diseases. Prog. Retin. Eye Res. 28, 155.
Denu, J.M., Gottesfeld, J.M., 2012. Minireview series on sirtuins: from biochemistry to health and disease. J. Biol. Chem. 287, 42417.
Desouza, M., Gunning, P.W., Stehn, J.R., 2012. The actin cytoskeleton as a sensor and mediator of apoptosis. Bioarchitecture 2, 75.
Diamond, J.S., 2011. Calcium-permeable AMPA receptors in the retina. Front. Mol. Neurosci. 4, 27.
Domogatskaya, A., Rodin, S., Tryggvason, K., 2012. Functional diversity of laminins. Annu. Rev. Cell Dev. Biol. 28, 523.
Doughty, M.J., 2012. Goblet cells of the normal human bulbar conjunctiva and their assessment by impression cytology sampling. Ocu. Surf. 10, 149.
Dua, H.S., Faraj, L.A., Said, D.G., Gray, T., Lowe, J., 2013. Human corneal anatomy redefined: a novel pre-Descemet's layer (Dua's layer). Ophthalmology 120, 1778–1785. https://doi.org/10.1016/j.ophtha.2013.01.018.
Dua, H.S. (2019) Dua's Layer (Personal communications).
Edwards, M.M., Lefebvre, O., 2013. Laminins and retinal vascular development. Cell. Adh. Migr. 7, 82.
Efron, N., Hollingsworth, J.G., 2008. New perspectives on keratoconus as revealed by corneal confocal microscopy. Clin. Exp. Optom. 91, 34.
Efron, N., 2017. Contact lens wear is intrinsically inflammatory. Clin. Exp. Optom 100, 3–19. https://doi.org/10.1111/cxo.12487.
Eichmann, A., Simons, M., 2012. VEGF signaling inside vascular endothelial cells and beyond. Curr. Opin. Cell Biol. 24, 188.
Elagouz, M., Stanescu-Segall, D., Jackson, T.L., 2010. Uveal effusion syndrome. Surv. Ophthalmol. 55, 134.
Elder, M., Srinivasan, S., 2018. Meibomography. www.opticianonline.net/cet-archive/4737 February 2018.
Elhawy, E., Kamthan, G., Dong, C.Q., Danias, J., 2012. Pseudoexfoliation syndrome, a systemic disorder with ocular manifestations. Hum. Genom. 6, 22.
Fanaei, M., Monk, P.N., Partridge, L.J., 2011. The role of tetraspanins in fusion. Biochem. Soc. Trans. 39, 524.
Fischbarg, J., 2012. Water channels and their roles in some ocular tissues. Mol. Aspects. Med. 33, 638–641.
Florey, O., Overholtzer, M., 2012. Autophagy proteins in macroendocytic engulfment. Trends Cell. Biol 22, 374.
Freddo, T.F., 2013. A contemporary concept of the blood–aqueous barrier. Prog. Retin. Eye Res. 32, 181.
Forrester, J.V., McMenamin, P.G., Dando, S.J., 2018. CNS infection and immune privilege. Nat. Rev. Neurosci 19, 655–671. https://doi.org/10.1038/s41583-018-0070-8.
Giorgio, M., Trinei, M., Migliaccio, E., Pelicci, P.G., 2007. Hydrogen peroxide: a metabolic by-product or a common mediator of ageing signals? Nat. Rev. Mol. Cell Biol. 8, 722.
Glasgow, B.J., Abduragimov, A.R., 2018. Interaction of ceramides and tear lipocalin. Biochim. Biophys. Acta Mol. Cell Biol. Lipids 1863, 399–408. https://doi.org/10.1016/j.bbalip.2018.01.004.
Gnana-Prakasam, J.P., Martin, P.M., Smith, S.B., Ganapathy, V., 2010. Expression and function of iron-regulatory proteins in retina. IUBMB Life 62, 363.
Gonzalez-Fernandez, F. et al. Pericellular interphotoreceptor matrix dictates outer retina critical surface tension. Exp Eye Res 167, 163–173. https://doi.org/10.1016/j.exer.2017.10.014.
Gogola, A., Jan, N.-J., Lathrop, K.L., Sigal, I.A., 2018. Radial and circumferential collagen fibers are a feature of the peripapillary sclera of human, monkey, pig, cow, goat, and sheep. Invest. Opthalmol. Vis. Sci. 59, 4763–4774. https://doi.org/10.1167/iovs.18-25025.
Goldmann, W.H., 2018. Intermediate filaments and cellular mechanics. Cell. Biol. Int 42, 132–138. https://doi.org/10.1002/cbin.10879.
Gouyer, V., Gottrand, F., Desseyn, J.L., 2011. The extraordinarily complex but highly structured organization of intestinal mucus-gel unveiled in multicolor images. PLoS ONE 6:e18761.
Govindarajan, B., Gipson, I.K., 2010. Membrane-tethered mucins have multiple functions on the ocular surface. Exper. Eye Res 90, 655.
Guillaumet-Adkins, A., Heyn, H., 2017. Single-Cell genomics unravels brain cell-type complexity. Adv. Exp. Med. Biol 978, 393–407. https://doi.org/10.1007/978-3-319-53889-1_20.
Han, M.K.L., de Rooij, J., 2016. Converging and unique mechanisms of mechanotransduction at adhesion sites. Trends Cell. Biol. 26, 612–623. https://doi.org/10.1016/j.tcb.2016.03.005.
Hadj-Said, W., et al., 2017. Taurine promotes retinal ganglion cell survival through GABAB receptor activation. Adv. Exp. Med. Biol 975 (Pt 2), 687–701. https://doi.org/10.1007/978-94-024-1079-2_54.
Hafezi, F., Kanellopoulos, J., Wiltfang, R., Seiler, T., 2007. Corneal collagen crosslinking with riboflavin and ultraviolet A to treat induced keratectasia after laser in situ keratomileusis. J. Cat. Refract. Surg. 33, 2035.
Hardarson, S.H., 2013. Retinal oximetry. Acta Ophthalmol. 91, 1, Thesis 2.
Heo, G. Y. et al. Conversion of 7-ketocholesterol to oxysterol metabolites by recombinant CYP27A1 and retinal pigment epithelial cells. J Lipid Res 52, 1117–1127, https://doi:10.1194/jlr.M014217 (2011).

Houtkooper, R.H., Pirinen, E., Auwerx, J., 2012. Sirtuins as regulators of metabolism and healthspan. Nat. Rev. Mol. Cell Biol. 13, 225.
Huang, C.W., et al., 2016. Lys-315 at the interfaces of diagonal subunits of delta-crystallin plays a critical role in the reversibility of folding and subunit assembly. PloS. One 11:e0145957. https://doi.org/10.1371/journal.pone.0145957.
Inder, K.L., Davis, M., Hill, M.M., 2013. Ripples in the pond – using a systems approach to decipher the cellular functions of membrane microdomains. Mol. Biosystems 9, 330.
Invergo, B.M., Montanucci, L., Laayouni, H., Bertranpetit, J., 2013. A system-level, molecular evolutionary analysis of mammalian phototransduction. BMC Evol. Biol. 13, 52.
Iozzo, R.V., 1998. Matrix proteoglycans: from molecular design to cellular function. Annu. Rev. Biochem. 67, 609.
Ito, S., Pilat, A., Gerwat, W., Skumatz, C.M., Ito, M., Kiyono, A., et al., 2013. Photoaging of human retinal pigment epithelium is accompanied by oxidative modifications of its eumelanin. Pigment. Cell. Mel. Res. 26, 357.
Jaqaman, K., Grinstein, S., 2012. Regulation from within: the cytoskeleton in transmembrane signaling. Trends Cell. Biol. 22, 515.
Jan, N.J., et al., 2018. Crimp around the globe; patterns of collagen crimp across the corneoscleral shell. Exp. Eye Res. 172, 159–170. https://doi.org/10.1016/j.exer.2018.04.003.
Ji, F., Jung, J., Koharudin, L.M., Gronenborn, A.M., 2013. The human W42R gammaD-crystallin mutant structure provides a link between congenital and age-related cataracts. J. Biol. Chem. 288, 99.
Kadler, K.E., Hill, A., Canty-Laird, E.G., 2008. Collagen fibrillogenesis: fibronectin, integrins, and minor collagens as organizers and nucleators. Curr. Opin. Cell Biol. 20, 495.
Kannan, R., Sreekumar, P.G., Hinton, D.R., 2012. Novel roles for alpha-crystallins in retinal function and disease. Prog. Retin. Eye Res. 31, 576.
Kaushik, S., Cuervo, A.M., 2012. Chaperone-mediated autophagy: a unique way to enter the lysosome world. Trends Cell. Biol. 22, 407.
Keenan, T.D., Clark, S.J., Unwin, R.D., Ridge, L.A., Day, A.J., Bishop, P.N., 2012. Mapping the differential distribution of proteoglycan core proteins in the adult human retina, choroid, and sclera. Invest. Ophthalmol. Vis. Sci. 53, 7528.
Kevany, B.M., Palczewski, K., 2010. Phagocytosis of retinal rod and cone photoreceptors. Physiol. (Bethesda) 25, 8.
Kim, C., Ye, F., Ginsberg, M.H., 2011. Regulation of integrin activation. Annu. Rev. Cell Dev. Biol. 27, 321.
Kitagishi, Y., Matsuda, S., 2013. RUFY, Rab and Rap family proteins involved in a regulation of cell polarity and membrane trafficking. Int. J. Mol. Sci. 14, 6487.
Kiyota, N., et al., 2018. The impact of intraocular pressure elevation on optic nerve head and choridal blood flow. Invest. Ophthalmol. Vis. Sci 59, 3488–3496. https://doi.org/10.1167/.iovs.18-23872.
Klingberg, F., Hinz, B., White, E.S., 2013. The myofibroblast matrix: implications for tissue repair and fibrosis. J. Pathol. 229, 298.
Koch, K.W., Dell'orco, D., 2013. A calcium-relay mechanism in vertebrate phototransduction. ACS Chem. Neurosci. 4, 909–917.
Koob, S., Reichert, A.S., 2014. Novel intracellular functions of apolipoproteins: the ApoO protein family as constituents of the Mitofilin/MINOS complex determines cristae morphology in mitochondria. Biol. Chem. 39, 285–296.
Kumar, V., et al., 2017. Homeostatic nuclear RAGE-ATM interaction is essential for efficient DNA repair. Nucleic. Acids. Res 45, 10595–10613. https://doi.org/10.1093/nar/gkx705.
Lancaster, M.A., Knoblich, J.A., 2014. Organogenesis in a dish: modeling development and disease using organoid technologies. Science 345, 1247125. https://doi.org/10.1126/science.1247125.
Last, J.A., Liliensiek, S.J., Nealey, P.F., Murphy, C.J., 2009. Determining the mechanical properties of human corneal basement membranes with atomic force microscopy. J. Struct. Biol. 167, 19.
Lee, Y.S., Tresguerres, M., Hess, K., Marmorstein, L.Y., Levin, L.R., Buck, J., et al., 2011. Regulation of anterior chamber drainage by bicarbonate-sensitive soluble adenylyl cyclase in the ciliary body. J. Biol. Chem. 286, 41353.
Lee, J.G., Ko, M.K., Kay, E.P., 2012. Endothelial mesenchymal transformation mediated by IL-1beta-induced FGF-2 in corneal endothelial cells. Exp. Eye Res. 95, 35.
Leitinger, B., 2011. Transmembrane collagen receptors. Annu. Rev. Cell Dev. Biol. 27, 265.
Levin, M.H., Verkman, A.S., 2006. Aquaporins and CFTR in ocular epithelial fluid transport. J. Mem. Biol. 210, 105.
Liton, P.B., Gonzalez, P., Epstein, D.L., 2009. The role of proteolytic cellular systems in trabecular meshwork homeostasis. Exp. Eye Res. 88, 724.
Liu, W., Chun, E., Thompson, A.A., Chubukov, P., Xu, F., Katritch, V., et al., 2012. Structural basis for allosteric regulation of GPCRs by sodium ions. Science 337, 232.
Liu, Z., Kocaoglu, O.P., Miller, D.T., 2016. 3D Imaging of retinal pigment epithelial cells in the living human retina. Invest. Ophthalmol. Vis. Sci 57, OCT533–543. https://doi.org/10.1167/iovs.16-19106.
Logue, S.E., Cleary, P., Saveljeva, S., Samali, A., 2013. New directions in ER stress-induced cell death. Apoptosis 18, 537.
Lorenzi, M., 2007. The polyol pathway as a mechanism for diabetic retinopathy: attractive, elusive, and resilient. Exper. Diab. Res. 2007, 61038.
Luensmann, D., Jones, L., 2012. Protein deposition on contact lenses: the past, the present, and the future. Cont. Lens Ant. Eye 35, 53.
Mainali, L., O'Brien, W.J., Subczynski, W.K., 2018. Detection of cholesterol bilayer domains in intact biological membranes: Methodology development and its application to studies of eye lens fiber cell plasma membranes. Exp. Eye Res. 178, 72–81. https://doi.org/10.1016/j.exer.2018.09.020.
Malhotra, A., Minja, F.J., Crum, A., Burrowes, D., 2011. Ocular anatomy and cross-sectional imaging of the eye. Semin. Ultrasound CT MR 32, 2.
Mandal, A., Shahidullah, M., Delamere, N.A., 2018. TRPV1-dependent ERK1/2 activation in porcine lens epithelium. Exp. Eye Res 172, 128–136. https://doi.org/10.1016/j.exer.2018.04.006.
Mann, G.E., Yudilevich, D.L., Sobrevia, L., 2003. Regulation of amino acid and glucose transporters in endothelial and smooth muscle cells. Physiol. Rev. 83, 183.
Mao, Y., Schwarzbauer, J.E., 2005. Fibronectin fibrillogenesis, a cell-mediated matrix assembly process. Matrix Biol. 24, 389.
Massague, J., 2012. TGFbeta signalling in context. Nat. Rev. Mol. Cell Biol. 13, 616.
Mathias, R.T., Rae, J.L., Baldo, G.J., 1997. Physiological properties of the normal lens. Physiol. Rev. 77, 21.
Mathias, R.T., White, T.W., Gong, X., 2010. Lens gap junctions in growth, differentiation, and homeostasis. Physiol. Rev. 90, 179.

Maurice, D.M., 1957. The structure and transparency of the cornea. J. Physiol. 136, 263.
Maxfield, F.R., van Meer, G., 2010. Cholesterol, the central lipid of mammalian cells. Curr. Opin. Cell Biol. 22, 422.
McBrien, N.A., 2013. Regulation of scleral metabolism in myopia and the role of transforming growth factor-beta. Exper. Eye Res. 114, 128–140.
McCaig, C.D., Rajnicek, A.M., Song, B., Zhao, M., 2005. Controlling cell behavior electrically: current views and future potential. Physiol. Rev. 85, 943.
Meek, K.M., Boote, C., 2009. The use of X-ray scattering techniques to quantify the orientation and distribution of collagen in the corneal stroma. Prog. Retin. Eye Res. 28, 369.
Meek, K.M., Hayes, S., 2013. Corneal cross-linking – a review. Ophthalmic. Physiol. Opt. 33, 78.
Mercer, A.J., Thoreson, W.B., 2011. The dynamic architecture of photoreceptor ribbon synapses: cytoskeletal, extracellular matrix, and intramembrane proteins. Vis. Neurosci. 28, 453.
Meyer, P.A.R., 2018. Re-orchestration of blood flow by micro-circulations. Eye 32, 222–229. https://doi.org/10.1038/eye.2017.315.
Mimura, T., Yamagami, S., Amano, S., 2013. Corneal endothelial regeneration and tissue engineering. Prog. Retinal. Eye Res. 35, 1–17.
Miyagishima, K.J., et al., 2016. In pursuit of authenticity: induced pluripotent stem cell-derived retinal pigment epithelium for clinical applications. Stem Cells Transl. Med. 5, 1562–1574. https://doi.org/10.5966/sctm.2016-0037.
Moreau, K.L., King, J.A., 2012. Protein misfolding and aggregation in cataract disease and prospects for prevention. Trends Mol. Med. 18, 273.
Mosher, D.F., Adams, J.C., 2012. Adhesion-modulating/matricellular ECM protein families: a structural, functional and evolutionary appraisal. Matrix Biol. 31, 155.
Mostowy, S., Cossart, P., 2012. Septins: the fourth component of the cytoskeleton. Nat. Rev. Mol. Cell. Biol. 13, 183.
Mueckler, M., Thorens, B., 2013. The SLC2 (GLUT) family of membrane transporters. Mol. Aspects Med. 34, 121.
Muiznieks, L.D., Keeley, F.W., 2012. Molecular assembly and mechanical properties of the extracellular matrix: a fibrous protein perspective. Biochim. Biophy. Acta 1832, 866–875.
Muller-McNicoll, M., Neugebauer, K.M., 2013. How cells get the message: dynamic assembly and function of mRNA-protein complexes. Nat. Rev. Gen. 14, 275.
Nagaraj, R.H., Linetsky, M., Stitt, A.W., 2012. The pathogenic role of Maillard reaction in the aging eye. Amino Acids 42, 1205.
Naslavsky, N., Caplan, S., 2018. The enigmatic endosome–sorting the ins and outs of endocytic trafficking. J. Cell. Sci. 131. https://doi.org/10.1242/jcs.216499.
Neisch, A.L., Fehon, R.G., 2011. Ezrin, Radixin and Moesin: key regulators of membrane-cortex interactions and signaling. Curr. Opin. Cell Biol. 23, 377.
Neuhuber, W., Schrodl, F., 2011. Autonomic control of the eye and the iris. Autonom. Neurosci. 165, 67.
Nickla, D.L., Wallman, J., 2010. The multifunctional choroid. Prog. Retin. Eye Res. 29, 144.
Niehrs, C., 2012. The complex world of WNT receptor signalling. Nat. Rev. Mol. Cell Biol. 13, 767.
Nightingale, T.D., Cutler, D.F., Cramer, L.P., 2012. Actin coats and rings promote regulated exocytosis. Trends Cell Biol. 22, 329.
Nikitovic, D., et al., 2012. The biology of small leucine-rich proteoglycans in bone pathophysiology. J. Biol. Chem. 287, 33926.
Okada, Y., et al., 2010. Neurotrophic keratopathy; its pathophysiology and treatment. Histol. Histopathol. 25, 771.
Olins, D.E., Olins, A.L., 2018. Epichromatin and chromomeres: a 'fuzzy' perspective. Open. Biol 8. https://doi.org/10.1098/rsob.180058.
Orth, J.D., McNiven, M.A., 2003. Dynamin atFiguren the actin-membrane interface. Curr. Opin. Cell. Biol. 15 (1), 31–39. https://doi.org/10.1016/s0955-0674(02)00010-8 (2003).
Orban, T., Gupta, S., Palczewski, K., Chance, M.R., 2010. Visualizing water molecules in transmembrane proteins using radiolytic labeling methods. Biochemistry 49, 827.
Ostrin, L.A., 2019. Ocular and systemic melatonin and the influence of light exposure. Clin. Exp. Optom. 102, 99–108. https://doi.org/10.1111/cxo.12824.
Palczewski, K., 2012. Chemistry and biology of vision. J. Biol. Chem. 287, 1612.
Palfy, M., Remenyi, A., Korcsmaros, T., 2012. Endosomal crosstalk: meeting points for signaling pathways. Trends Cell. Biol. 22, 447.
Palkovits, S., et al., 2013. Measurement of retinal oxygen saturation in patients with chronic obstructive pulmonary disease. Invest. Ophthalmol. Vis. Sci. 54, 1008.
Palty, R., Hershfinkel, M., Sekler, I., 2012. Molecular identity and functional properties of the mitochondrial Na+/Ca2+ exchanger. J. Biol. Chem. 287, 31650.
Panaser, A., Tighe, B.J., 2012. Function of lipids – their fate in contact lens wear: an interpretive review. Cont. Lens Ant. Eye 35, 100.
Parker, R., Wang, J.S., Kefalov, V.J., Crouch, R.K., 2019. Interphotoreceptor retinoid-binding protein as the physiologically relevant carrier of 11-cis-retinol in the cone visual cycle. J. Neurosci. 31, 4714.
Peti, W., Nairn, A.C., Page, R., 2013. Structural basis for protein phosphatase 1 regulation and specificity. FEBS J. 280, 596.
Piccirelli, M., Bergamin, O., Landau, K., Boesiger, P., Luechinger, R., 2012. Vitreous deformation during eye movement. NMR Biomed. 25, 59.
Pickard, G.E., Sollars, P.J., 2012. Intrinsically photosensitive retinal ganglion cells. Rev. Physiol. Biochem. Pharmacol. 162, 59.
Prince, J., Chuck, R.S., 2012. Refractive surgery after Desçemet's stripping endothelial keratoplasty. Curr. Opin. Ophthalmol. 23, 242.
Quinlan, R.A., et al., 2013. Changes in the quaternary structure and function of MjHSP16.5 attributable to deletion of the IXI motif and introduction of the substitution, R107G, in the alpha-crystallin domain. Phil. Trans. R. Soc. Lond B. Biol. Sci. 368:0120327.
Rabbani, N., Thornalley, P.J., 2012. Glycation research in amino acids: a place to call home. Amino Acids. 42, 1087.
Rachel, R.A., Li, T., Swaroop, A., 2012. Photoreceptor sensory cilia and ciliopathies: focus on CEP290, RPGR and their interacting proteins. Cilia 1, 22.
Racioppi, L., Means, A.R., 2012. Calcium/calmodulin-dependent protein kinase kinase 2: roles in signaling and pathophysiology. J. Biol. Chem. 287, 31658.
Rask-Madsen, C., Kahn, C.R., 2012. Tissue-specific insulin signaling, metabolic syndrome, and cardiovascular disease. Arterioscl. Thromb. Vasc. Biol. 32, 2052.
Reinoso, R., et al., 2012. Topographical distribution and characterization of epithelial cells and intraepithelial lymphocytes in the human ocular mucosa. Mucosal Immunol. 5, 455.

Renard, H.F., et al., 2015. Endophilin-A2functions in membrane scission in clathrin-independent endocytosis. Nature 517, 493–496. https://doi.org/10.1038/nature14064.
Rhee, I., Davidson, D., Souza, C.M., Vacher, J., Veillette, A., 2013. Macrophage fusion is controlled by the cytoplasmic protein tyrosine phosphatase PTP-PEST/PTPN12. Mol. Cell. Biol. 33, 2458–2469.
Richdale, K., et al., 2013. Quantification of age-related and per diopter accommodative changes of the lens and ciliary muscle in the emmetropic human eye. Invest. Ophthalmol. Vis. Sci. 54, 1095.
Ritch, R., 2008. Exfoliation syndrome: beyond glaucoma. Arch. Ophthalmol. 126, 859.
Rizzuto, R., De Stefani, D., Raffaello, A., Mammucari, C., 2012. Mitochondria as sensors and regulators of calcium signalling. Nat. Rev. Mol. Cell Biol. 13, 566.
Rosenberg, L., Hellmann, W., Kleinschmidt, A.K., 1975. Electron microscopic studies of proteoglycan aggregates from bovine articular cartilage. J. Biol. Chem. 250, 1877–1883.
Rousselle, P., Beck, K., 2013. Laminin 332 processing impacts cellular behavior. Cell. Adh. Migr. 7, 122.
Rowan, S., Bejarano, E., Taylor, A., 2018. Mechanistic targeting of advanced glycation end-products in age-related diseases. Biochim. Biophys. Acta Mol. Basis. Dis. 1864, 3631–3643. https://doi.org/10.1016/j.bbadis.2018.08.036.
Saari, J.C., 2012. Vitamin A metabolism in rod and cone visual cycles. Annu. Rev. Nutr. 32, 125–145.
Sahin, E., DePinho, R.A., 2012. Axis of ageing: telomeres, p53 and mitochondria. Nat. Rev. Mol. Cell Biol. 13, 397.
Sanderson, J., et al., 2014. Purines in the eye: recent evidence for the physiological and pathological role of purines in the RPE, retinal neurons, astrocytes, Muller cells, lens, trabecular meshwork, cornea and lacrimal gland. Exp. Eye Res 127, 270–279. https://doi.org/10.1016/j.exer.2014.08.009.
Sasai, Y., 2013. Cytosystems dynamics in self-organization of tissue architecture. Nature 493, 318.
Schaffler, A., Buechler, C., 2012. CTRP family: linking immunity to metabolism. Trends Endocrinol. Metab. 23, 194.
Schlotzer-Schrehardt, U., Kruse, F.E., 2005. Identification and characterization of limbal stem cells. Exp. Eye Res. 81, 247.
Schmidt, T.M., Chen, S.K., Hattar, S., 2011. Intrinsically photosensitive retinal ganglion cells: many subtypes, diverse functions. Trends Neurosci. 34, 572.
Sebastian, C., Satterstrom, F.K., Haigis, M.C., Mostoslavsky, R., 2012. From sirtuin biology to human diseases: an update. J. Biol. Chem. 287, 42444.
Sharma, K.K., Santhoshkumar, P., 2009. Lens aging: effects of crystallins. Biochim. Biophys. Acta 1790, 1095.
Shin, K., Fogg, V.C., Margolis, B., 2006. Tight junctions and cell polarity. Annu. Rev. Cell Dev. Biol. 22, 207.
Shindou, H., et al., 2017. Docosahexaenoic acid preserves visual function by maintaining correct disc morphology in retinal photoreceptor cells. J. Biol. Chem. 292 (2017), 12054–12064. https://doi.org/10.1074/jbc.M117.790568.
Shoulders, M.D., Raines, R.T., 2009. Collagen structure and stability. Annu. Rev. Biochem. 78, 929.
Singh, P., Carraher, C., Schwarzbauer, J.E., 2010. Assembly of fibronectin extracellular matrix. Annu. Rev. Cell Dev. Biol. 26, 397.
Singh, S.R., et al., 2018. Wide-field choroidal vascularity in healthy eyes. Am. J. Ophthalmol. 193, 100–105. https://doi.org/10.1016/j.ajo.2018.06.016.
Sit, A.J., McLaren, J.W., 2011. Measurement of episcleral venous pressure. Exp. Eye Res. 93, 291.
Sneddon, M.W., Emonet, T., 2012. Modeling cellular signaling: taking space into the computation. Nat. Methods 9, 239–242. https://doi.org/10.1038/nmeth.1900 (2012).
Soboloff, J., Rothberg, B.S., Madesh, M., Gill, D.L., 2012. STIM proteins: dynamic calcium signal transducers. Nat. Rev. Mol. Cell. Biol. 13, 549.
Song, B., Zhao, M., Forrester, J., McCaig, C., 2004. Nerve regeneration and wound healing are stimulated and directed by an endogenous electrical field in vivo. J. Cell. Sci. 117, 4681.
Song, S., et al., 2009. Functions of the intermediate filament cytoskeleton in the eye lens. J. Clin. Invest. 119, 1837.
Song, M.S., Salmena, L., Pandolfi, P.P., 2012. The functions and regulation of the PTEN tumour suppressor. Nat. Rev. Mol. Cell Biol. 13, 283.
Sterling, P., 2013. Some principles of retinal design: the Proctor lecture. Invest. Ophthalmol. Vis. Sci. 54, 2267.
Strauss, O., 2005. The retinal pigment epithelium in visual function. Physiol. Rev. 85, 845.
Stringer, J.M., Barrand, S., Western, P., 2013. Fine-tuning evolution: germ-line epigenetics and inheritance. Reproduction 146, R37.
Su, X., Ohi, R., Pellman, D., 2012. Move in for the kill: motile microtubule regulators. Trends Cell. Biol. 22, 567.
Subczynski, W.K., Raguz, M., Widomska, J., Mainali, L., Konovalov, A., 2012. Functions of cholesterol and the cholesterol bilayer domain specific to the fiber-cell plasma membrane of the eye lens. J. Memb. Biol. 245, 51.
Summers, J.A., 2013. The choroid as a sclera growth regulator. Exp. Eye Res. 114, 120–127.
Suresh, S., 2007. Biomechanics and biophysics of cancer cells. Acta Biomat. 3, 413.
Swamynathan, S.K., 2013. Ocular surface development and gene expression. J. Ophthalmol. 103947.
Swaroop, A., Kim, D., Forrest, D., 2010. Transcriptional regulation of photoreceptor development and homeostasis in the mammalian retina. Nat. Rev. Neurosci. 11, 563.
Swarup, A., et al., 2019. Modulating GLUT1 expression in retinal pigment epithelium decreases glucose levels in the retina: impact on photoreceptors and Muller glial cells. Am. J. Physiol. Cell Physiol 316, C121–C133. https://doi.org/10.1152/ajpcell.00410.2018.
Tang, P.H., Kono, M., Koutalos, Y., Ablonczy, Z., Crouch, R.K., 2013. New insights into retinoid metabolism and cycling within the retina. Prog. Retin. Eye Res. 32, 48.
Teng, P.Y., Wanek, J., Blair, N.P., Shahidi, M., 2013. Inner retinal oxygen extraction fraction in rat. Invest. Ophthalmol. Vis. Sci. 54, 647.
Thevenin, A.F., et al., 2013. Proteins and mechanisms regulating gap-junction assembly, internalization, and degradation. Physiology (Bethesda) 28, 93.
Thoreson, W.B., Mercer, A.J., Cork, K.M., Szalewski, R.J., 2013. Lateral mobility of L-type calcium channels in synaptic terminals of retinal bipolar cells. Mol. Vis. 19, 16.
Toivola, D.M., Strnad, P., Habtezion, A., Omary, M.B., 2010. Intermediate filaments take the heat as stress proteins. Trends Cell. Biol. 20, 79.
Turei, D., Korcsmaros, T., Saez-Rodriguez, J., 2016. OmniPath: guidelines and gateway for literature-curated signaling pathway resources. Nat. Methods 13, 966–967. https://doi.org/10.1038/nmeth.4077.
Ujiie, H., Shibaki, A., Nishie, W., Shimizu, H., 2010. What's new in bullous pemphigoid. J. Dermatol. 37, 194.
Vallee, A., Lecarpentier, Y., Guillevin, R., Vallee, J.N., 2017. Aerobic glycolysis hypothesis through WNT/Beta-Catenin pathway in

exudative age-related macular degeneration. J. Mol. Neurosci 62, 368–379. https://doi.org/10.1007/s12031-017-0947-4.
van Reeuwijk, J., Arts, H.H., Roepman, R., 2011. Scrutinizing ciliopathies by unraveling ciliary interaction networks. Hum. Mol. Gen. 20, R149.
Varadaraj, K., Kumari, S.S., 2018. Molecular mechanism of Aquaporin 0-induced fiber cell to fiber cell adhesion in the eye lens. Biochem. Biophys. Res. Commun 506, 284–289. https://doi.org/10.1016/j.bbrc.2018.10.066.
Varma, S.D., Kovtun, S., Hegde, K.R., 2011. Role of ultraviolet irradiation and oxidative stress in cataract formation—medical prevention by nutritional antioxidants and metabolic agonists. Eye Cont. Lens 37, 233.
Vos, W.L., et al., 2010. Expression and structural characterization of peripherin/RDS, a membrane protein implicated in photoreceptor outer segment morphology. Eur. Biophys. J. 39, 679.
Wang, J., et al., 2011. Ultra-high resolution optical coherence tomography for imaging the anterior segment of the eye. Ophthal. Surg. Lasers Imag. 42, S15.
Wareham, L.K., Buys, E.S., Sappington, R.M., 2018. The nitric oxide-guanylate cyclase pathway and glaucoma. Nitric. Oxide 77, 75–87. https://doi.org/10.1016/j.niox.2018.04.010.
Wenzel, A., Grimm, C., Samardzija, M., Reme, C.E., 2005. Molecular mechanisms of light-induced photoreceptor apoptosis and neuroprotection for retinal degeneration. Prog. Retin. Eye Res. 24, 275.
Wright, A.F., Chakarova, C.K., Abd El-Aziz, M.M., Bhattacharya, S.S., 2010. Photoreceptor degeneration: genetic and mechanistic dissection of a complex trait. Nat. Rev. Genet. 11, 273.
Wu, S.Y., Zou, P., Mishra, S., McHaourab, H.S., 2018. Transgenic zebrafish models reveal distinct molecular mechanisms for cataract-linked alphaA-crystallin mutants. PloS. One 13:e0207540. https://doi.org/10.1371/journal.pone.0207540.
Yang, Y., et al., 2018. Protein SUMOylation modification and its associations with disease. Open Biology. 7(10). https://doi.org/10.1098/rsob.170167.
Yamada, K.M., Miyamoto, S., 1995. Integrin transmembrane signaling and cytoskeletal control. Curr. Opin. Cell. Biol. 7, 681.
Yamaguchi, M., Shiraishi, A., 2018. Relationship between eyelid pressure and ocular surface disorders in patients with healthy and dry eyes. Invest. Ophthalmol. Vis. Sci 59, DES56–DES63. https://doi.org/10.1167/iovs.17-23586.
Yamashiro, S., Gokhin, D.S., Kimura, S., Nowak, R.B., Fowler, V.M., 2012. Tropomodulins: pointed-end capping proteins that regulate actin filament architecture in diverse cell types. Cytoskeleton (Hoboken). 69, 337.
Young, T.L., Metlapally, R., Shay, A.E., 2007. Complex trait genetics of refractive error. Arch. Ophthalmol. 125, 38.
Yu, F.X., Guan, K.L., 2013. The Hippo pathway: regulators and regulations. Genes. Dev. 27, 355.
Yucel, Y.H., et al., 2009. Identification of lymphatics in the ciliary body of the human eye: a novel 'uveolymphatic' outflow pathway. Exp. Eye Res. 89, 810.
Zang, J., Neuhauss, S.C.F., 2018. The binding properties and physiological functions of recoverin. Front. Mol. Neurosci. 11, 473. https://doi.org/10.3389/fnmol.2018.00473.
Zheng, Z., et al., 2012. Sirtuin 1-mediated cellular metabolic memory of high glucose via the LKB1/AMPK/ROS pathway and therapeutic effects of metformin. Diabetes 61, 217.
Zhou, L., Beuerman, R.W., 2012. Tear analysis in ocular surface diseases. Prog. Retin. Eye Res. 31, 527–550.
Zhou, L., Zhao, S.Z., Koh, S.K., Chen, L., Vaz, C., Tanavde, V., et al., 2012. In-depth analysis of the human tear proteome. J. Proteomics 75, 3877–3885.

Chapter 5

Physiology of vision and the visual system

CONTENTS

Introduction
- Do I have good vision?
- Sensing colour
- Shape, form and depth perception (and more) help to 'shape' vision

Light detection and dark adaptation
- What are the limits of detectable light?
- Dark adaptation curve and retinal sensitivity
- Are two small stimuli equivalent to one large one (summation)?

Visual acuity and contrast sensitivity
- Visual acuity is not simply a function of cone activity
- Limits of and limitations on acuity
- Contrast sensitivity
- Best-corrected visual acuity: effects of external factors

Electrophysiology of the visual system
- The electrical response is initiated by phototransduction
- Electrophysiology of single retinal cells
- Retinal connections, circuitry and neurotransmitters
- Clinical visual electrophysiology
- Flicker

Colour vision
- Colorimetry and colour discrimination
- Different colours have different luminosity
- Colour detection requires more than one type of photoreceptor
- Convergence, Young–Helmholtz and Hering
- Colour blindness

Visual perception
- Monocular versus binocular vision
- Stereopsis and depth perception
- Retinal rivalry and ocular dominance
- Colour processing
- Shape detection
- Contribution of texture analysis and motion detection to depth perception

Division of labour in the visual system: parcellation and the human connectome
- Imaging studies
- The magnocellular and parvocellular pathways subserve different functions
- Parcelling out the processing in V1
- Is the visual cortex organized for hierarchical neural processing or for functional specialization?

Physiology of ocular movement
- Types of movement
- Control of eye movement
- The cerebellum

Conclusion

Introduction

Vision in all its aspects could arguably be described as the most important physiological function for survival. Vision encompasses detection of luminance (and, we now understand, irradiance), contrast sensitivity (visual acuity), discrimination of texture, colour, depth and motion disparities and integrates them into what we describe as perception. Most of these functions are located in the higher cortical centres where the retinal sensation (image) is converted into our personal view (percept) of the outside world. What is seen (conceived) might not be the same as what is perceived, or even detected, and the latter may be extensively edited through input from other non-visual centres, especially memory and previous visual experience.

Zeki's notion (1992) of the brain constructing an image of the world by segregating the component parts via cortical regions that are, for instance, directionally selective

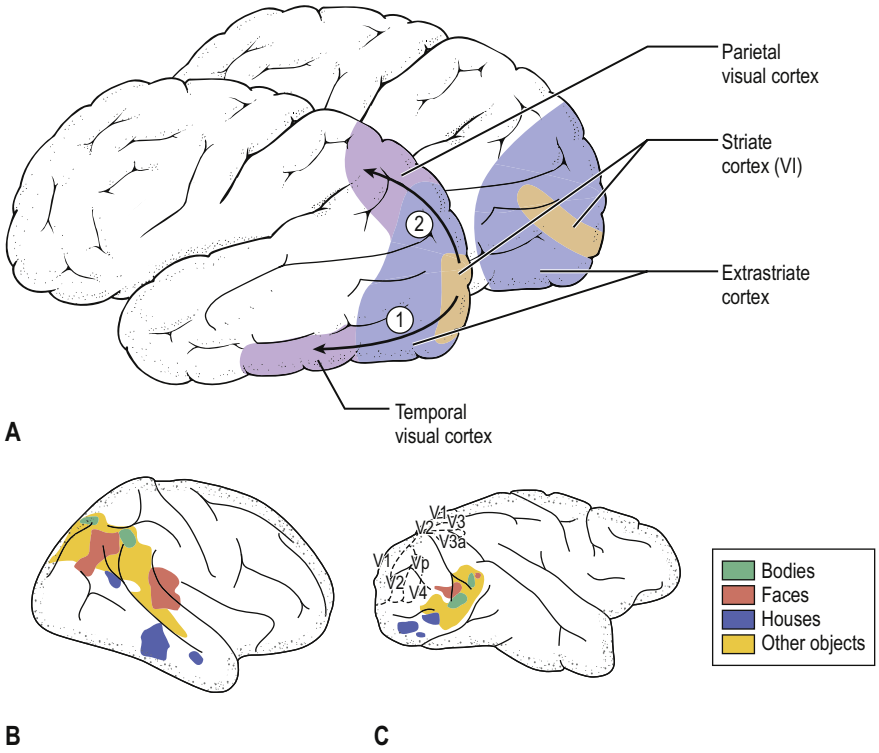


Fig. 5.1 Typical locations of category-selective regions in the human ventral visual cortex. (**A**) The location of visual regions in the human cortex, including the primary visual cortex (area V1 in the striate cortex) and the extrastriate cortex in the occipital lobe, and the traditional distinction into two visual cortical pathways that start in area V1 and extend into the temporal lobe (the ventral 'what' or 'object-vision' pathway (1)) or into the parietal lobe (the dorsal 'where' pathway (2)). (**B** and **C**) Ventral pathway regions in one individual that were activated significantly for selectivity of bodies, faces, houses or other objects. In addition, the *yellow* areas represent the regions that, in a group of people (n = 9), activated significantly in the contrast: intact objects > scrambled objects. All data were processed using SPM5 (Wellcome Department of Cognitive Neurology, London). (From Op de Beeck et al., 2008.)

(motion detectors), orientation selective, hue discriminators (colour) and that assess depth (stereopsis) has now been extended using functional magnetic resonance imaging (fMRI) to show how the segregated parts are integrated to provide a final but individual-specific image. Indeed, fMRI reveals which specific areas of the brain are activated when we visually register different objects (Fig. 5.1). At face value, it might be thought that spatial resolution would be most important to survival but in fact it has been shown that colour is what we see best.

There are further subtleties to the business of seeing. For instance, the manner in which we detect shape/form depends on more than activation of orientation-selective neurones. Remarkably, the recognition of faces and the recognition of the expression on a face are processed separately, as has been demonstrated in patients with damage to highly selective regions of the brain (Fig. 5.1). The recognition of texture is akin to a form of visual touch. This chapter provides a simplified and brief overview of aspects of the complex sensory and psychophysical responses to visual stimuli.

Do I have good vision?

A certain level of good vision is required for many daily activities, some of which may have a legal requirement, such as driving. However, good vision is a variable measure and depends on the set standard. Perhaps it is more valuable to have a concept of the limits of our visual capabilities. Vision can be considered in two ways: the optical requirements to achieve an image (i.e. refraction of light by the eye to focus the image on the retina, also known as physiological optics) and the neural processing of visual stimuli by the retina and the brain. The visual process is initiated by the detection of a light signal by photoreceptor cells in the outer retina. Photoreceptor cells convert light energy

to an electric stimulus, which is then transmitted to the bipolar cells and onwards to the ganglion cells in the retina (see Chapter 1, p. 45). The information is further transmitted in the axons of these cells (the optic nerves which, after 50% crossover in the optic chiasma, become the optic tracts) to the visual thalamic organ, the lateral geniculate nucleus (LGN). Synaptic contact with neurones in the LGN that project to the cerebral cortex permits onward transmission of the signals via optic radiation to the visual or striate cortex (V1), where they interact with many other neuronal connections from visual cortical cells in the prestriate cortex (V3–V5) and where parcelling out and processing of the signals takes place to build up the final perceived visual image. Input is also received by the visual cortex from many other areas, particularly those controlling general motor function and eye movement, cerebellar and spatial sense, memory and many other functions located in prefrontal cortex. This produces some of what is known as 'top down' modulation of visual responses whereby signals received and interpreted in the visual cortex can be influenced by input to the final image (the perceived image) from other areas, such as visual area 4 (V4), which combines elements of object recognition with visual attention.

An example of 'top down' integration of visual attention with visual cortical input from focus of attention in visual field is available online at: https://expertconsult.inkling.com

The knowledge that vision and visual acuity are modulated by top down and other influences has led to novel approaches to visual restoration in some blinding disorders (Sabel, 2018).

In biophysical terms a photoreceptor is capable of detecting a single photon of light (see Chapter 4, p. 251), but in practice what are the limits of detection of a visual stimulus? This depends on the nature of the stimulus and the nature of the ambient conditions in which it is presented. Sensing light is a function of all regions of the retina but the foveal region is specialized for high spatial resolution (visual acuity) and colour detection, served by the small 'midget', slow-transmitting ganglion cells (the parvocellular or P system). In contrast, luminance and motion detection are served by the large, fast-transmitting ganglion cells (the magnocellular or M system) that dominate the remaining retina and thus incorporate the entire visual field. According to Barlow's single neurone doctrine, it should take only one neurone to detect a visual stimulus (Barlow, 1972); however, the question is whether the signal received from the one rod in a thousand stimulated in the dark is sufficient to activate this neurone. Psychophysical studies have shown that both the luminance and colour thresholds for vision are different by orders of magnitude for P neurones between monkeys and humans, suggesting that more than one neurone is involved in detecting a light stimulus. In fact, continuous pooling of information occurs both in the excitatory and inhibitory neuronal activity that is present at all times and that, after a visual stimulus, the changes in the response rate of many neurones are 'sampled' by the brain until they reach a certain threshold level, at which point they register and the stimulus is 'recognized' (Hurlbert and Derrington, 1993). This perhaps explains how we can sometimes look at an object and yet not 'see' it; furthermore, these psychophysical considerations are highly relevant to methods for testing vision, for instance with regard to setting luminance thresholds for studies of visual fields using small transient targets.

Whether or not we have good vision at any moment in time depends on our level of awareness, consciousness and attention to visual stimuli, which have many properties such as depth, shape, form, colour texture and more besides, each with its own rate of detection/discrimination.

Flicker can be used to determine limits of vision

Detection of a stationary target or spot depends on the size and brightness of the spot relative to the background. The limits of detectability of the target are therefore determined by the spatial resolution and the anatomical relationships between stimulated receptors (see later). Spatial resolution is highest at the fovea and declines sharply towards the peripheral retina; this is clearly demonstrated by the detection threshold at different eccentricities in the visual field.

The threshold for spatial resolution is, however, considerably higher than that for detecting light; this latter parameter can be measured by flicker detection, which is the ability to detect two stimuli separated in time. This function is normally subserved by rod photoreceptors, while spatial resolution is subserved by cones, with some input from rods.

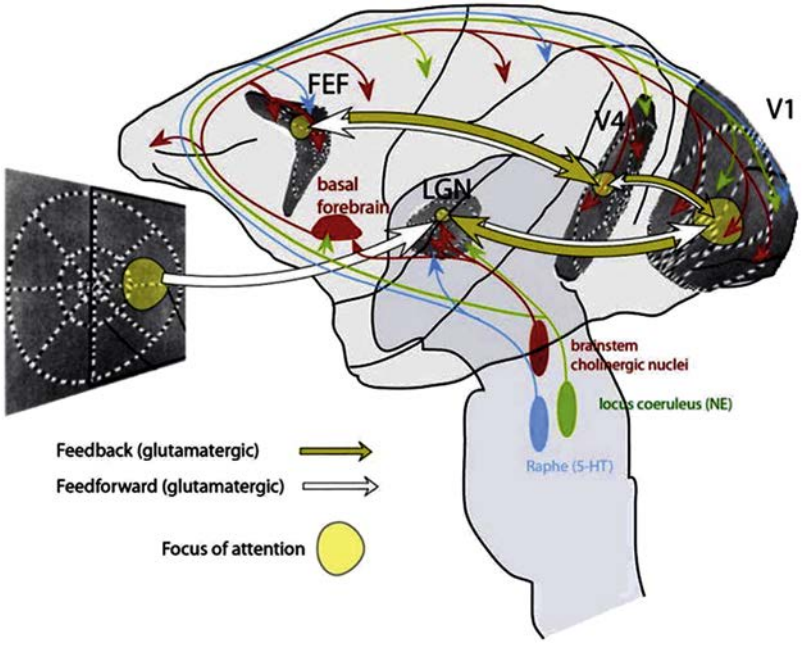
The critical flicker fusion (CFF) frequency test (see later) may be a useful predictor of cataract surgery outcome in cases of co-morbidity of lens opacity and macular disease because the CFF (see later) is relatively unaffected by image degradation resulting from cataracts but would be affected by foveal disease. In addition, as a neurophysiological test, it has been used in the early detection of hepatic encephalopathy.

Motion detection is also a feature of rod vision

It is clear, therefore, that the ability to detect a standard small bright spot in specified regions of the visual field is in fact a much more complex task than would at first appear. Not only does it depend on the absolute brightness of the stimulus but also on the background on which the stimulus is presented and thus on contrast. It also depends on whether the target is moving or stationary and, if stationary, for how long the target is presented. Its detection depends

TOP DOWN INPUT FROM FRONTAL EYE FIELDS TO VISUAL CORTEX (V1)

The brain is constantly shifting its position to meet the requirements of changes in behaviour. It continuously adapts its processing machinery to behavioural demands. Information is therefore transformed, modulated and rechannelled through different neural cortical wiring circuits, which, it has been revealed in series of novel experiments in various animal models, can assess small changes in activity occurring at low frequency. Harris and Thiele suggest that processes involved in selective attention are similar to those involved in ground state changes; these are summarized in eFig. 5.1 (Harris and Thiele, 2011). They include increased activity of cortical neuromodulatory afferents (red arrows: cholinergic; blue arrows: serotonergic; and green arrows: noradrenergic), which causes a general desynchronization and reduction in spontaneous fluctuation but might lack the spatial selectivity to desynchronize the patch of cortex representing the attended stimulus. Focused glutamatergic inputs arising from feedback connections could provide this specificity (yellow arrows), causing enhanced desynchronization and sensory responses in the regions of cortex activated by the object of attention. The yellow circle in the visual display (visual field) indicates the focus of attention, which affects processing in thalamic and cortical areas at specific locations (indicated by the yellow patches). The distorted replication of the visual world in the different areas illustrates (approximately) the known retinotopic organization of these different areas.



eFig. 5.1 Processes involved in attention are thought to be similar to changes occurring in state changes in cortical activity for specific functions. Neuromodulatory effects are identified by the *red, blue and green arrows* and are explained in the text. (From Harris and Thiele, 2011.)

on the density of photoreceptors and thus the region of retina stimulated. If it is a moving target, it will stimulate different cortical neurones depending on which direction it is moving. This functional segregation of visual input is retained at several levels within the cortex before construction of the final visual image.

Sensing colour

Cone photoreceptors are built to sense colour, which they do through cone opsin proteins. There are three types in humans: long (L, red), medium (M, green) and short (S, blue) wavelength cones, each with its own specific opsin (see p. 311). Early colour matching experiments in which the colour of a test stimulus is matched by adding together stimuli composed of the three primary colours verified that three colours were all that was required to detect the full spectrum of white light (see Colorimetry, p. 279). This is known as the trichromatic theory of colour vision. Each colour has properties such as hue and chromaticity.

There are many hues but only three primary colours

Hue is an idealized term for the colour produced by light of a single wavelength. In spite of having only three cone photoreceptors, we are able to distinguish many hues of colour, e.g. lilac and violet. It is therefore clear that any single colour is recognized by a mixture of the three primaries and that there must be overlap in the spectral sensitivity for each primary colour (see later). Theoretically, it should be possible to produce light of a single wavelength using a narrow slit on a device such as a monochromator, but photoreceptor sensitivity is also subject to the intensity of the light, and narrow wavebands of this degree of selectivity are not sufficiently intense to produce a stimulus.

The hue-discrimination curve (Fig. 5.2) describes the physiological limits at which a shift in wavelength can be discriminated as a change in colour. It was derived empirically by using fixed amplitude selective wavelength stimuli (1.5 cycles/300 nm), which found that there were two peaks of discrimination, one in the yellow/orange and a second in the blue/violet range. Monochromatic light therefore is not a practical reality; most colours are in fact tints, i.e. they are unsaturated hues, the degree of unsaturation being determined by the amount of additional white light they require to match them to a hue.

Chromaticity is semi-quantified 'colouredness'

Chromaticity refers to 'colouredness' and depends on hue, saturation and intensity of light (luminosity). Indeed, hue itself is not independent of the luminosity of the stimulus and chromatic shifts occur as the intensity increases until all hues appear yellow–white (the Bezold–Brucke phenomenon) or as the intensity decreases, when all hues appear achromatic (the Purkinje shift; see later). Any colour can thus be matched by a mixture of the three primary colours plus or minus a proportion of white light to account for unsaturation; these are formally described in the chromaticity chart and can be determined at different levels of lower or higher colour metrics (discrimination) (Fig. 5.3). The International Commission on Illumination (CIE) has developed standard colour 'observers' (colorimeters), which have proved valuable in defining true colours but do not fully take into account the minimally perceptible differences in colour discrimination (just noticeable differences, JND) based on aspects of photoreception affected by background noise.

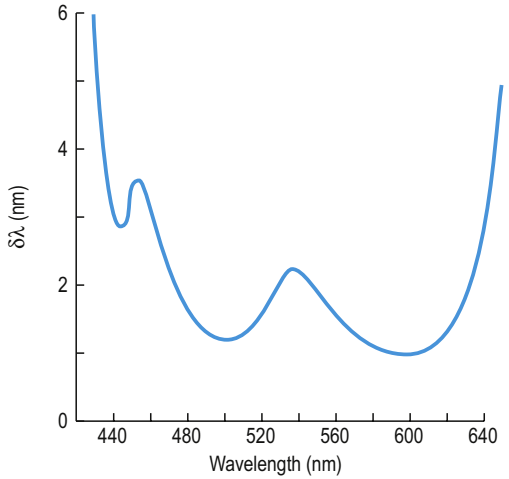


Fig. 5.2 Hue-discrimination curve comparing wavelength discrimination (*y*-axis) with changing wavelength (*x*-axis). Discrimination of hues varies for any given wavelength, being best at 455 and 535 nm.

However, the chromaticity chart has true practical value, for instance when applied to colour-mixing techniques used routinely in computer programs to produce different colours digitally for image creation and other purposes, and have been developed into a computerized colour vision test. Since the colour of any object can be categorized by the chromaticity chart based on the dominant wavelength and the colour saturation, detection of colour by 'machine vision' (robotics) via computer algorithms based on colour discrimination thresholds is now possible (Palchikova, 2018).

Clinically, colour vision can be tested using hue-discrimination techniques (e.g. the Farnsworth–Munsell 100 hue test) and normal values vary with age (affected by JND effects), peak ability occurring around the age of 19 years. Some effect by rods on cone vision has also been shown by rod function studies (background noise) (Buena-Atienza et al., 2018).

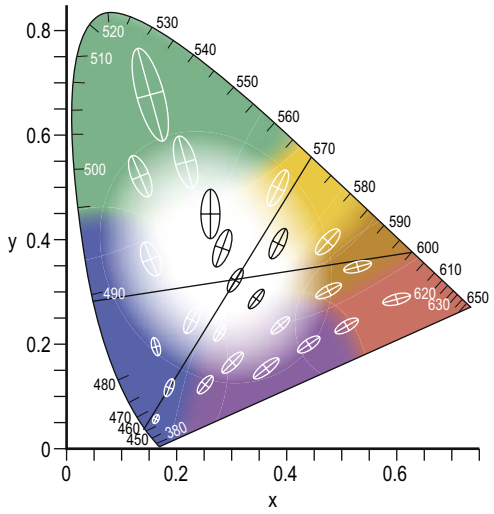


Fig. 5.3 Lower versus higher colour metrics. The chromaticity diagram is long established in the field of colour discrimination. It represents the laws of colour mixing in terms of (*x*, *y*). This is the domain of lower colour metrics. The straight lines through $x = 0.305$, $y = 0.323$ indicate that this white colour can be obtained, for instance, by mixing, in suitable quantities, light of wavelengths 570 and 465 nm or of wavelengths 600 and 489 nm. The ellipses, drawn here at 10 times their true size, are contours of just noticeably different colours from their central colour. The description of the differences in size, shape and orientation of these 'just noticeable differences' ellipses is the domain of higher colour metrics. (From Vos, 2006.)

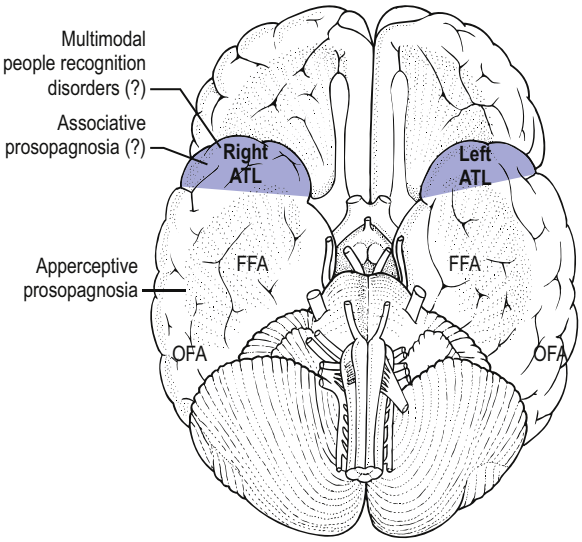


Fig. 5.4 Processing information for features (of the face) involves a specific neural cortical network, the occipital face area *(OFA)*, the face fusiform area *(FFA)* and the anterior temporal lobe *(ATL)*, which have a right hemisphere dominance. This region underlies neurological perceptual defects such as prosopagnosia (lesion in the posterior region) and person recognition disorders (lesions in the anterior temporal lobe). (From Gainotti, 2013.)

Shape, form and depth perception (and more) help to 'shape' vision

The discrimination of shape and form is highly developed in primate vision and the cortical localizations that define these functions are now well established. The brain recognizes and can categorize objects according to shape irrespective of the angle or distance from which they are viewed, the ambient lighting conditions, or other factors, indicating a complex and sophisticated level of neural processing.

Shape processing is achieved by specialized orientation-sensitive cells in the visual cortex, but the extra dimension of form recognition, as in recognition of facial features, requires additional processing. Studies of patients with specific visual defects such as prosopagnosia (inability to recognize familiar faces), however, are powerful indicators of the localization of visual functional sites. Even this apparently specific defect can be subdivided into a perceptual form and an associative form, the latter arising when the patient can perceive the image but cannot draw on visual memory sufficient to 'remember' the face. These separate functions have been ascribed to different regions of the brain, although associative face recognition may also involve other senses such as voice recognition (Fig. 5.4).

Much is also known about depth perception. For instance, it has been shown that there are specific cortical cells responsive only to disparate but simultaneous orientations of an object, presented to non-corresponding regions of the retina. In addition, clues on the nature of 3D structures can be obtained from motion detection (structure-from-motion). However, although no specific 'depth appreciation' cortical region has been identified, the lateral occipital cortex is a favoured area showing much activity. The appreciation of depth is more than simply stereopsis and is built up from many other cues (see later). An understanding of how these modify perception can only come after a description and appreciation of the different types of visual stimuli that separately induce discrete responses in the brain.

Light detection and dark adaptation

What are the limits of detectable light?

As in all biological systems, there is no precise answer to the above question. Light energy comes in quanta (small packets) and it has been estimated that between 50 and 150 quanta of light are required to strike the cornea for a discrete signal to be detected. Of these quanta, only about 10% actually reach the photoreceptors. The detection of

this stimulus is not simply a function of photoreceptor stimulation but is subject to 'dark light' (effectively background noise in rhodopsin photoisomerization) and is also dependent on higher neural function, and the concept of a visual threshold is more or less a statistical function dependent on how large the stimulus has to be to reach a level of recognition. This is well recognized by anyone who performs a visual field test using an automated visual field analyser.

Thresholds and the frequency of seeing

A distinction must be made between the theoretical estimate of the number of quanta required to produce an electric stimulus in a patch-clamped photoreceptor cell and the psychophysical conversion of the light stimulus to a perceived sensation. The latter depends on a defined measure, termed the 'frequency of seeing', which is the number of times a repetitively presented minimal stimulus is detected, and is a probability function that varies between and within individual observers.

The former is theoretically a single photon of light. However, there is considerable 'noise' in the system owing, for instance, to random opening and closing of ion channels as a result of thermal isomerization of rhodopsin, or to scatter from background and/or stray light energy from the stimulus itself. These effects can account for up to 1000 quanta/degree, which is well above the absolute threshold for light stimulation. It is thought that some of this is 'smoothed' by coupling between photoreceptors. More recently, it has been shown that the time taken to process the stimulus (temporal resolution) is apparently much finer (more precise) than the duration of the light stimulus itself while 'thermal noise' from rhodopsin activation is less than previously imagined suggesting that the limits on detection of a photon are less affected by noise and more determined the terminal resolution of the response after all (Field et al., 2019).

What is the minimal stimulus for vision?

Even when theoretical biophysical considerations such as signal-to-noise ratio are taken into account, this deceptively simple question depends on many factors such as background illumination, spatial frequency, summation, wavelength, dark adaptation and optical qualities of the image-gathering system. The specific conditions have to be stated, therefore, before this quantity can be expressed.

In addition, consideration of whether a single rod can detect a single photon of light *in vivo* has to take into account the different routes that a rod can take to stimulate a ganglion cell and convert this into a behavioural response (see pp. 297–299). Furthermore, the minimal stimulus size for a cone is also an important measure to define with physiological and clinical relevance.

Box 5.1 **Dark adaptation**

The normal dark adaptation curve *(a)* varies if the conditions are varied: with a very small central white target, rods fail to become stimulated at all and the curve flattens out *(b)*. If the cones are first light adapted by weakly stimulating them to maximum sensitivity or by adapting subjects to red light before placing them in the dark, the cone component can be 'lost' *(c)*; subjects without cone vision also have no cone component (rod monochromats).

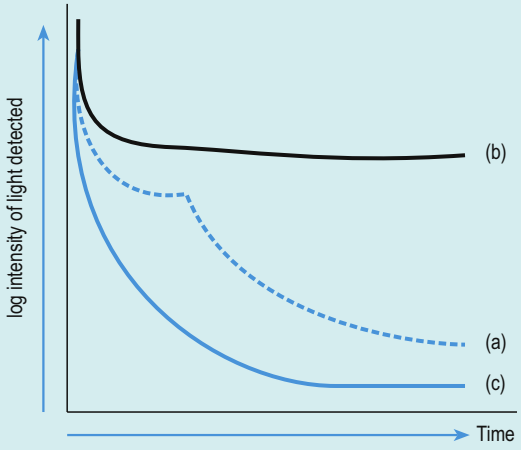


Figure outlining dark/light adaptation responses: *(a)* mixed rod and cone response of physiological dark adaptation, *(b)* pure cone response, *(c)* pure rod response.

Figure courtesy H. Dawson.

Dark adaptation curve and retinal sensitivity

The minimum visual stimulus varies depending on ambient light conditions, i.e. whether the stimulus is viewed in the dark or under normal/bright light conditions. In the dark, the eye becomes progressively more sensitive to light stimulation until the light threshold reaches a minimum after about 30 minutes. This is demonstrated in the dark adaptation curve (Box 5.1), which has two components: an early one resulting from increases in the cone sensitivity and a second produced by increases in rod sensitivity. There is also a light adaptation curve for cones in which the sensitivity to light varies as the luminosity increases or decreases within a wide range of high ambient illumination (see later). Thus, the shape of the curve can be varied by altering the conditions.

Light and dark adaptations are the psychophysical correlates of visual pigment bleaching and regeneration and can be measured by reflection densitometry. This technique is based on the assumption that light reflected from the unbleached retina will contain lower amounts of 500 nm (peak sensitivity for rods) light than that reflected from the bleached retina, since there will be considerable absorption of 500 nm light by the dark-adapted retina. Slow reflectance changes (6–30 minutes) can also be detected and appear to arise in cones. Reflection densitometry studies permit an evaluation of the photosensitivity of the retina, i.e. the rate at which bleaching takes place for a given intensity of illumination. It has been estimated that the normal retina absorbs 50% of the quanta of light striking the retina but, as discussed previously, this is not necessarily associated with a perceived visual stimulus because the absorption by a single rod of a photon of light can have at least three outcomes.

Regeneration of rhodopsin after dark adaptation is slow, taking 30 minutes for completion, with a half-time in humans of 5 minutes. This varies significantly between species. Clearly, the sensitivity of the retina in any individual will depend on the total amount of rhodopsin, and this relationship has been delineated in the Dowling–Rushton equation:

$$\log(z) / A = aB$$

where A is the threshold in complete dark adaptation, B is the fraction of bleached rhodopsin and a is a constant of proportionality. This sort of mathematical relationship has been used to estimate the rhodopsin content of the retinas of patients with certain forms of retinal disease, such as Oguchi's disease, fundus albipunctatus and especially vitamin A deficiency. However, it is important to realize that receptor sensitivity and rhodopsin content are not equivalent and that sensitivity to light is markedly reduced after partial bleaching, long before there is a reduction in rhodopsin content. This is clear in the isolated retina where photosensitivity is permanently reduced even after full recovery in the dark. These changes reflect the level of rhodopsin intermediates (metarhodopsin I and II) in the retina, which remain after bleaching (see Fig. 4.84, p. 259).

These effects are important in the determination of photosensitivity of discrete regions of the retina where it has been shown that reduced sensitivity can be detected in regions of the retina not exposed to point sources of light. Although this has been attributed to light scatter, there are probably other mechanisms operative here, particularly related to convergence of neural input (see later).

What does adaptation mean at a molecular level? There is considerable evidence to show that dark adaptation and regeneration of rhodopsin are dependent on the local concentration of 11-*cis* retinal, and the limiting factor for recovery after a large bleach is the rate at which 11-*cis* retinal is delivered to opsin in the bleached photoreceptors (Fig. 5.5)

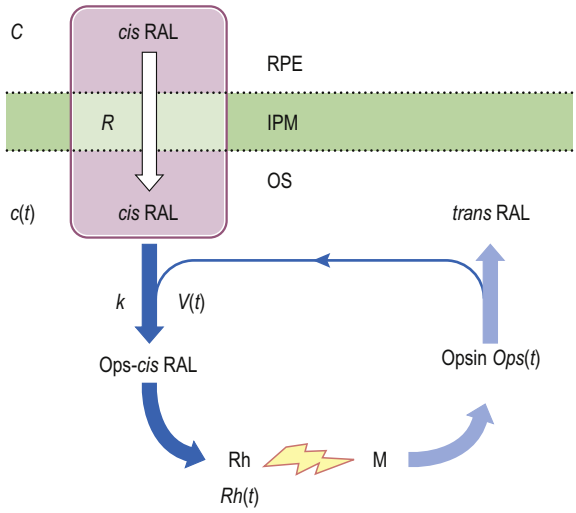


Fig. 5.5 Schematic of the MLP rate-limited model. Removal of photoproduct and regeneration of visual pigment are rate-limited by the delivery of 11-*cis* retinal (*cis* RAL) from the retinal pigmented epithelium *(RPE)* to opsin in the outer segment *(OS)*. *IPM*, Interphotoreceptor matrix; *M*, metarhodopsin; *Rh*, rhodopsin. (From Lamb and Pugh, 2004, with permission from Elsevier.)

despite some more recent evidence that some of the retinoid conversion steps also occur in the Müller cells (see Chapter 4, p. 244, Fig. 4.69). Thus, because a healthy retinal pigment epithelium is central to this process, age-related decline in dark adaptation can be explained on this basis. It is also dependent on termination of the photon-induced signal brought about by efficient phosphorylation of the enzyme rhodopsin (Chapter 4, p. 266) via rhodopsin kinase, with the subsequent docking of arrestin to the complex, so that free opsin can be made available to bind more 11-*cis* retinal and respond to a new photon. Absence of rhodopsin kinase (also known as G protein-coupled receptor kinase 1, Grk1) or of arrestin underlies the pathology of stationary night blindness (Oguchi's disease) while absence of the equivalent cone opsin kinase (Grk7) causes enhanced S cone syndrome.

Light and dark adaptation are also under some degree of regulation at the photoreceptor ribbon synapse via the SNARE proteins, complexins ¾, which regulate synaptic vesicle release (fewer released in the light than in the dark) (Babai et al., 2016).

Psychophysical evaluation of rhodopsin bleaching has been experimentally tested in humans by comparing a range of dark adaptation curves to different background light levels with the amplitude of the *a* wave of the electroretinogram (ERG), also known as the rod current (see p. 294), since both are desensitized to varying degrees by the amount of ambient light. It appears that rhodopsin regeneration and a-wave recovery rates match well.

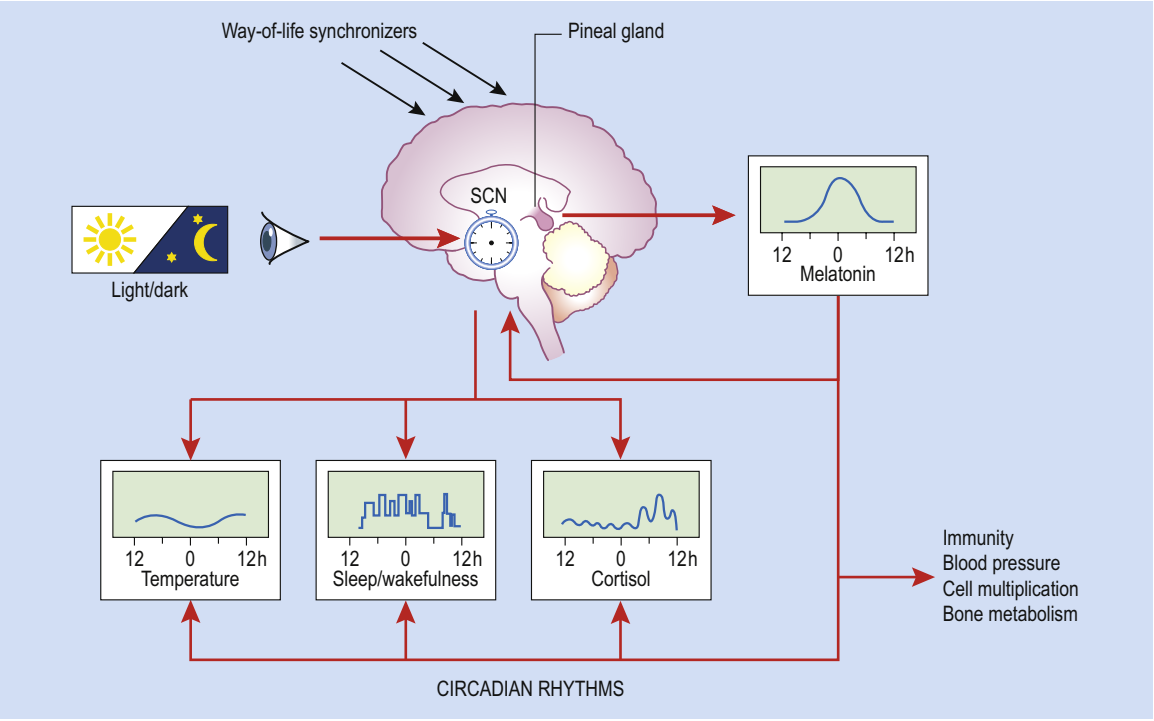


Fig. 5.6 Melatonin acts as an endogenous synchronizer. (From Claustrat et al., 2005, with permission from Elsevier.)

Cones also regulate their sensitivity in photopic conditions (see above under reflectance densitometry), but it is much more difficult to saturate this response, i.e. cones still adapt at high intensities of steady illumination and the recovery time is very short (100 milliseconds compared with 20–30 minutes for rods). This is probably furnished by Müller cell-derived 11-*cis* retinal, also under the control of retinal pigmented epithelial (RPE) 65 in the Müller cell (see Chapter 4, p. 258, Fig. 4.82).

In summary, adaptation is exactly what it means: the retina rapidly adapts to changes in background illumination such that it can respond to increasingly strong or weak stimuli. However, the dynamic range of responses (normally a range from zero to a few hundred impulses per second) over which it functions at any specific level of illumination remains the same and the intensity of the response when it makes one is also the same. Put simply, the retina adapts rapidly to new lighting conditions when there is plenty of light but slowly when light levels are low.

Melatonin and circadian rhythms

The circadian clock is a process whereby genes regulating various functions such as the sleep–wake cycle, body temperature, immune cell function and behaviour are expressed in a rhythmical manner. At least 11 core clock genes have been discovered, including a set of period genes (*PER* 1,2,3) and clock genes (*CLOCK*). CLOCK and BMal-1 form a complex in the cytoplasm which translocates to the nucleus, activates PER and CRY genes and acts as transcriptional activators or repressors. PER:CRY genes have multiple downstream effects on other important regulatory transcription factors such as POPα/β/γ important in immune cell function and genes involved in the synthesis of melatonin. The light/dark (sleep/wake) cycle, generated by pacemaker cells in the suprachiasmatic nuclei, drives the production of the pineal gland secretory product, melatonin as well as melatonin produced at other sites, including the retina and bone marrow. Melatonin is synthesized in a circadian cycle from tryptophan via serotonin in two steps involving the enzymes serotonin-*N*-acetyltransferase (NAT) and hydroxyindole-*O*-methyltransferase (HIOMT) (see Chapter 4, p. 254) and provides information to the organism which allows the daily organization of various physiological functions. Because melatonin secretion can adapt to night length, it can promote a seasonal (photoperiod) as well as a diurnal rhythmicity (Fig. 5.6). Apart from its obvious physiological functions, such as sleep–wake patterns, blood pressure levels, temperature and colonic activity, melatonin influences diurnal variations in innate immune defence functions such as antioxidation, glucose

regulation, blood coagulation enzyme systems and several ocular functions such as control of aqueous secretion.

Melatonin is a methoxyindole, synthesized and secreted principally by the pineal gland at night under normal environmental conditions and binds to two receptors (M1 and 2). The endogenous rhythm of secretion is generated by the suprachiasmatic nuclei and entrained to the light/dark cycle (Fig. 5.6). Light is able either to suppress or to synchronize melatonin production according to the light schedule. The nyctohemeral rhythm of this hormone can be determined by repeated measurement of plasma or saliva melatonin or urine sulphatoxymelatonin, the main hepatic metabolite.

The primary physiological function of melatonin is to convey information concerning the daily cycle of light and darkness to body physiology. As indicated above, this information is used for the organization of physiological functions in response to changes in the photoperiod (seasons). There is still, however, only limited evidence for seasonal rhythmicity of physiological functions in humans related to possible alteration of the melatonin message in temperate areas under field conditions, although there is a reported link between seasonal affective disorder, clinical depression and its control with novel antidepressant drugs based on melatonin receptor stimulation, such as agomelatine. Major clinical depressive illness has also been linked to markedly reduced function of the core clock genes in the brain in a recently reported post-mortem microarray analysis. The centre for regulation of the photoperiod has been suggested to lie in the pars tuberalis of pituitary under control of thyrotrophin (Korf, 2018; Wood and Loudon, 2018).

The onset of daily melatonin secretion is a very robust biochemical signal of night, regulating circadian rhythms. Melatonin stabilizes and strengthens the coupling of circadian rhythms, especially of core temperature and sleep–wake rhythms. However, regulation of melatonin secretion is complex, following central and autonomic pathways, and is readily disturbed. The resulting alteration may increase predisposition to disease or add to the severity of symptoms of existing diseases, modifying the course and outcome of the disorder.

Melatonin is also produced by photoreceptors where it can act on melatonin receptors (MRs) in an autocrine manner, as well as on MR+ ganglion cells and other retinal neural cells in a paracrine manner. Thus, it regulates the activity of photoreceptors, acting on horizontal cells stimulated by cones to reduce their responsiveness, but heightens ON-bipolar cells and ganglion cells in some species. In this way, melatonin is thought to fine-tune visual function, especially in cone cells under varying ambient light conditions.

Further information of fine tuning of photoreceptor function by local production of melatoninis available online at: https://expertconsult .inkling.com

Box 5.2 **Light energy**

Light energy is measured subjectively by its brightness and luminance can be measured in:

- trolands
- candelas
- luxes.

Specific measures of brightness are as follows:

(a) Intensity of illumination of a surface (L) = intensity of the light source/square of the distance between the source and the surface.

$$L = I / r$$

(b) Unit of L = foot-candela or metre-candela
1 lux = 1 metre-candela
1 phot = 1 cm-candela
1 lambert = 1 candela at 1 cm distance for a perfectly diffusing light source on a surface at 1 cm
1 troland = a unit of retinal illumination that results when a surface luminance of 1 candela/m^2 is viewed through a pupil area of 1 mm^2
1 lumen = one unit of flux C, the spherical illumination from a point source of light of intensity of 1 metre-candela or 1 foot-candela.

Interestingly, MRs are present on many other ocular tissue cells such as ciliary epithelium, RPE cells, lens cells, corneal endothelium and keratocytes, and stromal cells in the sclera and choroid. MRs are also widely distributed throughout the body.

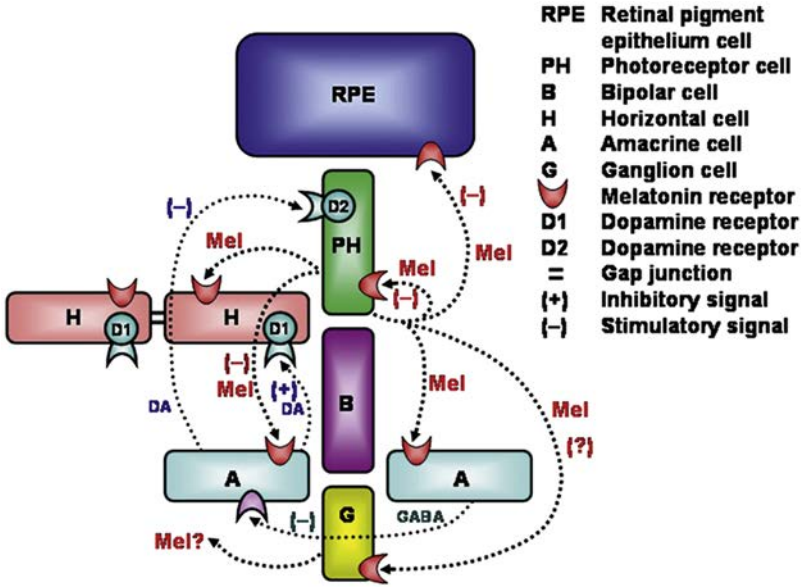
Are two small stimuli equivalent to one large one (summation)?

The threshold for light detection can be measured arbitrarily by setting certain conditions of stimulus size, brightness, pupil size and level of background illumination and recording how often a subject detects the stimulus. An empirically set level of 'hits' or positive detection responses (e.g. 55%) can then be set and expressed in trolands (Box 5.2). Experimentally it has been estimated that at the limit of light detection in the fully dark-adapted eye, the retina is illuminated to a level of 4.4×10^{-5} trolands, which is equivalent to the stimulation of only 1/5000 rods per second. However, if the light is concentrated on one area, it will more readily elicit a response and it therefore becomes less practical to think of light energy in terms of area of retinal illumination; instead the minimum flux in light energy required to induce a detectable response is commonly accepted as the threshold and is around 120 quanta per second or, if the stimulus is instantaneous, between 5 and 15 quanta of light.

From this it is clear that stimulation of a single rod is insufficient to produce a visual sensation (even though an electrical response may occur in terms of a change in

LOCAL CONTROL OF PHOTORECEPTOR BY MELATONIN

This occurs via activation of specific receptors on amacrine, horizontal and photoreceptor cells. Wiechman and Summers (2008) have proposed a working hypothesis for melatonin paracrine signalling in the retina (eFig. 5.2). Melatonin is normally produced by photoreceptors at night and diffuses to target cells within the retina that have specific receptors on cells such as GABA-ergic and/or dopaminergic amacrine cells, which work in a reciprocal manner to some degree since GABA inhibits dopamine release from amacrine cells. A lower rate of dopamine release from amacrine cells results in lower stimulation of D1 receptors on horizontal cells, which in turn leads to increased coupling of horizontal cells. This would result in an increase in receptive field size and increased sensitivity to light. Lower levels of binding of dopamine to D2 receptors on photoreceptor cells induces an increase in melatonin synthesis. Meanwhile, melatonin may bind to horizontal cells to inhibit directly the cellular response to D1 receptor binding. Melatonin may also bind to receptors located on the photoreceptor membrane, which could directly increase rod sensitivity to light and/or regulate synthesis of melatonin.



eFig. 5.2 Diagram outlining mechanism of how melatonin fine-tunes visual function (see text for details). (From Wiechmann et al., 2008.)

hyperpolarization of the cell membrane). Approximately 10–15 rods must be stimulated and the summed response must be collected either at the bipolar cell level or within the ganglion cells to induce a visual sensation. These determinations are approximate as fluctuations occur at all levels from the stimulus itself to the responses in each of the different cell types, and the final analysis is based on probabilities of a response taking place.

Spatial summation

As indicated previously, the empirical determination of the absolute threshold of light detection depends on the stimulus size; therefore, spatial summation must be important in setting this threshold. Each ganglion cell has a receptive field in which a light stimulus falling on a point within that field will produce a response. Receptive fields are the result of convergence of several photoreceptors to synapse with one bipolar cell and of several bipolar cells to synapse with a single ganglion cell (see next section).

Some limited general rules have therefore emerged concerning summation. Ricco's law states that the threshold intensity of a stimulus is inversely proportional to the area of the stimulus, provided the total stimulus area is sufficiently small to fit within the receptive field of a single ganglion cell. In terms of quanta, however, the amount of energy is independent of the area. As the receptive field size increases at greater distances from the fovea, Ricco's law also varies in the area in which it can be applied. In overlapping receptive fields, Ricco's law applies only partially in that larger stimuli require more quanta to reach the absolute threshold. This has led to further attempts to formulate equations that would provide a general solution for these phenomena, but in practice no simple solution covers all possibilities and summation is best explained by probability theory (see earlier).

Temporal summation

When the retina is stimulated in rapid succession by a target, the level of response is the same as when the target is presented continuously for the same total period of time. This is known as temporal summation and is formulated by Bloch's law, which states: the intensity of the threshold stimulus is inversely proportional to the duration of the stimulus. To a degree this is a difficult psychophysical measurement since at very short intervals it is difficult to distinguish different contrast and duration. Bloch's law holds true only for a defined period of time since, if the interval between the stimuli were long, the effect would be rapidly lost. Bloch and others found that, in fact, a 'plateau' effect was observed. However, Broca and Sulzer found that there was a peak in perception and then there was a decay before a plateau effect. This long-standing conundrum has been repeatedly examined and challenged in the literature (see letter, Gorea and Tyler, 2013).

Laws concerning temporal summation are described online at: https://expertconsult.inkling.com

In practice, the peak described by Broca and Sulzer occurred at about 50–100 milliseconds, and beyond this time there is still some degree of summation, known as partial summation, which decays exponentially.

The question of temporal summation has been revisited in the context of artificial light (which accounts for about 20% of our energy consumption). Using an experimental design in which intrinsic bias from previous learned experience was omitted, Broca's peak was detected and was attributed to a differentiation in duration versus contrast, which is eliminated probably by higher neural mechanisms ensuring that the same rapid stimulus is identified with short flashes of light. Bloch's law therefore represents a smoothed-out perception in which the peak of detection/contrast is eliminated by prior learned experience at a subconscious level. If artificial lighting systems were optimally tuned to these temporal summation effects in human vision, for instance by using DC light-emitting diodes, a 20% saving in energy consumption has been estimated, which is not insubstantial.

Is there a visual response to melanopsin stimulation?

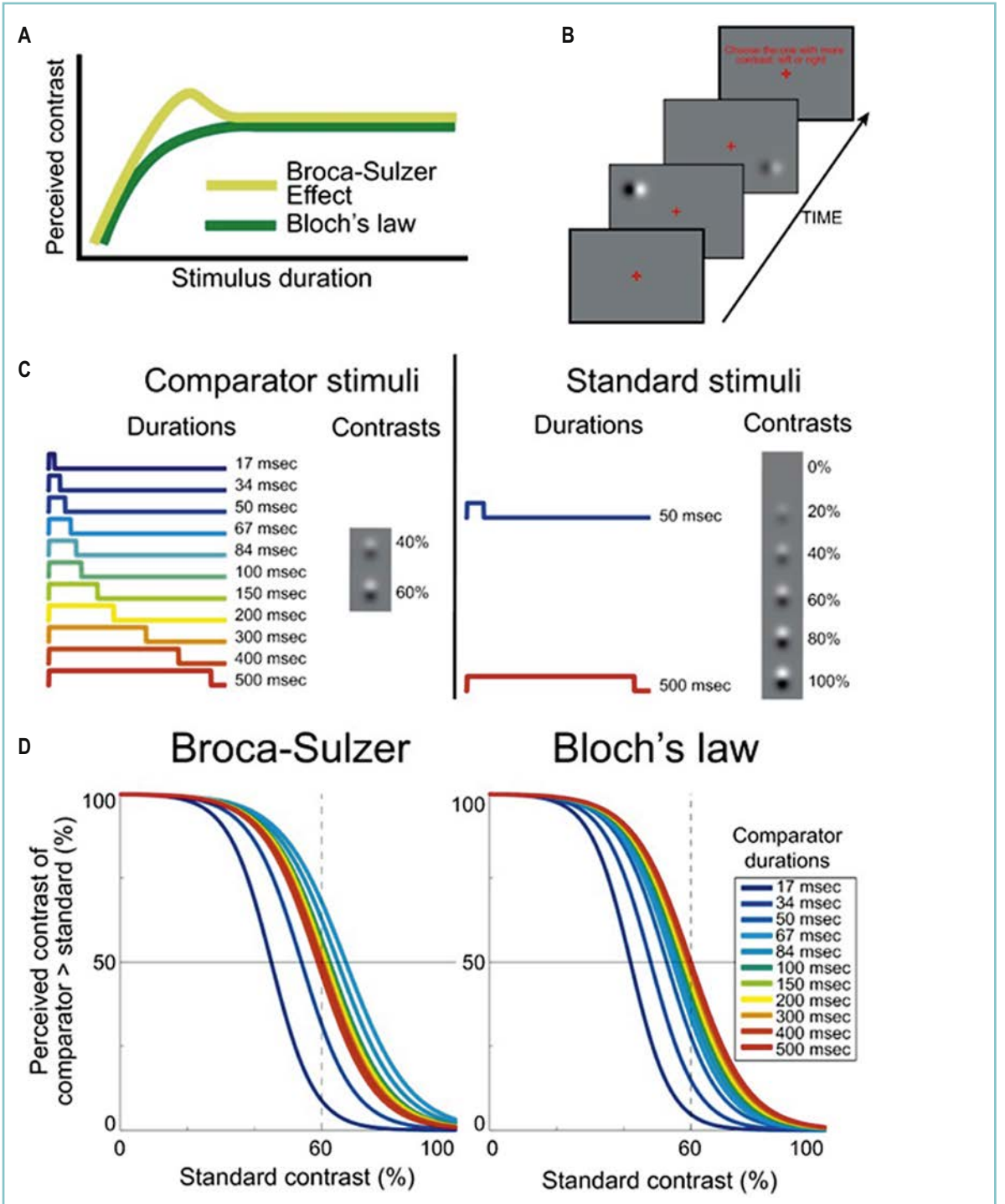
The above details address issues around the limits of detection of temporal (rod) and spatial (cone) stimuli. The role of melanopsin in pupil responses and circadian rhythms is well established. However, it is not clear whether there is also an associated visual sensation. Recent fMRI studies with a 'cone-silenced' melanopsin-selective stimulation technique, using pulsed (3-seconds) spectral modulations around a photopic background, have revealed signals in VI accompanied by the perception of a transient blurred 'visual sensation' (Spitschan, 2017).

Detection of minimum stimulus for motion displacement

An extension of these concepts has been developed to evaluate the minimum detectable motion stimulus test, since motion detection is a major function of the magnocellular ganglion cell, i.e. rod-dominated pathway (see later). The test (motion displacement test, MDT) is based on the minimum positional displacement of a standard line stimulus, which is detected as a sensation of motion. Since it is based on a square wave stimulus that oscillates back and forth between two points, it is considered to be the summation of the ON–OFF receptive field responses of the stimulated M cells. The threshold for detection varies with the square root of the stimulus energy and has been described as the threshold energy displacement law (TED).

UNDERSTANDING TEMPORAL SUMMATION

Investigation of potential different effects can be modelled experimentally. For instance, the following experiments reported recently by Rieiro and colleagues demonstrate the actual effects of the different stimuli as varying potential outcomes (Rieiro et al., 2012). In eFig. 5.3, (**A**) represents two competing models of temporal vision. Bloch's law postulates a monotonic increase in perceived contrast with increased duration, whereas the Broca–Sulzer effect postulates a peak in perceived contrast with increased duration. In (**B**) subjects fixated on a central cross and two Gabor patches flashed in succession on opposite sides of the screen. Following stimulus presentation, subjects reported which Gabor had higher contrast. Part (**C**) shows the physical contrasts and stimulus durations used for the comparator and standard stimuli. In the unblocked experiment, all possible combinations were randomized. In the blocked experiment, the different conditions were grouped into four sequential sets of trials or blocks, each with a constant comparator contrast and standard duration and an internally randomized trial sequence. Finally, (**D**) shows the psychometric curve models of the two possible experimental outcomes, colour-coded for different comparator durations. If contrast perception has a peak, as in the Broca–Sulzer effect, the curves will first shift right and then left as stimulus duration increases. If contrast perception follows Bloch's law, the curves will shift monotonically to the right.



eFig. 5.3 Studies of the potential different effects which may occur in temporal summation (see text for details). (From Rieiro et al., 2012.)

Binocular summation

Summation may occur in visual stimuli received by corresponding retinal regions when using both eyes. In practice, mostly because of optical aberrations (see later), the effect is not considered significant. However, aberrations can be removed using wavefront technologies and it has been shown that such aberrations account for between 5% and 15% of loss in visual discrimination. In a recent study, binocular summation and inhibition, defined as seeing five or more or fewer than five letters on the ETDRS visual acuity chart with both eyes compared with best visual acuity with each eye individually, occurred at a prevalence rate of 21% and 2%, respectively, which has considerable relevance to driving vision. In addition, the effect of amblyopia may be such that the loss of binocular summation has a distinct effect on overall visual acuity. A meta-analysis of summation data has shown that binocular summation may be more accurate than monocular but depends on the ratio of temporal to spatial frequency and that methodological factors influence the outcome (Baker, 2018).

Visual acuity and contrast sensitivity

Visual acuity is not simply a function of cone activity

Vision varies for each individual because of refractive errors and visual physiologists have therefore restricted discussion of normal visual physiology to the emmetropic idealized eye.



Information on optics and the Idealised eye is available online at: https://expertconsult.inkling.com

(See also Chapter 4, p. 215 Emmetropia, myopia and hyperopia.)

Visual acuity is a measure of the ability to discriminate between two stimuli separated in space. Clinically, this is determined by distinguishing letters on a chart, but this task also requires recognition of the form and shape of the letters, processes that involve higher centres of visual perception. Discrimination at a retinal level may therefore be determined by less complex stimuli such as contrast sensitivity gratings. The visual processes that allow discrimination between letters and gratings are fundamentally the same, with finer resolution contrast discrimination at lower luminance levels being provided by some newer test charts, such as the Mars contrast sensitivity charts, which are graded in log 0.04 units. Charts that have been customized to test visual acuity in different groups of people, such as the SKILL test (Smith–Kettlewell Institute Low Luminance test), might be a good predictor of eventual development of macular degeneration in older people.

Theoretically, the resolving power of the eye can be derived from an estimate of the angle subtended by a single photoreceptor and is about 1.5 μ or 20 minutes of arc (20 arcminutes, 20′) in the case of cones, because this represents the smallest unit distance separating two individually stimulated photoreceptors. This corresponds to about a pixel on a computer screen when viewed at half a metre. However, it is well recognized that the resolving power of the eye can be as great as 0.5′ of arc, for instance when looking for the gap in a Landholt C target, or 4″ of arc when viewing a thin line on an illuminated background. This hyperacuity, or Vernier acuity, is achieved by the complexities of retinal neuronal synaptic organization and is 5–10 times greater than 'standard' visual acuity, but the limits of acuity are still determined to some extent by the retinal photoreceptor mosaic or 'grain'.

The highest discriminatory capacity is subserved by cones, although a certain degree of resolution can be achieved by rods. The level of acuity, however, falls off rapidly the greater the distance from the fovea, such that at 5 degrees from the central fovea visual acuity is only one-quarter of foveal acuity. As rod and cone longitudinal dimensions are not sufficiently different to explain the marked difference in acuity and as the resolving power of the eye is greater than the theoretical limits based on cell size, other mechanisms must underpin acuity. Visual acuity is affected by the luminance of the test object and the degree of adaptation of the observer; dark adaptation increases both rod and cone acuity and therefore is not affected by the sensitivity of cones *per se*. In contrast, light adaptation increases sensitivity of cones but not rods (see earlier).

Vernier acuity is used in everyday life, for instance in measuring distance with a ruler or detecting the time on a mechanical clock. Vernier acuity is not present in infancy but reaches its highest level of function around the age of 14 years. It is absent in strabismic amblyopia but may be present in patients with anisometropic amblyopia. Vernier acuity can be measured objectively using sweep-mode visual evoked responses (Swept VEP see p. 306) in which the pattern or grating stimulus is rapidly changed over a short time period (10 seconds).

Vernier acuity is different from the recently recognized state of supervision, which has been revealed by the use of adaptive optics. Adaptive optics were developed for use in astronomy to minimize optical aberrations and correct higher-order dynamic aberrations caused by such aspects as angle of viewing and accommodation, compared with correction of static aberrations such as astigmatism and defocus (see later). In essence, by using a wave form sensor, adaptive optics measures phase aberrations in reflected light produced by the imaging light source. When applied to the eye, for instance in the use of wavefront aberrometers and wavefront-guided vision correction in refractive

THE IDEALIZED EYE

The eye has the power to refract (bend) light waves and, as for any lens, this is measured in dioptres. A dioptre (D) is a unit of measurement that describes the strength, or 'power', of a lens to bend (refract) light a set amount (degree); the optical power of a lens or curved mirror is equal to the reciprocal of the focal length measured in metres (that is, 1/focal length (m)). It is therefore a unit of reciprocal length. Thus, a three-dioptre lens brings parallel rays of light to focus at 1/3 metre. The overall refractive power of the eye is around 60D for the normal healthy emmetropic eye and much of its refractive power is attributed to the lens-like (focusing) properties of the cornea (which amount to about 40D), with the remaining 20D due to the ocular lens (eFig. 5.4). The effect of this refractive property of the ocular media (i.e. the tissues through which the light passes) is to focus light rays on the retina, and specifically the fovea, when visual acuity is being measured. Many refractive errors occur in the healthy population, including myopia (short-sightedness), hyperopia (long-sightedness) and astigmatism (non-spherical aberrations of the eye's refractive power).

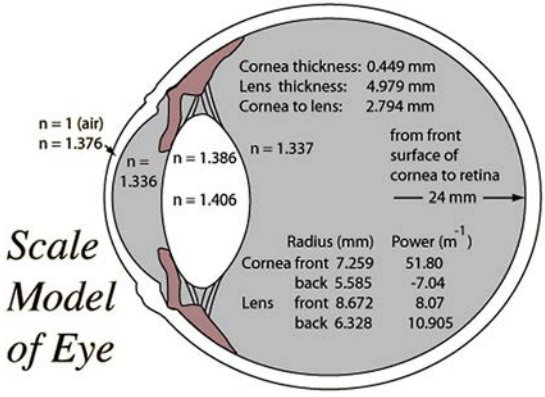
The scientific discipline dealing with the optics of the eye is known as optometry. Many textbooks discuss physiological optics in addition to optical devices such as spectacle correction, contact lenses and the optics of intraocular lenses. There are large treatises that deal in depth with physiological optics and shorter text books that summarize refraction and refractive errors for those not intending a career in optometry.

The reader is referred to the following texts as examples of (1) short comprehensive practical texts and (2) a three-volume in-depth treatise:

(1) Hunter DG, West MD. Last-minute optics: a concise review of optics, refraction, and contact lenses. 2010 (Also available online as podcasts and U-tube video presentation); Thorington, J. *Refraction and How to Refract*: (2016)

(2) von Helmholtz H. *Treatise on Physiological Optics*, vols I, II and III. Dover Phoenix Editions; 2005

In addition, several texts are available that deal with the optics of the pseudophakic eye (i.e. the eye with a prosthetic intraocular lens), including official publications of professional bodies such as the American Academy of Optometry and the British College of Optometry.



eFig. 5.4 Standard dimensions of the eye as related to the model eye. (From Hecht, 1987.)

corneal surgery, adaptive optics can theoretically increase acuity to 'supervision' levels.

Limits of and limitations on acuity

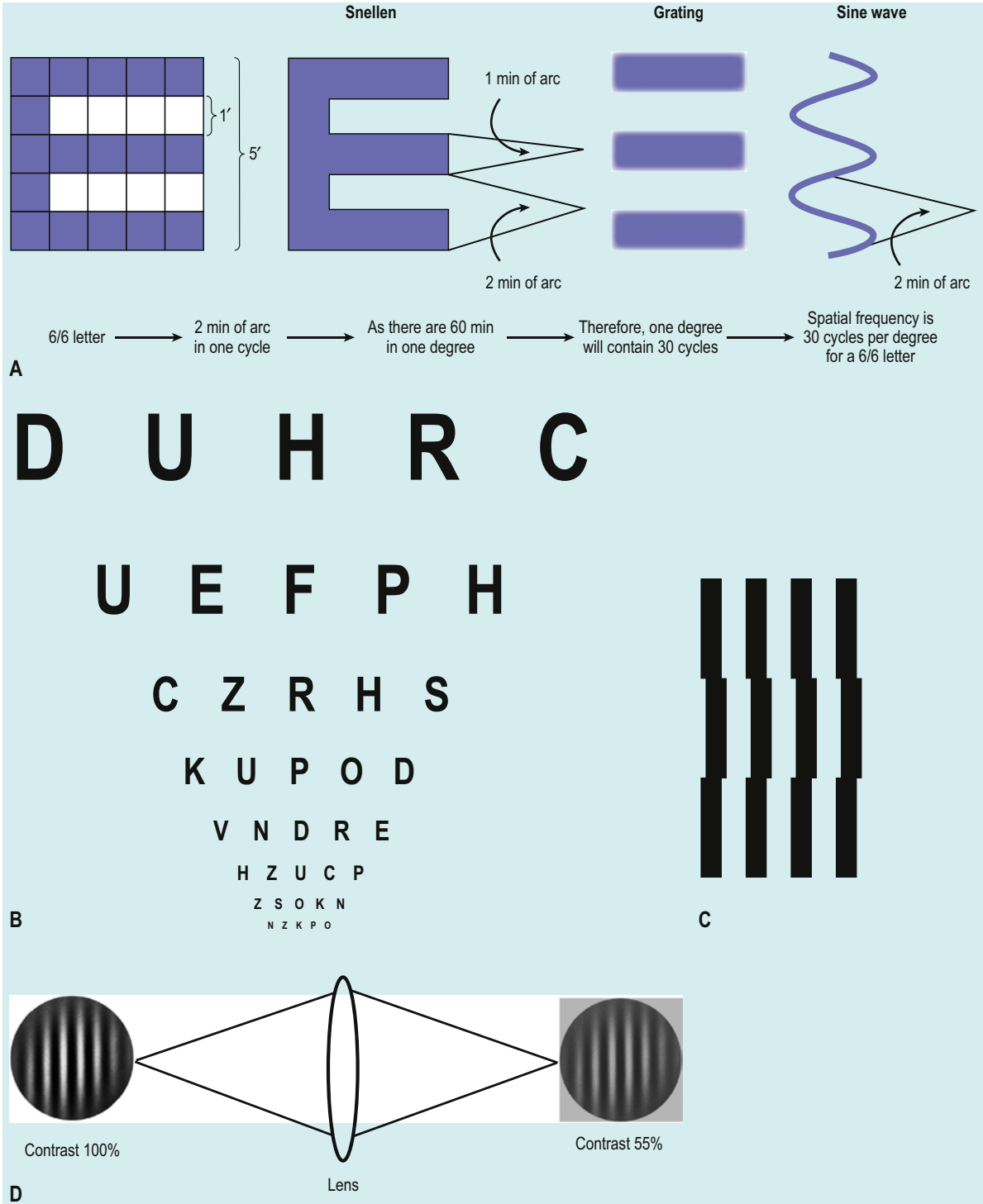
The letters on reading charts such as the Snellen test type and the ETDRS chart (Box 5.3) have been constructed on the assumption that the average person can resolve two points separated by 1′ of arc. If the limit on acuity is in part determined by the single photoreceptor theory (earlier), then a one-to-one relationship between the photoreceptor and the nerve cell must exist if there is to be no downstream loss of acuity. For foveal cones such a relationship exists between cone cells, midget bipolar cells and midget ganglion cells (see section on retinal connections, pp. 297–305), but even midget cells have some interconnections with diffuse bipolar and ganglion cells. In spite of these connections, summation of information should not occur for cells subserving the highest levels of acuity and, indeed, is absent from foveal cone cells but is characteristic of rod cells; furthermore, it has been suggested that the improved visual acuity that occurs under conditions of light adaptation is the result of inhibition of these subsidiary connections.

Visual acuity is also limited by the physical behaviour of light, such as diffraction and chromatic/spheric aberration. A single point of light small enough to stimulate a single cone will produce diffraction rings in its traverse through the pupil sufficient to stimulate more than one cone. Similarly, the prismatic separation of white light into its constituent wavelengths will lead to the stimulation of several cones of different types. It is clear therefore that resolution of images must be achieved at a post-receptor level and is in fact a function of the receptive field of each ganglion cell unit. Where there is minimal convergence of information from each receptor, i.e. where the one-to-one relationship between receptor and bipolar cell is maintained, then resolution is at its highest and this occurs at the fovea. However, where there is increasing convergence of information, such as with several parafoveal cones synapsing with one bipolar cell, resolution obviously decreases.

The one-to-one relationship, however, does not adequately explain hyperacuity or Vernier acuity. Diffraction and spheric/chromatic aberrations have ruled out the concept of single unstimulated cones occurring between neighbouring stimulated cells. However, it is likely that discrimination is more a matter of degree than absolute responses, i.e. that resolution is achieved by certain receptors being *less stimulated* than their neighbouring receptors on either side. This is likely to occur with diffraction, where alternating light and dark rings emanate from a point source of light, and with chromatic aberration, where different wavelengths of light are likely to stimulate their respective neighbouring cones to different degrees (Box 5.4). In fact, this fits with the nature of the hyperpolarization response being a graded one in retinal neural cells: as indicated in Chapter 4 (see p. 266), only ganglion cells fire 'classic' depolarizing action potentials, while other retinal cells have a graded analogue-type 'tunable' electrical response. This differential stimulation from cones is registered with the respective bipolar and ganglion cells and, if combined with a minimal degree of receptor convergence, can explain high levels of discrimination. In this way diffraction could explain, at least partly, the ability to resolve a break in a line subtending an angle of less than 10′ of arc, since partial diffraction lines deriving from the edge of the break would ensure differential stimulation of cone receptors over a very small area. Stimulation of any particular cone is also likely to induce local inhibition (via receptive field mechanisms; see later) in neighbouring cones, thus enhancing resolution and sharpening the image further.

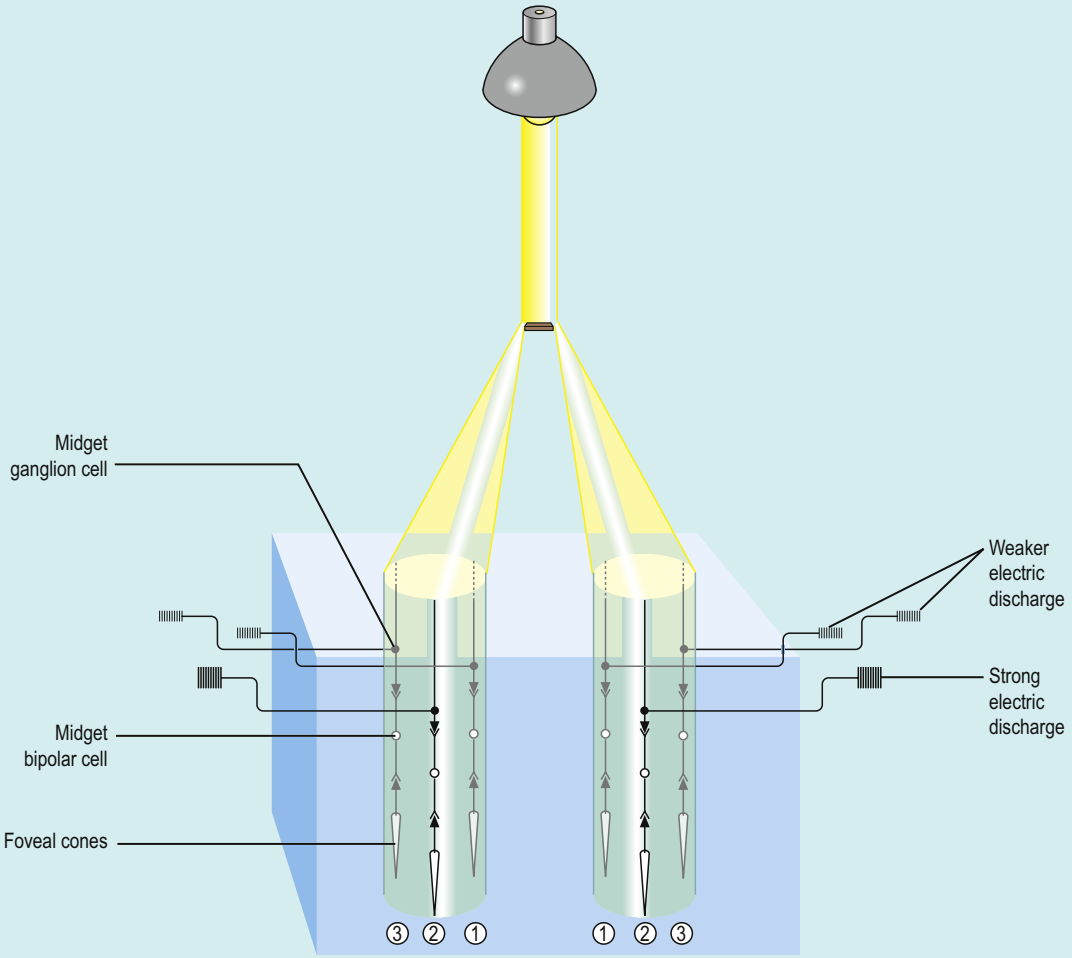
Box 5.3 **Assessment of visual acuity using standard letter charts**

Visual acuity in clinical practice is determined as an empirical value based on the assumption that the cone photoreceptor has the ability to discriminate between two objects in space subtended by an angle of 1 minute of arc at the nodal point of the eye (**A**). This is measured using a set of charts (optotypes) and standard normal visual acuity equates to the vision of 6/6 or 20/20 (i.e. 1.0 or 100%) when viewing a predetermined standard target (size of optotype letter) at 6 m (UK) or 20 feet (USA). Test conditions describing the ambient illumination and the illumination of the letters on the chart to provide contrast are also arbitrarily set. The Snellen chart is based on the concept that the smallest spatial target that can be resolved subtends 1 minute of arc at the nodal point of the eye (see earlier) and although theoretically inaccurate, it serves as a useful parameter. The Snellen chart has rows of letters of decreasing size and is arbitrarily set to produce the standard test at 6 m, although other charts with proportionally smaller letters can be used at shorter distances. The LOGMAR (LOGarithm of the Minimal Angle of Resolution) (**B**) is more precisely designed with definitive sizing and spacing of the letters and can provide a more quantitative evaluation of visual acuity. It therefore tends to be the standard for use in clinical trials. As indicated in the text, spatial acuity better than 100% can be achieved, for instance when discriminating the 'offset' of a line or edge (**C**). This is termed Vernier acuity. In addition, visual acuity is modified by such factors as glare and contrast and can be measured as in the contrast sensitivity test using a sine wave grating as shown in (**D**), where diffraction and aberrations have degraded the contrast of a sinusoidal grating pattern.



From Schweigerling, 2000, with permission from Elsevier.

Box 5.4 **Differential activation of neighbouring cones determines the limits of visual acuity**



A beam of light interrupted by a small target will produce diffraction rings at its edges: the central rays in each ring will stimulate one cone (2) more than the weaker peripheral rings will stimulate its neighbours (1) and (3). Our ability to detect the break in the light beam is determined by comparing the differential responses in all the cones in the illuminated region and finding two that produce a similar response to correspond to the edges of the target. Visual acuity is therefore a measure of the retina's ability to produce different graded responses and not absolute responses.

These concepts are embedded in the canon of knowledge going back to the time of Rayleigh in the late nineteenth century. However, Rayleigh's law does not fully explain the real condition where diffraction spreading from a double line is slightly greater than that from a single line for the same total amount of light energy. Information theory, adaptive optics and modern electro-optical devices for generating discrete stimuli might help to explain this anomaly. Studies have, in fact, shown a light/dark background to the target stimulus affects a light stimulus more than a dark stimulus and creates a 'neuronal blur', which is due to differences in stimulation of the ON/OFF pathways (see later) and has been suggested to influence the progression of myopia (Pons et al., 2017). This phenomenon has been known to artists for centuries.

These considerations also have a number of implications. In particular, the resolving power of the eye is limited by the distance between two images such that a single cone or set of cones is appreciably less stimulated than the rest; the limit of resolution is therefore not an absolute

determinant, but depends on conditions such as light and dark adaptation, background illumination and other factors. Most importantly, it depends on the degree of dendritic connections that occur between the affected cones and the neural cells and activation of ON vs OFF neurones.

The resolving power of the eye therefore depends on:

- the distance between two objects
- the degree of light and dark adaptation
- the background illumination
- the extent of the dendritic connections between the cone and neurones.

Contrast sensitivity

Visual acuity is also affected by contrast (sharpness). The finest limits of resolution have been determined by the ability to discriminate a thin white line against a uniform background illumination (0.5′ of arc). The effects of diffraction are such that detection of this line depends on the liminal brightness increment (l.b.i.). This increment represents the endpoint at which the differential in brightness between the individual dark/bright oval diffraction rings produced at the edge of the line can be detected; if they are not sufficiently different from the background luminosity, the line will not be detected. The l.b.i. is determined by the contrast between the light and dark lines and can be measured quantitatively with a sinusoidal grating (see Box 5.3): a spatial pattern where the average luminance remains the same but the contrast between the light and shaded areas can differ.

The degree of contrast (C) is relative to the background luminance (L) and is described in terms of the maximum (L_{max}) and the minimum (L_{min}) as follows, also known as the Michelson contrast:

$$C = (L_{max} - L_{min}) \ / \ (L_{max} + L_{min})$$

A higher Michaelson contrast of pupil over iris has been shown to determine how we assess a person's direction of gaze (Antsis, 2018).

Alternative measures are the Weber contrast, where the difference between the maximum and the minimum is compared against the background luminance, and RMS contrast, where the mean luminance is a factor of the standard deviation in luminance. The Michelson contrast is used for gratings and a threshold is reached when the target is detected reproducibly. Sensitivity is the inverse of the threshold and thus contrast sensitivity is a measurable quantity, usually described in cycles per degree (the grating frequency) and visual acuity is equivalent to 1/grating frequency. Using this method, it has been shown that there is a peak response in the middle range of frequencies (Fig. 5.7).

Contrast sensitivity is therefore set by the limits of the grating frequency and is affected by both the optics of the

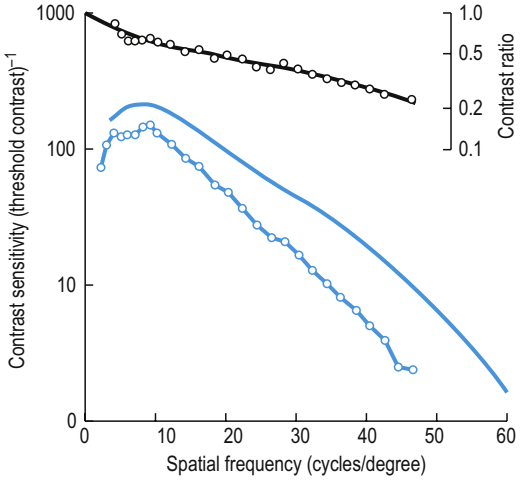


Fig. 5.7 Contrast sensitivity curve showing peak response at midspatial frequencies.

system and the direction of the grating lines, being most sensitive in vertical or horizontal directions. Remarkably, threshold contrast for many targets sits around 1% independently of target size or brightness, which as Peli and Bex observed remains unexplained since originally described by Fechner in 1860 (Peli and Bex, 2013).

Contrast sensitivity above threshold is, as for any measure of acuity, affected by luminance. In addition, bar width, length and grating motion all affect sensitivity. In the latter there is likely to be significant cortical processing at this level, as there is for 'line modulation sensitivity', a technique whereby grating bars are composed of wavy lines and the subject is asked to determine whether the line is straight or not. It has been shown that the depth of curvature of the wavy lines is a major determinant of decision-making in this test and has been described as the curve frequency sensitivity function (CFSF). This technique can provide highly sensitive measures of acuity.

Contrast sensitivity measurements, while an excellent measure of acuity, are sensitive to phase shifts and grating orientation and for absolute measures of object detection, object contrast is critical. In this context, mesopic low-contrast letter acuity is the most sensitive method for revealing small differences in retinal image 'quality', which influences recognition as opposed to detection of an object. The contrast sensitivity function (CSF) has been arbitrarily measured using a set of five spatial frequencies and has been found to be relatively robust. Age and decreased luminance cause a shift to larger frequencies in the CSF. Glare, which is often a side-effect of refractive surgery, affects the CSF at low rather than high spatial frequencies. For clinical purposes, measurement of the CSF (as opposed to measuring the threshold as in most contrast sensitivity charts) is very time consuming and impractical, but recent developments

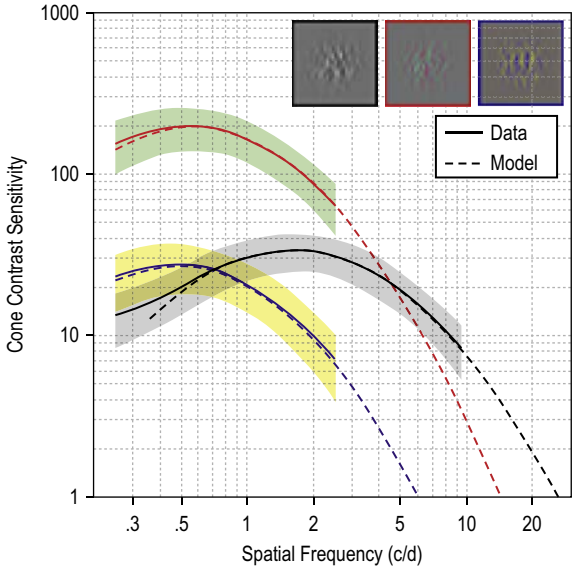


Fig. 5.8 Measured CCS as a function of spatial frequency for the Ach (*solid black line*), RG (*solid red line*), and BY (*solid blue line*) conditions under monocular viewing. The average across the 51 subjects is shown. The *dotted lines* indicate the log-parabola model estimation, which is reconstructed with the average estimated values for each of the three parameters by the qCSF. The averaged model parameters are reported in the Table. The *shaded regions* represent 6 SD. (from Kim et al., 2017.)

using customized and selective spatial frequencies and contrasts such as Gabor patches allow tailored CSF tests to be applied to specific conditions, e.g. macular degeneration. Internet/computer-based CSF tests are now used routinely. In addition, a new tool has also been developed (Curveball), which correlates CSF with tracking smooth eye movements (see p. 346) and allows the test to be used in individuals who cannot communicate verbally (Mooney et al., 2018).

Wavelength also affects contrast sensitivity such that at high spatial frequencies the gratings appear to be of the same colour, whereas at low frequencies (i.e. with coarse gratings), colour differences can be detected. CSF measurements were highest for L/M (red–green) cones while S cones (blue–yellow) were similar to achromatic (black–white) CSF measurements (Fig. 5.8). Discrimination is poorest with red–green, however, suggesting that for low frequencies rod–cone interactions are important in achieving best visual acuity. Interestingly, contrast sensitivity appears to induce more electrical signal responses in M cells, generally thought to subserve rod function, than in P cells, which are linked to cone function (see later).

Does the retinotopic arrangement of fibres in the cortex have a bearing on acuity?

The representation of retinal ganglion cells in the LGN and cortex is disproportionately larger for foveal midget cells than for ganglion cells elsewhere in the retina. This produces a cortical magnification factor for foveal cones over other cones. However, the magnification factor is not related solely to the reduced convergence of foveal cones on ganglion cells (see later), but also to a disproportionate LGN and cortical representation of neurones served by foveal cones. fMRI studies have revealed the retinotopic map of the human visual cortex and demonstrated this magnification factor in technicolour (see also p. 333) (Fig. 5.9).

Best-corrected visual acuity: effects of external factors

Visual acuity is, of course, affected by factors that do not relate directly to the retinal stimulus. These include pupil size, eye movements and binocular viewing.

Pupil size and visual acuity testing in infants

The size of the pupil affects the level of visual acuity in that a reduction in pupil size reduces aberrations but increases the effects of diffraction. Below 3 mm, these effects tend to cancel each other out and visual acuity is independent of pupil size, although wavefront aberrometry reveals that a pupil size of 2.5–3.0 mm produces the best image quality.

The level of visual acuity attained may also have the reverse effect on the size of the pupil. Luminance affects the level of acuity and the size of the pupil is affected by the light level via well-characterized pupil reflexes (Box 5.5). The size of the pupil also indirectly affects the visual acuity by reducing the amount of light entering the eye when the light stimulus is intense and, conversely, increasing light capture under dim lighting conditions. The three-way relationship between light capture, lighting conditions and stimulus intensity has been used to develop an objective measure of visual acuity, which may be useful in assessing vision in infants and others who are not able to cooperate in standard visual acuity testing. The test uses a high-resolution infrared pupillometry device to show changes in the amplitude of constriction in response to sine-wave gratings presented on a uniformly illuminated test background. As for contrast sensitivity, there is a peak response in the middle range of frequencies and the threshold for response correlates well with contrast sensitivity estimates of acuity and with objective sweep VEP tests. This pupil response is governed by higher visual pathways, being altered in patients with

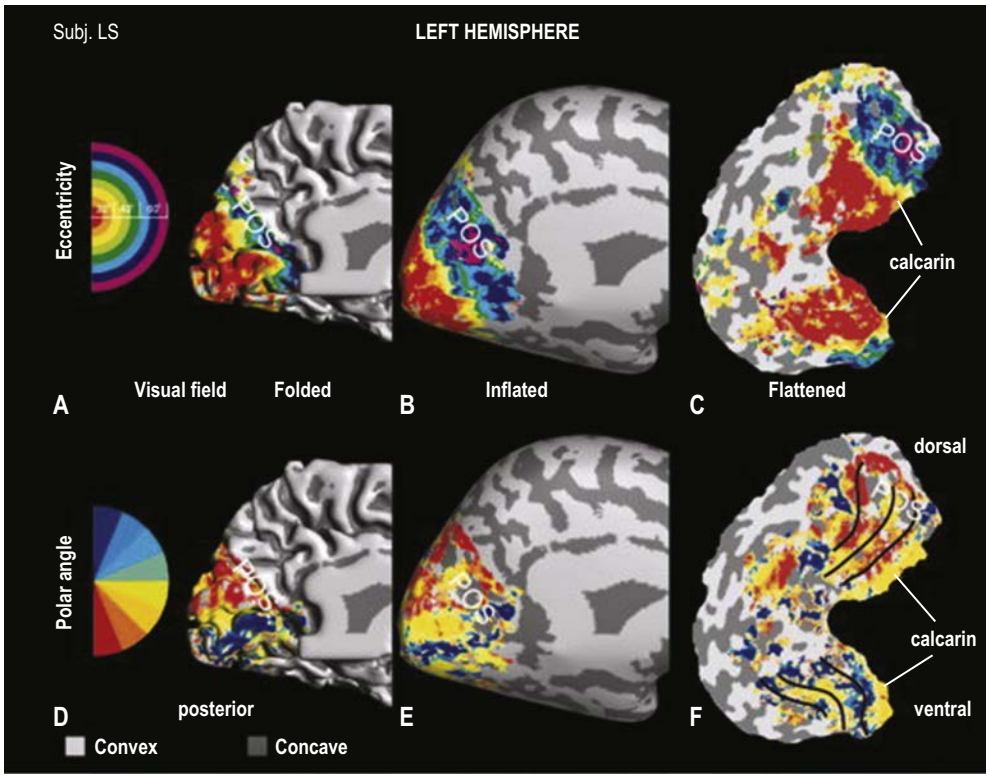


Fig. 5.9 Dipolar map and isopolar angle maps of human visual areas. These images were prepared using fMRI technology and computer modelling to 'flatten out' the visual cortex (**C** and **F**). The top row shows a map of the occipital cortex indicating the retinotopic location of the stimulus as its eccentricity increases from the fovea; the dipole is coded by colour (*brown* [fovea] -> *orange* -> *blue* -> *cyan* [periphery]) displayed on (**A**) the original cortical surface, (**B**) the unfolded cortical surface, and (**C**) the cut and flattened cortical surface. The bottom row shows the polar angle of the stimulus: *blue* (upper vertical meridian) -> *green* (horizontal meridian) -> *red* (lower vertical meridian) plotted on the same three surfaces (**D–F**). (From Wu et al., 2012.)

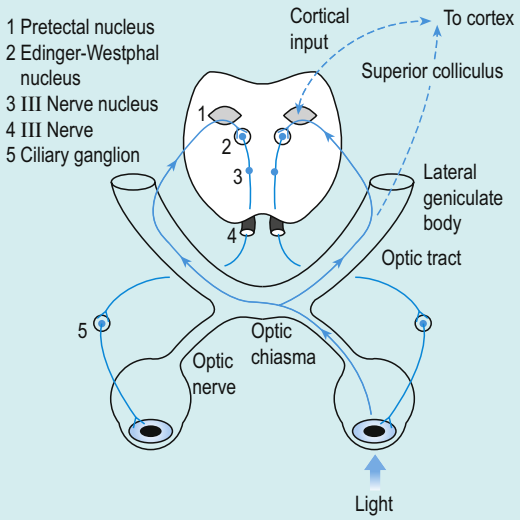
hemianopia but normal pupil light reflexes; indeed, the phenomenon is well recognized by clinical neuro-ophthalmologists. Infrared pupillometry has been shown to be valuable in studies of delayed visual maturation and to be significantly more reliable than the Rosenbaum card method in which subjective comparisons of pupil size are made.

The pupillary response via the intrinsically photosensitive retinal ganglion cells (ipRGCs) through melanopsin (see Chapter 4, p. 256) can be measured by chromatic pupillometry (stimulation of pupil responses at different wavelengths) to separate the contributions from rods, cones and ipRGCs. Rod- and melanopsin-driven responses correlate with stimulus luminance and size as might be expected. However, cone-driven pupillary responses are strongly dependent stimulus intensity but much less so on stimulus size (Park and McAnany, 2015). There are several types of ipRGCs with several different functions. For instance, the sleep/wake circadian rhythm responders connect with the suprachiasmatic nucleus, while the pupillary responsive neurones synapse in the pretectal (olivary) nucleus. In terms of response to light, ipRGCs are considered to register irradiance or radiating light. They may also be responsible for photoallodynia (the photophobia/light aversion response to very bright light).

Do fine eye movements maintain retinal stimulation under steady fixation?

The concept that the continuous fine eye movements that occur as part of normal viewing (ocular trembling) are important in ensuring constant stimulation of the photoreceptors to maintain image perception has been mooted since it was first recorded in 1738 by Jurin (Martinez-Conde and Macknik, 2017). Indeed, it has been shown that images received by peripheral receptors fade rapidly if fixation is deliberately maintained in one

Box 5.5 **Pupillary light reflexes**



- The afferent response commences in photoreceptors, is transmitted to retinal ganglion cells, enters the optic nerve, decussates at the chiasm, traverses the optic tract and terminates in the pretectal nucleus (bypassing the lateral geniculate nucleus).
- Both crossed (via posterior commissure) and uncrossed fibres pass from pretectal nucleus to Edinger–Westphal nucleus (parasympathetic).
- Parasympathetic fibres pass to the III nerve nucleus and leave the brainstem via the III nerve. Fibres synapse in the ciliary ganglion before supplying sphincter pupillae of iris (constriction) via short ciliary nerves.
- Uniocular light stimulus therefore gives rise to bilateral and symmetrical pupillary constriction.
- Melanopsin signals through ipRGCs (see text) to the suprachiasmatic nucleus (circadian response) and the pretectal nucleus (irradiance response).

position – the Troxler phenomenon. There are three different types of ocular movement during steady gaze: microsaccades, drift and tremor (Fig. 5.10). The threshold or perception is raised at the peak of the microsaccade (microsaccade suppression) as well as in classic saccadic movements (see later). Drift occurs between microsaccades and is also important for preventing image fade while tremor and drift are simultaneous. Tremor is the smallest of the three movements and is about one photoreceptor wide in amplitude. It is considered the most difficult to investigate but the most important for maintaining image. Although this was originally considered to be a mechanism for enhancing the central image by inhibiting peripheral images, use of a stabilized retinal image has shown that elimination of these fine movements does not necessarily lead to a reduction in visual acuity. However, these findings were obtained using high-frequency gratings and it is possible that fine eye movements may be important at lower spatial frequencies in improving contrast.

Fine eye movements occur during different visual tasks: for instance, during reading, the fixation time on the target letter is around 200–250 milliseconds and the average saccade is about 8–9 letters. This increases in skim reading but the level of cognition (the perceptual span) is reduced. Useful information is gathered from a region about 3–4 letters to the left of foveal fixation and 8–9 letters to the right.

Binocular viewing and the probability theory of visual perception

Perception is a relative occurrence and depends on many factors to achieve optimal levels (there is a significant element of chance in achieving this optimum, which can

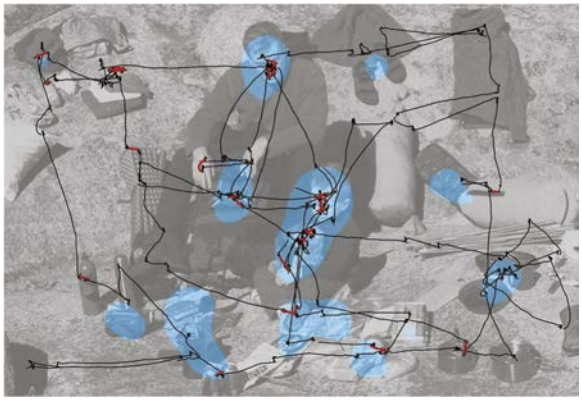


Fig. 5.10 Microsaccades concentrate on scene locations that are consistently fixated across observers. Consistently fixated regions are shown in *blue*; inconsistently fixated regions are shown in *grey*. *Black* traces indicate the eye positions of one subject over 45 seconds of free exploration; *red lines* indicate microsaccades. (From Martinez-Conde, 2017.)

be expressed as a linear transformation of log odds of frequency and/or probability). Interestingly, determination of some stimuli such as negative (concave) contours versus positive (convex) contours has a greater chance of detection. The differences have also been explained to be stereoptic (see later) depending on angle of view.

It follows therefore that two eyes are better than one, at least in increasing the chances of the highest level of visual processing of the same image.

Electrophysiology of the visual system

The transmission of nerve impulses in retinal receptors and neurones is mediated, as might be expected, by recordable changes in electric potential across the cell membrane (see Chapter 4, p. 166) and is accompanied by electric discharge. The action potential is usually an all-or-nothing event and, in muscle tissue, does not occur in the resting state. However, in neural tissue, continuous discharge may be taking place and information is relayed by changes in the frequency or rate of electric discharge in the nerve, an increase in frequency usually representing stimulation and a decrease representing inhibition, thus emphasizing the essential binary nature of biological information systems (just like computers!). This applies for all nerves in any system: the character of the received sensation is determined not by the type of nerve but by the site of information relay in the cortex and its subsequent processing in the brain.

In the retina, these general principles hold true for retinal ganglion cells, but in bipolar, horizontal, amacrine and photoreceptor cells the electrical response is more of a tonic or graded response, and the direction of the response can be positive or negative. For instance, it is this graded response that permits spatial discrimination via differential responses to diffraction rings, as described previously (see p. 286). However, the graded response in the bipolar cell becomes an ON/OFF response in the ganglion cell. In a sense retinal information is converted from an analogue signal to a digital signal at the final stage of retinal processing, i.e. at the connection between ganglion and bipolar cells.

The electrical response is initiated by phototransduction

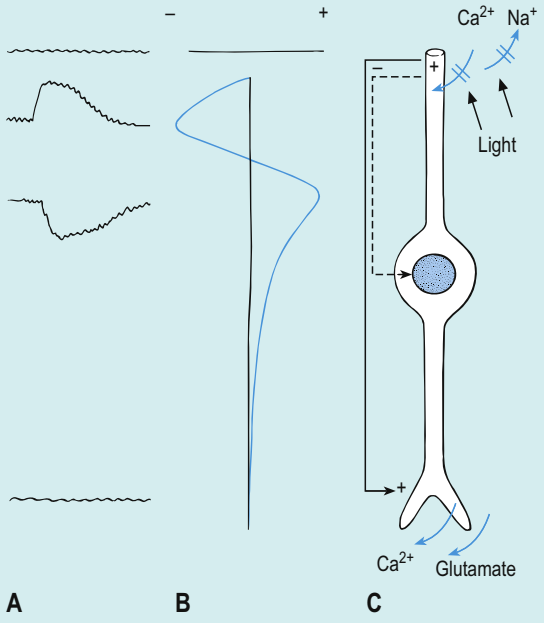
As we have seen (see Chapter 4, p. 267), when a photon of light strikes the photoreceptor outer segment, conversion of rhodopsin to the activated molecule induces a series of molecular events culminating in an electrical response. Cells, and particularly neurones, normally exist in a 'charged' state in that the inside of the cell is 'negative' with respect to the extracellular environment, creating an electrical potential difference across the cell membrane. This condition is maintained by differential distribution of Na^+ and K^+ ions on either side of the cell membrane, regulated by the $Na^+/Ca^{++},Ka^+$ exchange channel protein NCKx2 (Chapter 4, p. 268). When a neurone is stimulated, there is an initial period of gradually increasing positivity (the generator potential), which culminates in a spike discharge characterized by a rapid depolarization response of the cell. This is achieved by the rapid influx of Na^+ through ion channels that are 'opened'.

In the photoreceptor, the reverse situation occurs. Under resting conditions in the dark, the outer segment is maintained in a depolarized state through open ('leaky') Na^+ channels, which permit the influx of sodium ions from the extracellular space. When light stimulates the outer segment, the sodium channels are abruptly closed, stopping the influx of sodium and thereby leading to a reduced level of depolarization, i.e. a relative hyperpolarization (Box 5.6). This is a direct result of rhodopsin isomerization and is mediated by amplification mechanisms involving cyclic guanosine monophosphate (cGMP) (see Chapter 4, p. 268).

The hyperpolarization response is transmitted by a flux in calcium ions along the length of the photoreceptor to the synapse with the bipolar cell (the Ca^{2+} wave), which is then induced to release its transmitter (glutamate). Bipolar cells may then adapt to one

Box 5.6 **Dark currents**

Dark currents occur in the resting state (dark adapted eye) owing to 'Na+-leaking' outer segments.



The conversion of light energy to an electric response is dependent on specialized ion channels that tightly control the permeability of the cell membrane to Na^+ and Ca^{2+}. Light stimulation reverses the dark current by closing Na^+ channels in outer segments and releasing Ca^{2+} (and glutamate) at synapses. The cGMP-gated Ca^{2+} channel and the Na^+/Ca^{2+}, K^+ exchanger are located in the plasma membrane of the photoreceptor, not in the disc stack, but are complexed together with peripherin/rds-rom-1, an integral protein of the disc rim.

of two responses to glutamate, depending on which type of receptor is induced: an ON response, which is a hyperpolarized state, and an OFF response, which is a depolarization response (see later). Indeed, the hyperpolarized state conferred on the bipolar cell is also transmitted to the horizontal cells in the same region. However, the hyperpolarization response of the bipolar cell is not as steep as that of the photoreceptor in the excited state. The protein caveolin 1.3 appears to be important in rod-bipolar cell synaptic plasticity and glutamate release and has been implicated in Usher's syndrome, while Caveolin 1.3 knockout mice have reduced dark adapted ERGs.

It will be obvious, therefore, that not only is there a resting potential difference across the photoreceptor cell membrane, but there is also a potential difference along the length of the photoreceptor in the dark between the relatively depolarized outer segment tip and the hyperpolarized synaptic region of the cell at its interaction with the bipolar cell. This generates the 'dark currents' in the eye, which are reversed by the photic current on light stimulation when the photoreceptor tip becomes hyperpolarized (see Box 5.6).

Electrophysiology of single retinal cells

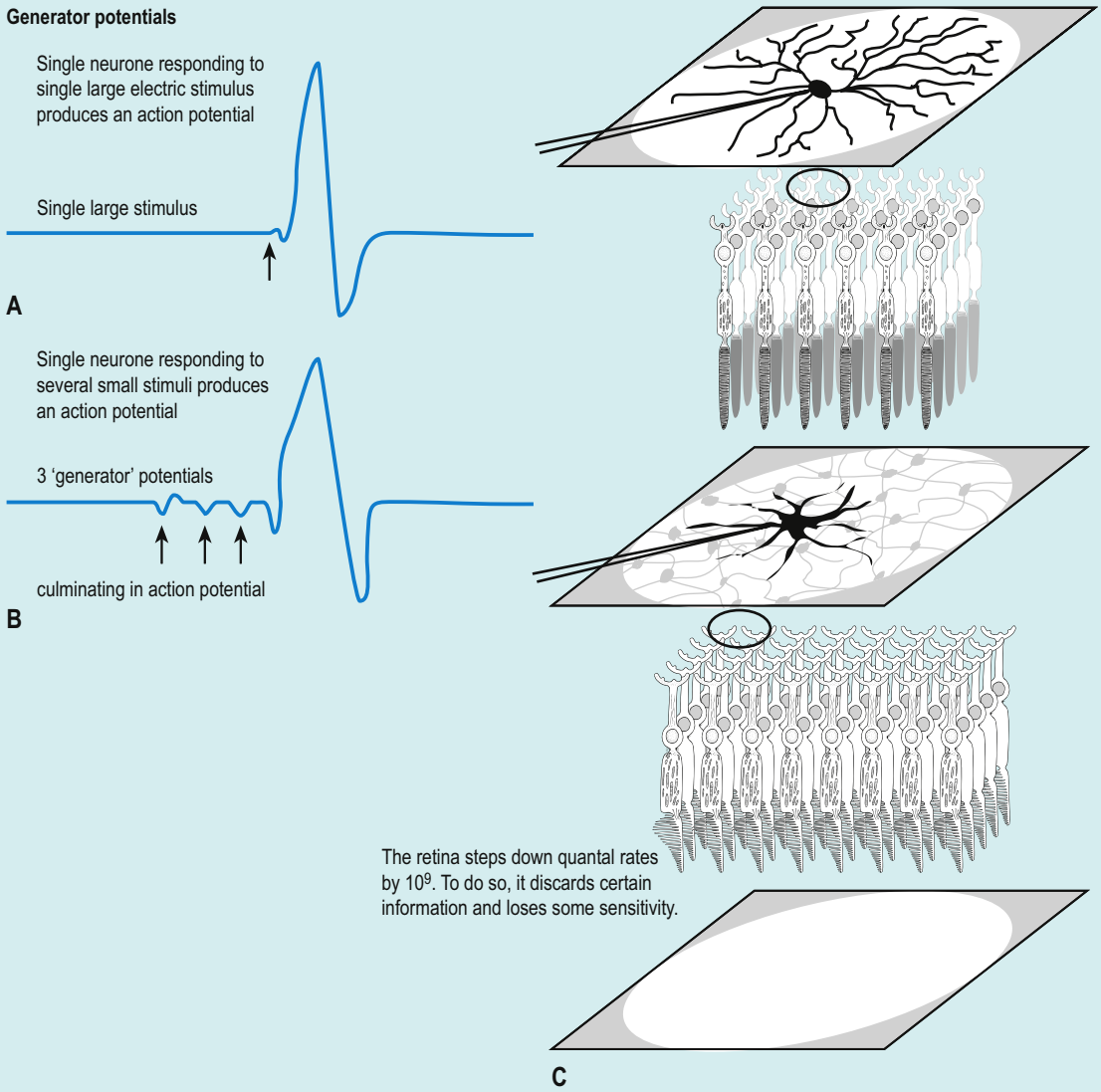
Early studies in this field concentrated on the large single neurones that could be obtained from invertebrate eyes and showed that typical action potentials could be obtained, usually preceded by a generator potential (Box 5.7). Surrounding neurones were usually inhibited when action potentials occurred in a single nerve.

Later studies of electrode-impaled optic nerves in vertebrates showed that the rate of discharge in certain nerve fibres increased (ON response) when a light stimulus was presented to the eye while it decreased in other fibres (OFF response). Yet others produced an ON/OFF response. As it was known that there were 150×10^6 photoreceptors but only 1×10^6 optic nerve fibres, it followed that many receptors must feed information into a single neurone, i.e. there must be convergence of signals and some of these must be inhibitory while others are stimulatory. On this basis, the concept of receptive fields was developed and confirmed by direct experimental testing on isolated optic nerve fibres using discrete spots of light to stimulate the retina (Box 5.8). Several phenomena, such as summation, which could previously be inferred only from psychophysical experiments, were directly confirmed. Indeed, summation effects could be compared with the effects of generator potentials in other systems that, in ganglion cells, are called synaptic potentials. The ON/OFF response in the optic nerve fibres has been shown to correlate with the centre/surround organization of the receptive field and is based on interneuronal interactions causing inhibition in surround cells.

ON/OFF receptive fields apply to cone–bipolar–ganglion cell circuitry while rod cells synapse directly into cone–cone circuits (see later). ON/OFF receptive field organization applies to several different types of signal, including light–dark, blue–yellow and red–green (see later). Amacrine cells and horizontal cells considerably modify the ON/OFF microcircuitry organization and, variably so far, each type of circuit. In addition, the receptive fields of retinal neurones depend on the size of the cell: big ganglion cells (magnocells, M cells) (Box 5.8) have large receptive fields and small ganglion cells (parvocells, P cells) have small ones. M cells receive information from

Box 5.7 **Generator potentials**

(**A**) Single neurone responding to single large electric stimulus produces an action potential.



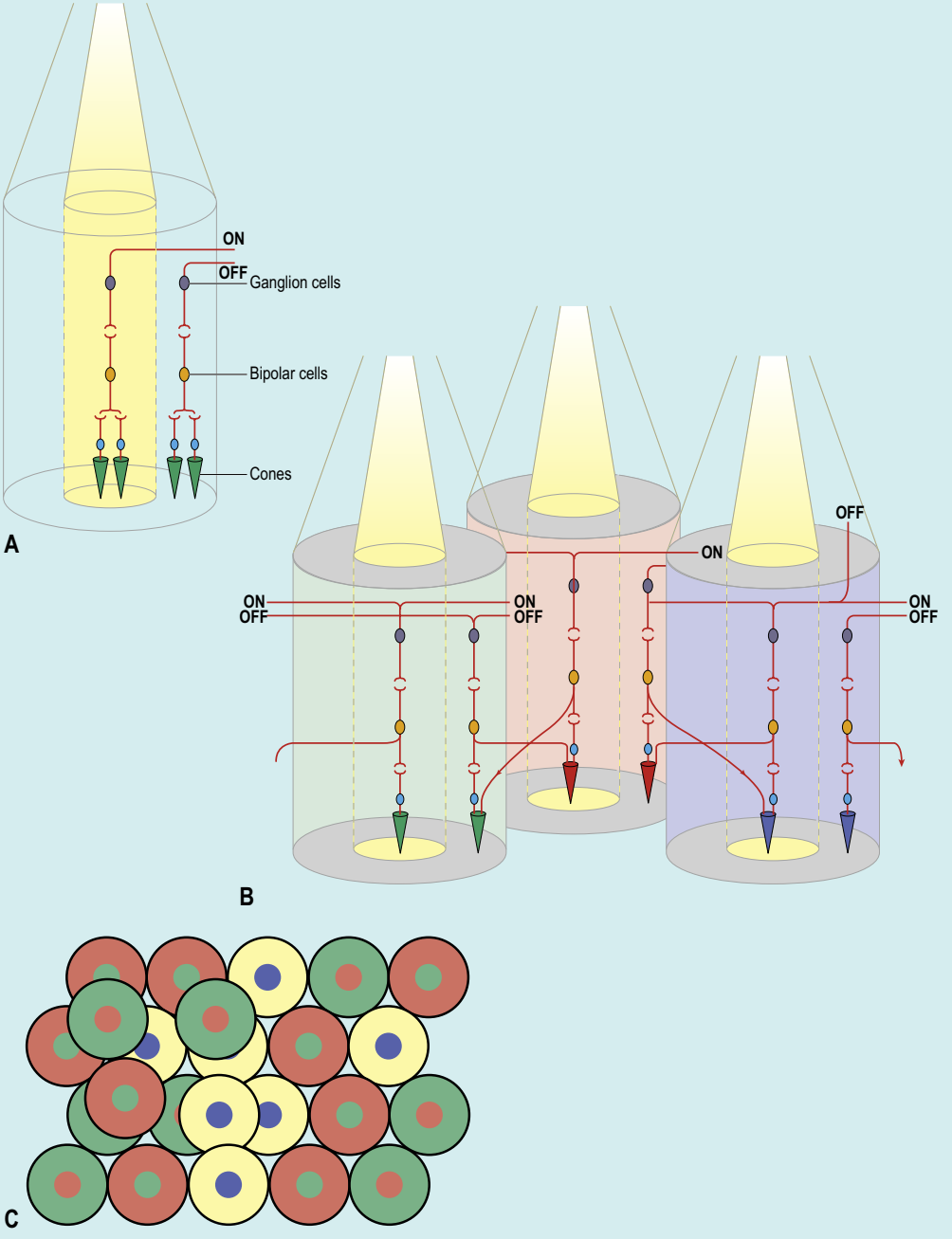
(**B**) Single neurone responding to several small stimuli produces an action potential.
(**C**) The retinal photoreceptor-bipolar cell electrical response steps down in quantal rates in a graded fashion. For instance, three 'summed' small light responses produce a typical ON response in ganglion cell neurones:*(top)* rod bipolar cell; *(middle)* cone bipolar cell.

many amacrine and bipolar cells, producing a high degree of convergence. P cells have small receptive fields receiving information directly from single or only a few bipolar cells (see later).

New data indicate that receptive fields are not static but are subject to change, probably as a result of synaptic pruning. In addition, the complexity in the types of ganglion cells (>30 types have been described) plus their many

Box 5.8 **Organization of visual information into discrete receptive fields**

The receptive field of a retinal ganglion cell neurone is the area of retina covered by that cell in which a light stimulus alters the frequency of discharge (**A**). The size of the receptive field varies from 200 to 600 μm in diameter. Receptive fields may have an ON centre (increased rate of discharge) and an OFF surround (decreased rate of discharge), an OFF centre and an ON surround, a double opponent ON/OFF centre/surround organization, or a pseudodouble opponent centre/surround organization (**B**); in addition, these arrangements occur for detection of dark–light, red–green and yellow–blue. Moreover, the dendritic connections of bipolar cells can overlap, allowing each cone to be contacted by more than one bipolar cell and, where overlap occurs, they share cones (**C**); this, in effect, means that at ganglion cell level, each ganglion cell may have its ON–OFF receptive field but this is based on input from a range of cones allowing some degree of colour mixing in what is essentially a hexagonal array (see Fig. 5.10).



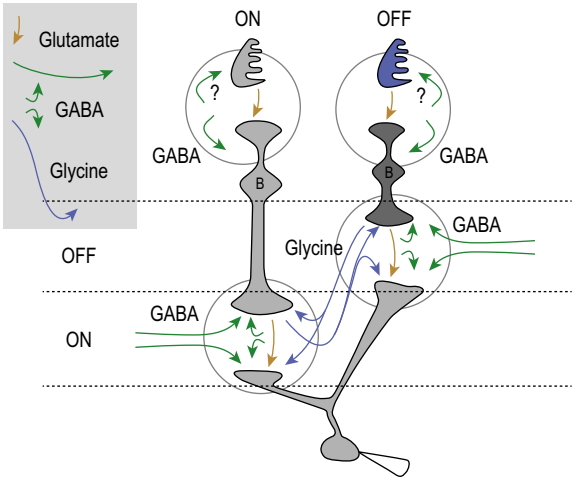


Fig. 5.11 Crossover inhibition occurs between bipolar, amacrine and ganglion cells. The *blue arrows* show the pathways for crossover inhibition acting at bipolar and ganglion cells. ON bipolar, amacrine and ganglion cells receive glycinergic OFF inhibition and OFF bipolar, amacrine and ganglion cells receive glycinergic ON inhibition *(blue arrows)*. Narrow field ON and OFF GABAergic amacrine cells *(short green arrows)* receive glycinergic inhibition. Wide field ON–OFF amacrine cells *(long lateral green arrows)* receive no inhibition. GABAergic amacrine cells *(all green arrows)* feedback to bipolar cells and forward to ganglion cells but not to other amacrine cells. *Red arrows* indicate excitatory pathways. This circuitry is verified by measurements of excitation and pharmacological block of inhibition in each cell type in many previous studies. All crossover signals could interact at the diad at the bipolar terminal as suggested in Fig. 5.2. (From Werblin, 2010.)

different cortical projections, apart from the visual cortex, are being charted.

How does the photochemical change in the receptor produce the spike discharge in the retinal neurone?

Horizontal cells at the receptor–bipolar cell interface and amacrine cells at the bipolar–ganglion cell interface are integrally involved in the organization of the retinal neuronal response to light. Single cell recordings have shown that the hyperpolarization response at the receptor level is a graded response (see previous discussion of visual acuity), not an ON/OFF response. Similarly, the hyperpolarizing horizontal cell response is a graded response, but in the horizontal cell there is a longer latency and the potential for summation over a wide range. The bipolar cell response is also graded but with a centre/surround effect where the centre hyperpolarizes and the surround depolarizes. In the amacrine cell, transient spikes can be observed, especially if correlated with the ON/OFF response, but only in the ganglion cell is a sustained spike discharge observed with a true depolarization occurring for ON, OFF and ON/OFF responses, depending on which type of ganglion cell is studied. This all-or-none response is graded only in the sense that the frequency or rate of discharge that occurs in the ganglion cell varies with the degree of depolarization. Different agonists are involved in different cells for ON, OFF and ON/OFF response in this locally integrated circuit. In addition, responses vary depending on whether wide field or narrow field amacrine cells are active (see next section) (Fig. 5.11).

Retinal connections, circuitry and neurotransmitters

Retinal connections

What is the basis of receptive field organization in the retina? As described previously, detection and processing of light stimuli by the retina are founded on the retinal receptor/neuronal network comprising the photoreceptors (rods and cones) and the neurones (bipolar cells and ganglion cells). The information finally transmitted to the lateral geniculate nucleus in the brain by the ganglion cells (ON, OFF and ON/OFF cells) is received directly from the bipolar cells but is modulated by horizontal cells and amacrine cells in the plexiform layers in the retina. This is canonically described in a simple arrangement of direct bipolar cell activity and lateral inhibition by horizontal cells and amacrine cells (see Chapter 1, p. 45 for details). In reality, retinal microcircuitry is more complex than this, underpinned by the fact that each retinal cell comes in several different varieties (Fig. 5.12). For instance, only one type of bipolar cell connects with rod photoreceptors but each of the 12 different types of cone bipolar cells connects with each cone photoreceptor in an integrated manner. In addition, the general arrangements found in most mammalian retinas are complicated by the presence of the fovea in primates, which is characterized by a single type of ganglion cell and where there is a one-on-one signal transmission from cone to ganglion cell. There are many types of ganglion cell serving different functions including two types of melanopsin sensitive ganglion cells (Fig. 5.13). In the periphery there is convergence of several cones onto one ganglion cell and this is why peripheral vision is less effective for fine detail, an effect previously attributed to rod signalling (Fig. 5.14). This lack of convergence at the fovea is also accompanied by a doubling of the 'time to peak' phototransduction at foveal cones compared with peripheral cones, for reasons that are not clear. A series of excellent reviews of this field has been published by Masland (2011, 2012a,b,c).

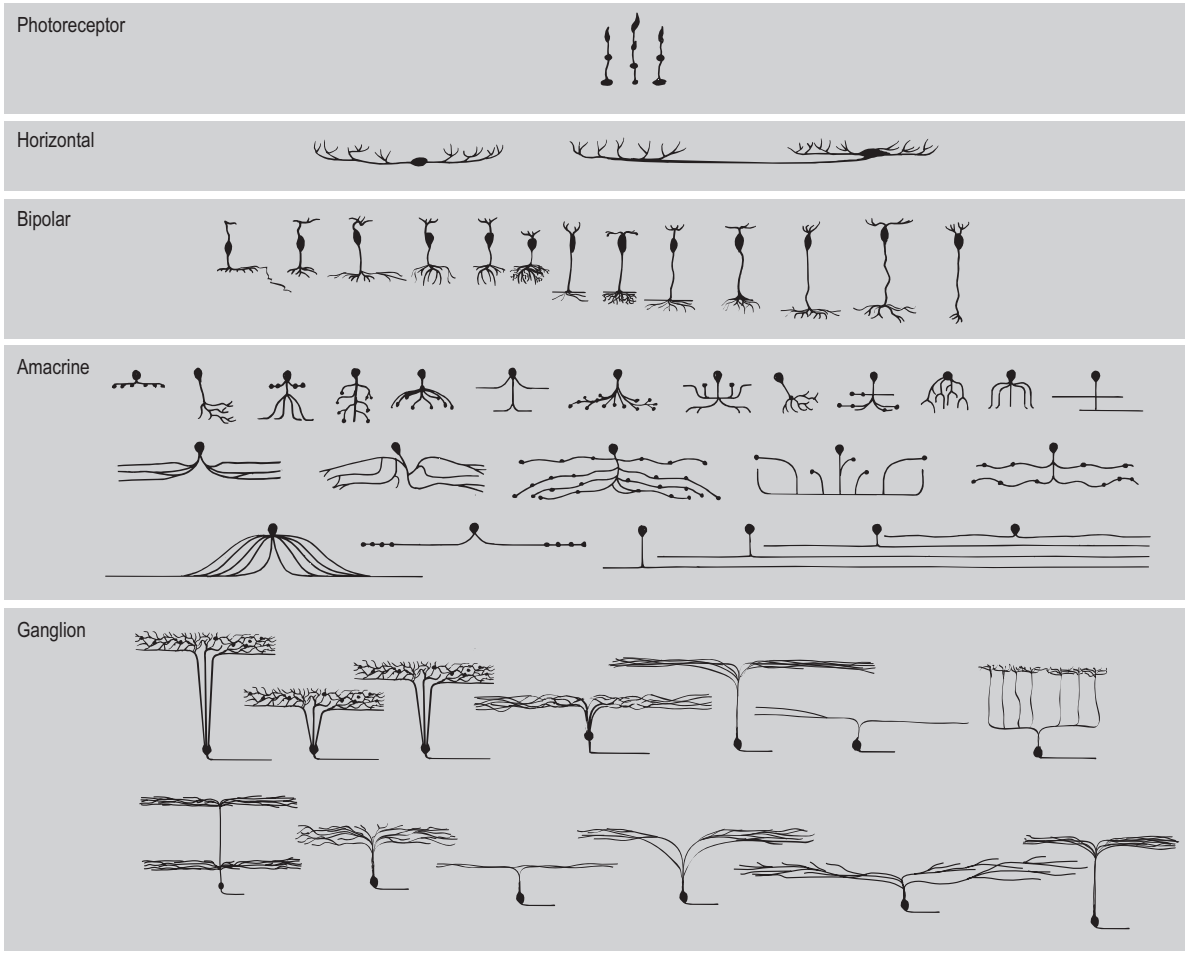


Fig. 5.12 The major cell types of a typical mammalian retina. (From Masland, 2012a.)

Further detail on the range and variety of retinal neural cells is available at: https://expertconsult.inkling.com

There are several retinal microcircuits emanating from the photoreceptors and acting in parallel. As indicated above, each bipolar cell connects all the cones within its branching territory and because of overlap, each cone connects with each of the 12 types of bipolar cells, each of which is transmitting different types of information to the next layer of cells (amacrine, horizontal and ganglion cells) (Fig. 5.15).

Thus, cones from several different types of microcircuits connect with a range of bipolar cells especially in the periphery and rod microcircuitry is minimal. In evolutionary terms, cones developed first, even though there are 20 times more rods than cones in the retina. Cone-bipolar cells synapse in the inner plexiform layer in a highly ordered set of stacks, each with a different number of connections and with its unique set of ionotropic, metabotropic (mGlur6), glycinergic and GABAergic receptors and calcium-binding proteins reflecting their inhibitory or excitatory output (translated into ON or OFF responses in the ganglion cell (see Box 5.7, Figs 5.11, 5.12 and 5.13, Video 5.1, and additional content concerning Masland available online at https://expertconsult.inkling.com).

There are several different types of ON/OFF bipolar cells, in part determined by the duration of the response (transient versus sustained); in addition, a single bipolar cell for these types of responses (non-chromatic ON/OFF) takes information from more than one cone (Fig. 5.15).

The organization of colour detection is somewhat different. A basic centre surround organization pertains here also. However, wavelength discrimination (see later) requires output from at least two cones to

TYPES OF RETINAL NEURAL CELLS

Masland's work has shown that the more than 60 different cell types in the retina are functionally organized into three sets that (1) sift the information from photoreceptors into roughly 12 separate channels or streams running alongside each other simultaneously; (2) transfer this information to specific types of ganglion cell; and (3) integrate information from bipolar cells and amacrine cells with ganglion cell output into approximately 20 different coded messages concerning the nature of the visual image which are transmitted to the brain via the ganglion cells (Masland, 2011, 2012a,b,c). The very large numbers of amacrine cells are especially interesting and appear to subserve many functions, many of which remain to be discovered but some of which are inferred simply from their positioning and architecture within the retina. For example, most recently, an 'interneuron' that in essence sends a blue OFF signal to the brain has been discovered, further underlying the complexity of neuronal cell integration within the retina.

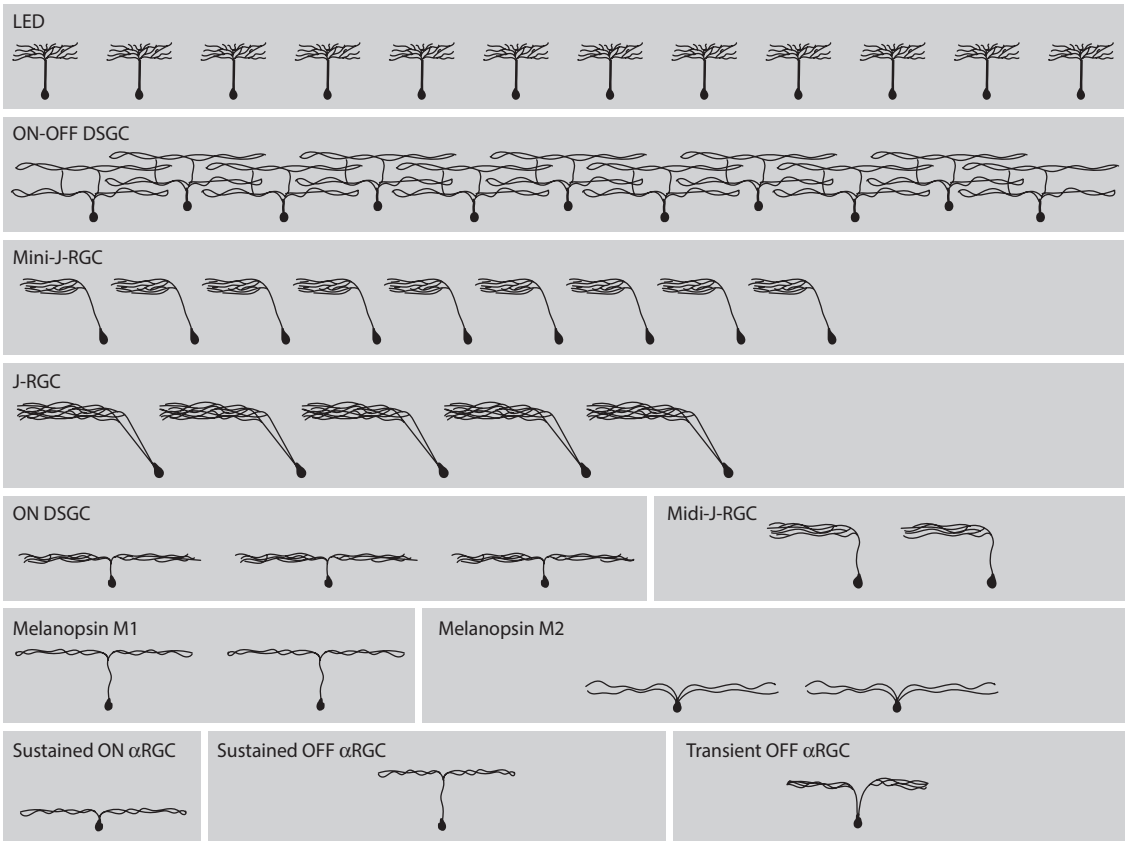


Fig. 5.13 A selection of mouse RGC types shown in cross section. The schematic images of the cells preserve their relative dimensions and stratification. The number of cells of each type is proportional to their approximate frequency in the mouse retina. *DSGC*, Directionally selective ganglion cell; *J-RGC*, junctional adhesion molecule B–positive RGC; *LED*, local edge detector; *RGC*, retinal ganglion cell. (From Sanes and Masland, 2015.)

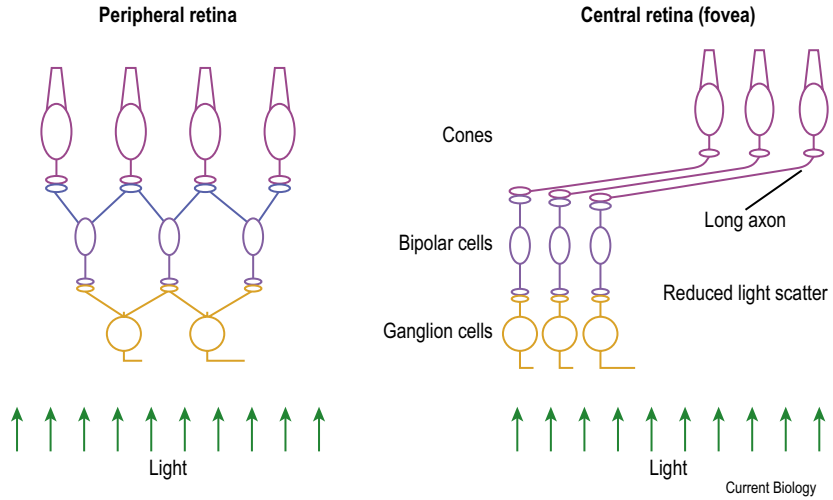


Fig. 5.14 Cones fire at different rates dependent upon their location in the retina: arrangement of neurons in the central retina and in the peripheral retina. There are two main specializations in the central retina. First, the central retina contains a specialized region, the fovea, in which each cone photoreceptor has a private line to the rest of the retina: when a cone synapses on a bipolar cell, that cell then synapses on one retinal ganglion cell. In the peripheral retina there is progressive convergence of several cones onto a bipolar cell. Second, in the fovea a long axon separates the synapse of a cone from the light -sensitive portion of the cone. Light can thus reach a foveal cone without having to pass through the neurons of the inner retina. Both of these specializations serve to increase the spatial resolution of the central retina. (From Masland, 2017.)

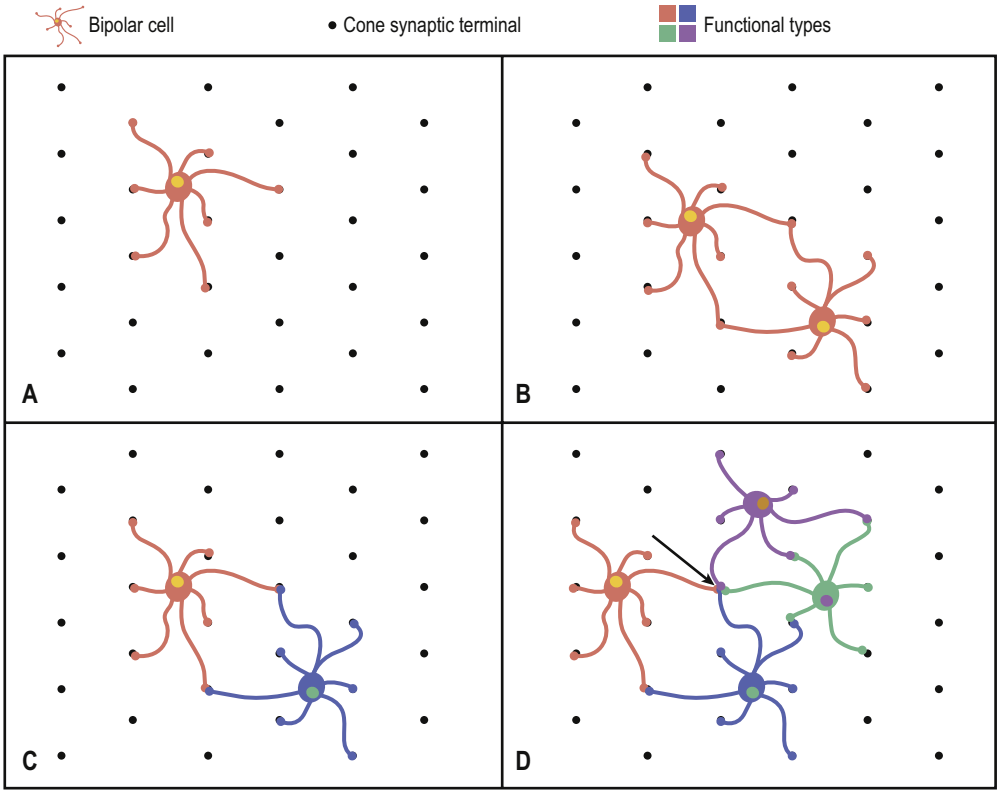


Fig. 5.15 (**A**–**D**) Each bipolar cell (represented by the different colours) connects all the cones within the area of its dendritic tree. The retina is organized such that each bipolar cell makes contact with all of the cones within the spread of its dendritic extensions, with the exception of the blue cone bipolar, which only connects with blue cones. They are numerically small and the principle remains that each of the different types of bipolar cell that connects with a single cone sends a unique message from that cone to the ganglion (or other) cell in the inner retina. (From Masland, 2012b.)

have something to compare against. 'Blue' (short-wavelength) cones make synapses with a single specialized type of cone-bipolar, while the remainder of the cones (long-wavelength red–green cones, comprising around 85%) connect with the several different types of bipolar cells described previously (Figs 5.12 and 5.15). The blue bipolar cell registers ON stimuli but blue OFF responses were generated by amacrine cells (see later), which convert a blue bipolar ON signal to an OFF signal through its role as an inhibitory interneuron (Fig. 5.16).

This particular amacrine cell is selectively sensitive to short-wavelength light and uses glycine as its inhibitory neurotransmitter. At the ganglion cell level, the blue ON ganglion cells receive input from the blue ON bipolar cell and the red–green OFF bipolar cell and the blue OFF ganglion cell may also receive input from the same amacrine inhibitory cells by sending its dendrites to that stratified layer of the retina.

Colour discrimination is therefore made by comparison of light detection by a short-wavelength cone (ON/OFF) with that from a long-wavelength cone (either red or green, ON/OFF), i.e. it is essentially a dichromatic system. In most mammalian retinas, there is no red/green discrimination (i.e. there is true dichromatism with comparison of only one long-wavelength cone with a short-wavelength cone), while in humans there is also red/green discrimination. However, in 5% of humans this is also true dichromatism (red–green colour blindness). These concepts are dealt with in more detail later in the chapter (p. 308).

So where does rod microcircuitry fit into this neuronal organization? Rods detect dim light, while cones detect bright light. Detection of dim light by rods is not to be confused with the OFF response of cone bipolar OFF cells (i.e. the response to the absence of light) but is a positive response by rod cells to very low levels of

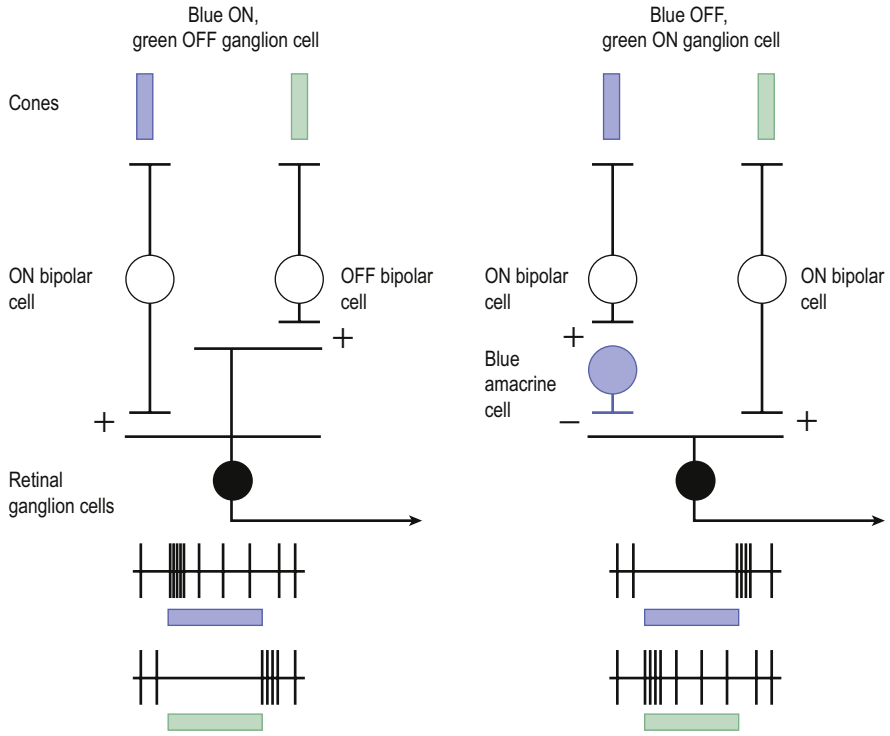


Fig. 5.16 The physiology of the ON:OFF ganglion cell as shown for the blue ganglion cells. Blue bipolar cells contacting blue-sensitive cones send a signal in response to the brightness of the stimulus (ON stimulus), which synapse directly to a particular ganglion cell *(diagram on left of figure)*. In contrast, if the bipolar cell receives a signal from the amacrine cell, this inhibitory signal generates a response from an OFF bipolar cell. In both cases, the signal is complemented by input from an inversely corresponding green cone. The stream of action potentials as the final output to the brain is shown in the *lower section* of the figure. (From Masland, 2012c.)

light. The majority of rods connect indirectly with ganglion cells through cone bipolar cells; large numbers of rods synapse with a single rod bipolar cell, which then connects via gap junctions with a cone bipolar cell through a further particular type of amacrine cell, the AII cell.

In primates, there is further specialization at the ganglion cell level related to the fovea and derived from the system of midget cells: midget ganglion cells connect directly with one cone bipolar cell and through this cell with one cone photoreceptor. Thus, there are a huge number of small bipolar cells and midget ganglion cells at the fovea, limited only by the packing of cones in this rod-free area. For instance, midget cells comprise about 70% of the ganglion cells in the monkey fovea. Moreover, in the primate fovea each cone cell connects with an OFF bipolar and an ON bipolar, leading to a greater density of bipolar cells at the fovea than cone photoreceptors. Each midget ganglion cell has a simple centre/surround organization underpinning the excellent spatial resolution of 'supervision' (see p. 314) ; in addition, it is believed that the midget cell system underpins the dual circuits required for red/green differentiation added to the existing blue cone system (see Box 5.8 and Figs 5.12 and 5.15). At the fovea, the colour opponency theory may not be as strict as suggested above but is modified by input from horizontal cells: thus short-, middle- and long-wavelength cones may compete against all the cones that contribute to the surround area after modulation by the horizontal cells.

Horizontal cells (Fig. 5.12) in the outer plexiform layer thus provide important feedback for both rod and cone outputs. For cones, this is classically believed to be in the form of contrast and mostly via generating inhibitory information to ganglion cells surrounding the activated cell (similar to the inhibitory effect of an action potential in a bundle of peripheral neurones on the neighbouring unstimulated neurones). This produces the centre/surround organization of ganglion cells (see Box 5.8). Horizontal cells may mediate their effects by subtracting the information on the average illumination of the entire retina from that stimulating a restricted

H^+ negative feedback onto cone terminals induced by HC polarization

Surround light	HC		V–ATPase at HC		Synaptic cleft [H^+]		Cone I_{Ca}		Transmitter release		OFF–BC	ON–BC
OFF	depolarize	→	↑	→	↑	→	↓	→	↓	→	hyperpolarize	depolarize
ON	hyperpolarize	→	↓	→	↓	→	↑	→	↑	→	depolarize	hyperpolarize

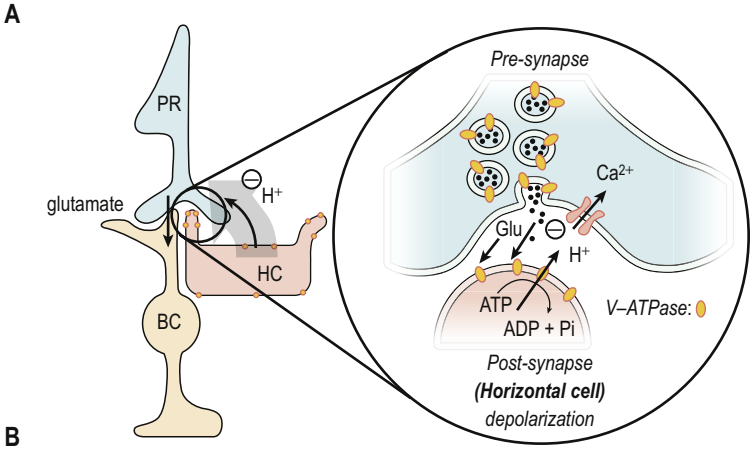


Fig. 5.17 Release of H^+ ions exert negative feedback onto the bipolar cells: this is achieved by changing the H^+ concentration at the cone photoreceptor:bipolar cell *(PR:BC)* synapse and is induced by horizontal cell *(HC)* polarization of BCs via change in H^+ concentration change at synaptic clefts of cone terminals induced by HC polarization. (**A**) The direction of change in the various parameters exerted by HC depolarization and hyperpolarization. (**B**) A schematic drawing of H^+ negative feedback onto BCs via release of H^+ from depolarized HCs illustrating the OFF case in (**A**). H^+ release occurs via V–ATPase (a proton pump, *turquoise circles* in the *inset* figure) resulting in suppression of glutamate *(Glu)* release from cones. V–ATPase is in synaptic vesicles at the cone terminals. (From Hirasawa et al., 2012.)

set of cones, thus underpinning the mechanism of light adaptation. Rods are separately modified by horizontal cells because of the anatomical location which separates the contact point with the rod far from the rod's contact with the cone bipolar cell.

One biochemical mechanism whereby horizontal cells direct the nature of the ON or OFF response appears to be mediated by acidification at the photoreceptor: bipolar cell synapse, mediated by a membrane-bound proton pump, and suppressing the effects of the glutamate neurotransmitter (Fig. 5.17). Modulation of cone responses is mediated by two types of horizontal cells, one acting locally on individual cones and the second acting globally.

Further detail on role of horizontal cells in cone / bipolar cell signalling is available at: https://expertconsult.inkling.com

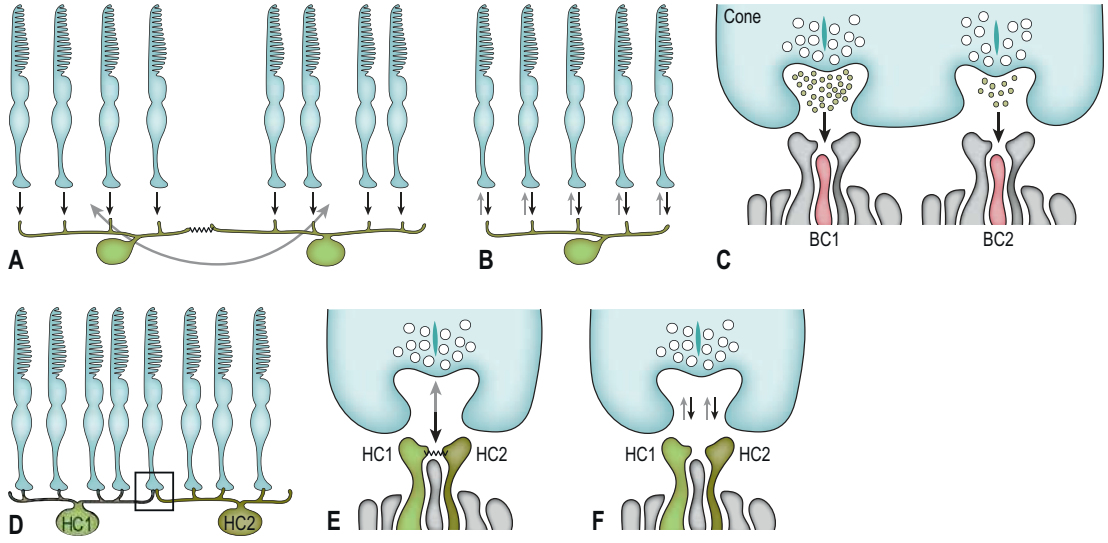
In contrast, amacrine cells are much more numerous than horizontal cells (Figs. 5.12) and have a wide range of functions. There are around 30 types of amacrine cells, some modifying ganglion cells respond over a wide area while others are much more restricted. In addition, they inhibit, enhance, entrain and refine through the large range of neurotransmitters and receptors discussed previously (e.g. dopamine, acetylcholine, glutamate, GABA, glycine, etc.). They thus affect many functions such as centre/surround organization, orientation selectivity, light–dark effects and colour discrimination. Some amacrine cells may have very small arborizations and function entirely within the receptive fields of a wide-field ganglion cell. For instance, starburst amacrine cells (SAC) have been identified as being ON–OFF direction sensitive, a feature that may involve only a section of the branching arborization in signalling. However, other amacrine cells have important roles in regulating direction sensitive ganglion cells (DSGC) either directly on the DSGC or indirectly by inhibiting the SAC (Figs 5.13, 5.15 and 5.18). In addition, other amacrine cells are sensitive to saccade-induced suppression, which allows differentiation of true object motion from motion of the entire visual field induced by fine eye movements. Amacrine cells have extensive paracrine effects via multiple neurotransmitter secretory functions, which include release of nitric oxide (see Chapter 7, Box 7.5, p. 397).

There is a growing awareness of the extent of neural processing that occurs at the retinal level particularly within the inner plexiform/ganglion cell layers (e.g. in contrast gain control) before it reaches the visual cortex.

Finally, there are about 20 different types of ganglion cells (Fig. 5.12). Originally described as X (slow, tonic) and Y (fast, transient) cells in the cat, and P (parvocellular, midget) and M (magnocellular, parasol) cells in the monkey, several others are now known to be responsible for centre/surround organization and direction/orientation selectivity. Each of the 20 different ganglion cells 'tile' the retina in its dendritic field so that the specific function of that ganglion cell is represented at that point in the retina (Fig. 5.19). In addition to morphological identification of ganglion cells, functional classifications based on transcriptomics and connectivity are developing, not only for GCs but for all retinal cells. The majority of ganglion cells (90%) are involved in visumotor guidance tasks; a small number of OPN4 (melanopsin) and

INHIBITION OF CONE/BIPOLAR CELL SIGNALLING

Horizontal cells (HC) are of two types. eFig. 5.5 shows a schematic representation of global feedback in an individual HC or in the electrically coupled HC network. Glutamate released by cones activates global HC whose inhibitory signals are spread widely via gap junction. In contrast, locally acting HC feedback onto cone-bipolar synapses via discrete single dendritic processes. It is suggested that the rate of glutamate release by the cone also has a regulatory effect on signal outcomes.



eFig. 5.5 Cross talk in the retina: Global and local signal processing at the cone-to-horizontal cell synapse. (**A**) Cones *(purple)* release glutamate *(black arrows)*, which is received by horizontal cells (HCs) (*red*) that in turn provide global feedback to cones *(grey arrows)* due to spread of signals in the HC network via gap-junctions *(black zigzag)*. (**B**) Cone synapse-specific; HC feedback. HC dendritic processes may feed back onto individual cones independently from neighbouring dendritic processes *(grey/black double arrows)*. (**C**) Hypothesized ribbon-specific presynaptic microdomains at different synaptic clefts in a single cone axon terminal. Two ribbons contacting dendrites of two different types of ON cone BC *(dark and light green)* differ in glutamate release rate *(black circles)* and synaptic strength *(arrows)*. (**D**) Overview of the connectivity between cones and two neighbouring HCs *(orange and red)*. The *black box* corresponds to the enlarged schemata shown in **E** and **F**. (**E** and **F**) Two alternative hypotheses for the processing of cone signals in HC distal dendrites (postsynaptic microdomains). (**E**) Two electrically coupled processes from two distinct HCs *(orange and red)* act as a single input–output structure *(double arrow)* because of their gap-junctional communication *(black symbol)*. (**F**) The two lateral HC processes may be independent of each other and act as individual 'post-synaptic microdomains' *(grey and black arrows)*. Note that only two ribbon synapses are shown for the mouse cone axon terminal in **C**, **E** and **F** instead of ~10 ribbon synapses per cone. *BC*, Bipolar cell. (From Chapot, 2017.)

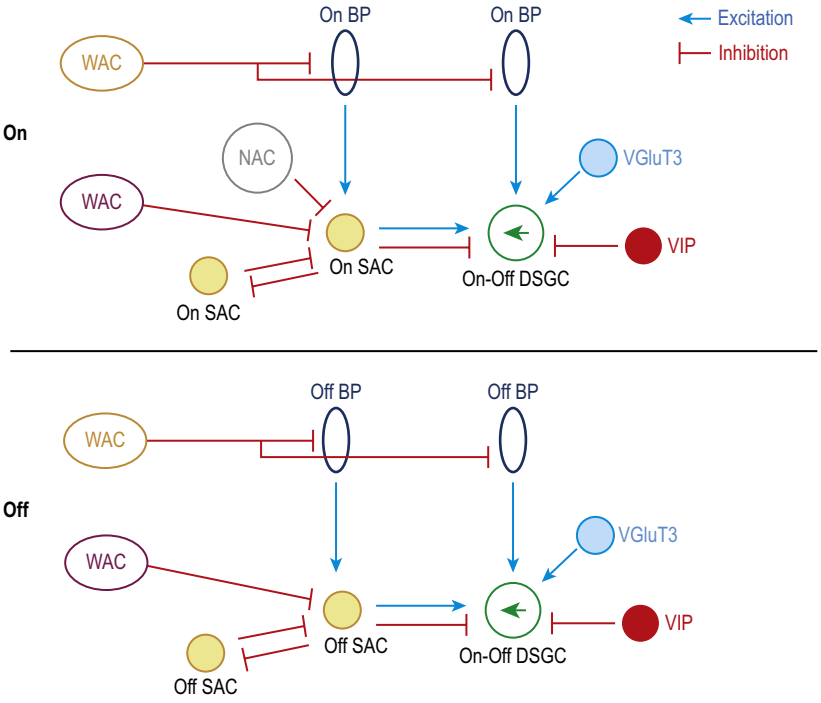


Fig. 5.18 Schematic of anatomically and functionally identified connections in the On and Off pathways of the direction-selective circuit. *BP*, Bipolar cell; *NAC*, narrow-field amacrine cells; *SAC*, starburst amacrine cell; *VGluT3*, vesicular glutamate transporter type 3; *VIP*, vasoactive intestinal polypeptide expressing amacrine cell; *WAC*, wide-field amacrine cells. (From Wei, 2018.)

OPN5 (neuropsin) sensitive ipRGCs (see Chapter 4, p. 256) connect with neurones in the pretectal and suprachiasmatic nuclei and responds to the level of 'irradiance' and controls pupil light responses as well as sleep/wake cycles (see Box 5.5). There are several types of ipRGCs (see eFig. 5.2) that regulate the production of melatonin by the pineal gland and are thus involved in entrainment of circadian rhythms (see p. 282). A number of other ganglion cells such as ON-OFF DSGCs, alpha RGCs, ON DSGCs, and more, deal with other visual functions such as photoentrainment and image stabilization (see Figs 5.12, 5.13, 5.18 and 5.19).

From all of these points, it is clear that there is considerable diversity in the retinal microcircuitry in which significant information processing occurs before its transmission to the higher centres in the brain.

Clinical visual electrophysiology

The electrical potentials that exist in the eye, by virtue of factors such as the dark currents in the retina and the transcorneal epithelial potential difference, can be altered by stimulation with bright flashes of light, producing mass responses of the tissue. These responses represent the resultants of many cellular potentials and the precise source of the discharges can only be inferred. However, extensive studies have located the source of retinal electric responses at different levels in the retina (Fig. 5.20). There are established internationally agreed standards describing normal responses set by the International Society for Clinical Electrophysiology of Vision (ISCEV).

Resting potential and the electro-oculogram

Since the eye acts as a dipole, it possesses a measurable resting potential, which is generated at the interface between the retinal pigment epithelium (RPE) and the photoreceptors (the resting retinal potential) and is about 60 mV in height. At a molecular level, the electro-oculogram (EOG) is representative of the trans-retinal pigment epithelial potential differences generated by separation of ionic gradients across the RPE by genes such as the bestrophin gene (see Chapter 4, Best1 gene, p. 157), and maintained by tight junctions. Similar transepithelial potential differences occur across all non-leaking epithelial layers. As the eye becomes light adapted, there is a steady rise in this potential, which is recordable as a reversible potential on horizontal eye movement and is known as the light rise (Box 5.9). Ratios less than 1.5 are abnormal, while a ratio >2.0 is regarded as normal. This effect is the result of an extracellular flow of current caused by changes in the potassium concentration in the interphotoreceptor matrix. The EOG is measured as a ratio of the light peak (i.e. peak amplitude when light adapted) to dark trough (i.e. lowest amplitude measured in the dark) and a normative set of responses based on ISCEV

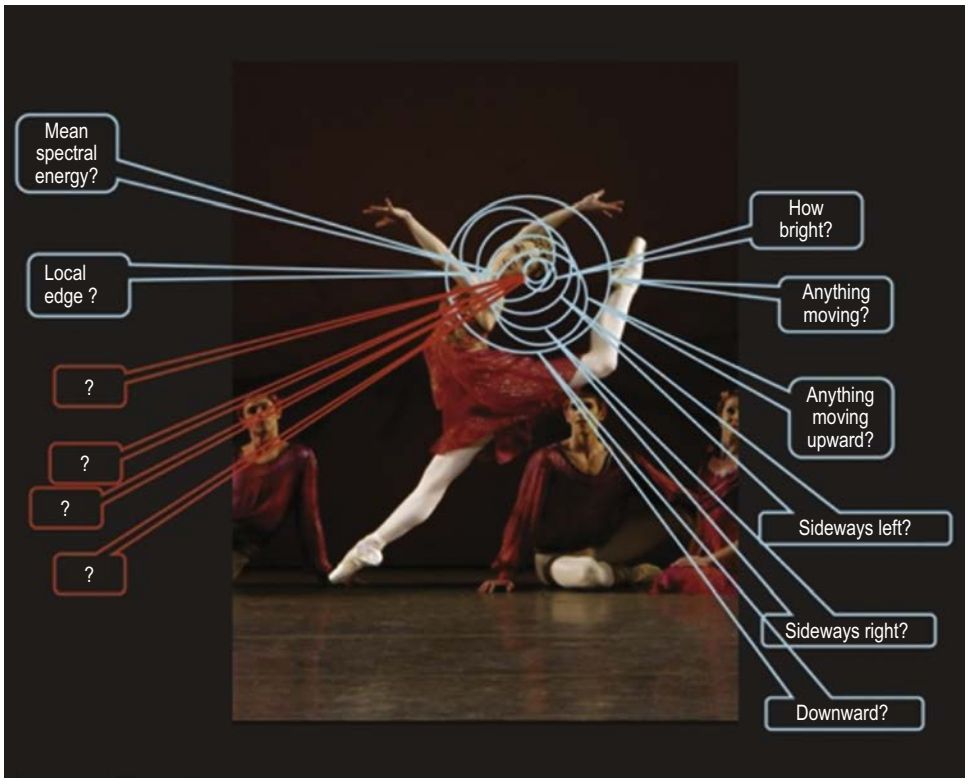


Fig. 5.19 The numerous components that contribute to an image are managed by the retina on a point-by-point basis. Twenty types of ganglion cells provide a unique inner retinal surface mosaic so that each is represented at any point in the retina corresponding to a point in space, which is the object being viewed. In the image, a small area on the dancer's head is assessed for various features, e.g. motion, direction, tilt, spectral composition (colour), tone, texture and more. (From Masland, 2011.)

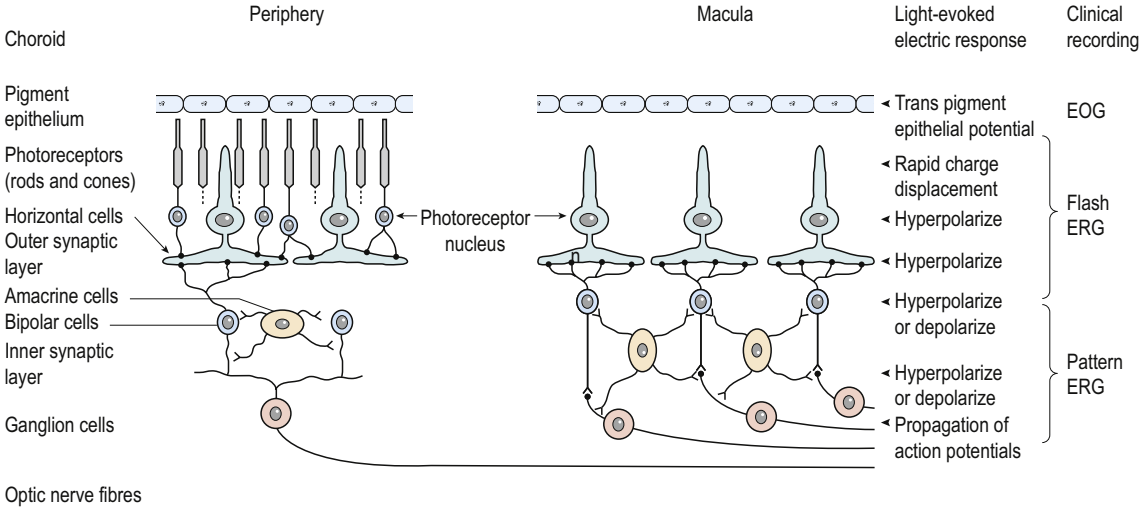
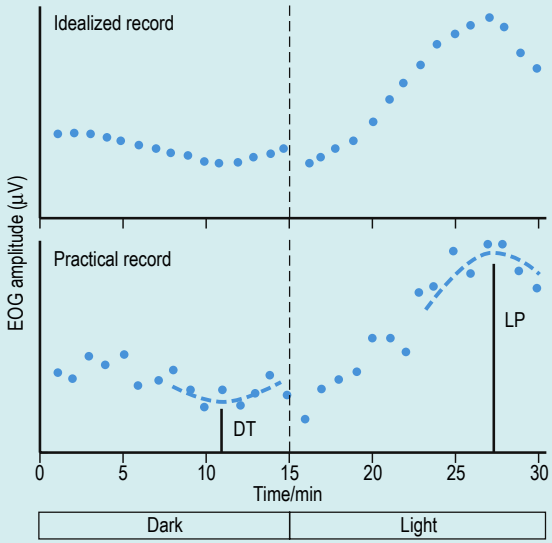


Fig. 5.20 Source of electrophysiological traces from retinal layers. The electro-oculogram lies at the RPE layer, while the ERG is retinal. The pattern ERG represents the response after processing in the bipolar cell layer. (Courtesy A.M. Halliday.)

Box 5.9 **The electro-oculogram**

The electro-oculogram (EOG) is a record of the electrical dipole occurring between the front and the back of the eye and reversing in current direction when the eye moves from side to side. The height of the potential difference increases in conditions of bright illumination.



An EOG recording: *vertical lines* represent the alternating dipole as eyes are moved from left (L) to right (R). The 'light rise' is seen as an increase in the height of the vertical lines as the light is switched on. The electro-oculogram (EOG) is measured in microvolts as a 'light rise', i.e. an increase in amplitude as the light stimulus is increased. The standard curve is shown in the upper two boxes on the left at low light levels and on the right as the light is increased. Actual EOG values are shown in the bottom two boxes. Idealized *(upper)* and practical (*lower)* EOG amplitude versus time curves for a healthy subject. The actual EOG is estimated from the points measured as in the *dashed line*. *DT*, Dark trough; *LP*, light peak.

methodology have been developed by meta-analysis of previous data (Constable et al., 2017). The electro-oculogram (EOG) is lost in conditions that disrupt the RPE–photoreceptor relationship, such as retinal detachment. Most patients with Best's disease or bestrinopathies have impaired EOGs but occasional normal EOGs have been reported.

The electroretinogram

The ERG is superimposed on the EOG and is the cumulative electrical response to a light stimulus from all the retinal elements. It is normally performed by presenting a very rapid light stimulus (the flash ERG) and the assumption is made that the stimulus is received over an infinitesimally small time. However, it is affected not only by the intensity and duration of the stimulus but also by the stimulus wavelength and pattern and the level of light–dark adaptation of the retina.

The ERG has several components (Fig. 5.21). Under mesopic conditions, the early receptor potential (ERP) is barely detectable and becomes apparent only with an extremely high-intensity flash stimulus in a deeply dark-adapted eye. It originates from the photochemical reactions in the rod outer segments on stimulation by light and is dependent on the density of rods and high levels of unbleached rhodopsin. The ERP may therefore be detectable in eyes where the inner retina has been destroyed but the outer retina is substantially intact, e.g. central retinal artery occlusion. However, it is not normally recorded because of the above physiological constraints.

The negative a-wave is generated by hyperpolarization in the photoreceptors' inner segments (Granit's PIII component), the a1 component coming from the cones and the a2 from the rods. In contrast, the b-wave (Granit's PII component) is believed to come from the bipolar cells either directly or indirectly via signal spread to the Müller cells; b1 is generated by cone-dominated and b2 by rod-dominated bipolar cells. The b wave is lost in certain retinal vascular conditions, such as central retinal vein occlusion.

Oscillatory potentials are thought to be generated by amacrine cells, while the slow rising c-wave (Granit's PI) depends

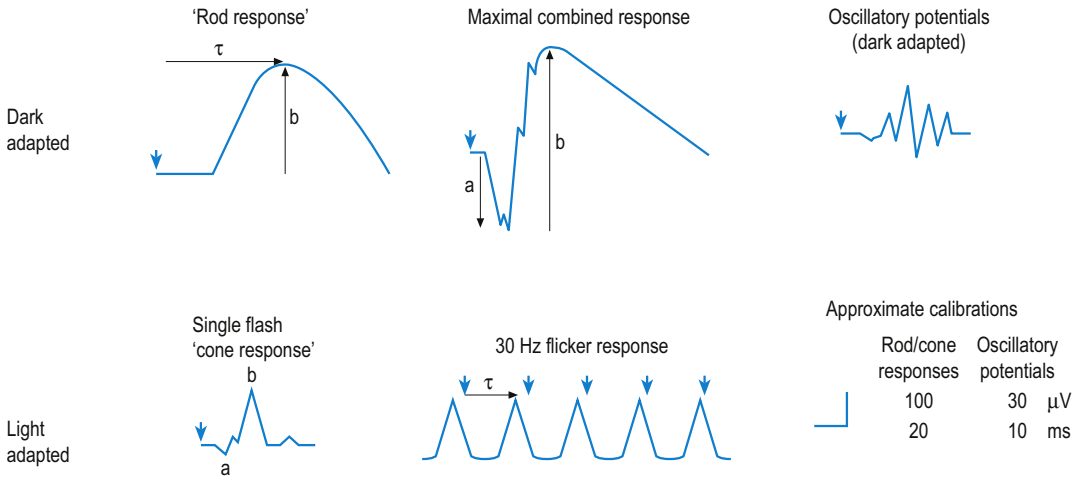


Fig. 5.21 Diagram of the five basic ERG responses defined under standard conditions. These waveforms are exemplary only and are not intended to indicate minimum, maximum or even average values. *Large arrowheads* indicate the stimulus flash. Dotted arrows exemplify how to measure time to peak (τ, implicit time), a-wave amplitude and b-wave amplitude.

on an intact pigment epithelium. However, the EOG provides a more effective estimate of the integrity of the pigment epithelium. Oscillatory potentials are lost in patients with diabetes.

The ERG as described previously is in essence a response to luminous intense stimuli. However, the pattern ERG (PERG), which is the ERG response to a reversing checkerboard of black and white squares of equivalent luminance, is thought to represent the electric response to spatial contrast, probably from ON-centre ganglion cells. ISCEV have also developed a protocol for a photopic negative light-adapted ERG, which represents a slow negative-going wave after the b-wave and is thought to be a result of ganglion cell activity (Fig. 5.22).

The multifocal ERG

The multifocal ERG records discrete electrophysiological responses from specified regions of the retina. An array of hexagonal elements in a changing frame is used to stimulate the retina rather than a full field flash of light and each element has a 50% chance of stimulating the defined retinal area (Fig. 5.23). Each stimulus is randomly illuminated over time, producing a continuous ERG response, which is correlated with the ON and OFF phases of each patch and so the final result is not truly an ERG response from a discrete retinal region; instead, the response is interpreted as a statistical/mathematical resolved factor and produces a typical waveform.

Further information on the normal multifocal ERG and accompanying data is available online at: https://expertconsult.inklingg.com

The multifocal ERG can be evaluated in conjunction with complementary function and imaging techniques that assess the visual field: for instance, microperimetry is a technique in which the visual sensitivity to discrete stimuli of a specific area of the retina, identified by scanning laser ophthalmoscopy, is correlated with optical coherence tomographic imaging of the corresponding retinal region.

How close is the flash ERG to normal retinal physiology?

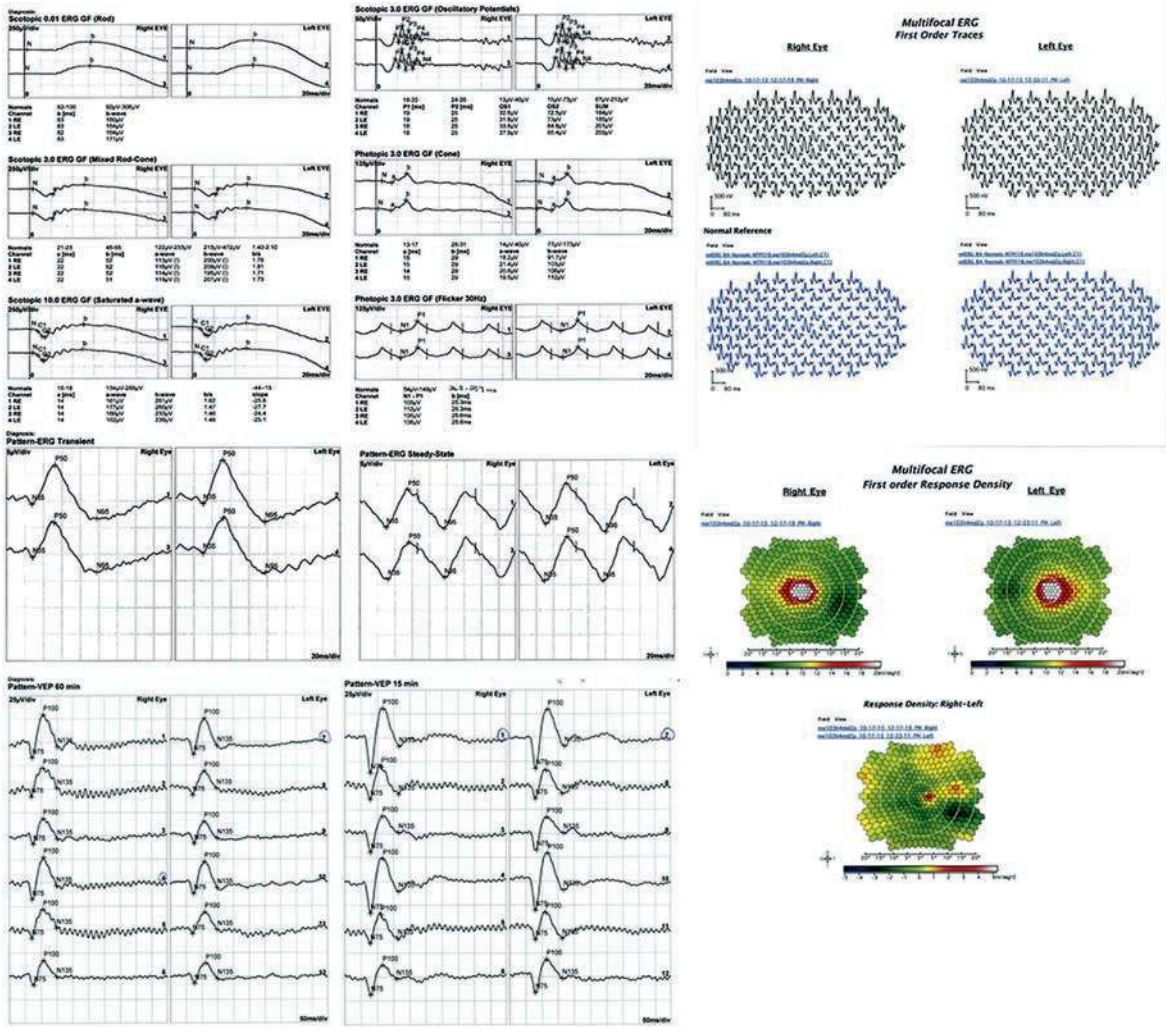
As previously described, the flash ERG is based on assumptions of an extremely short impulse time and is approximation of the impulse response function (IRF). In addition, errors in positioning the electrodes can introduce significant variability which affect reproducibility. There may also be significant background interference. An alternative approach involves the use of temporal white noise analysis to smooth out potential artefacts and is probably closer to physiological retinal responses (Fig. 5.24). It has also been shown to be possible to record selectively the rod response in a light adapted eye using 'silent substitution stimuli' which effectively block out all cone responses (Maguire et al., 2017). Such techniques will be of great value in studies of normal and abnormal rod function in humans.

The visual evoked potential

The visual evoked potential (VEP) records electric activity from the occipital cortex following presentation of a light stimulus to the retina and represents a limited electroencephalogram (EEG). Recordings are taken from a set of six electrodes placed

THE NORMAL MULTIFOCAL ERG IN PRACTICE

The multifocal ERG has greatly improved the detection of functional abnormalities in the retina and precise localization of such abnormalities to specific regions of the retina. In a standard examination protocol several data sets are taken, including: an ISCEV standard ERG (scotopic and photopic), the multifocal ERG, a pattern ERG (both transient and steady state) and the visual evoked potential (VEP) using a large check and a small check stimulus (see main text Figs 5.22–5.26 and eFig. 5.6). In addition, visual acuity and contrast sensitivity measurements are taken. Using this battery of tests, differentiation between generalized retinal dysfunction, disorders of the macula, optic nerve dysfunction and 'non-organic' disorders can be reliably made.



eFig. 5.6 Typical set of information obtained from electrophysiological evaluation of retinal and optic nerve function. (Courtesy Dr Fred Chen, Lions Eye, University of Western Australia.)

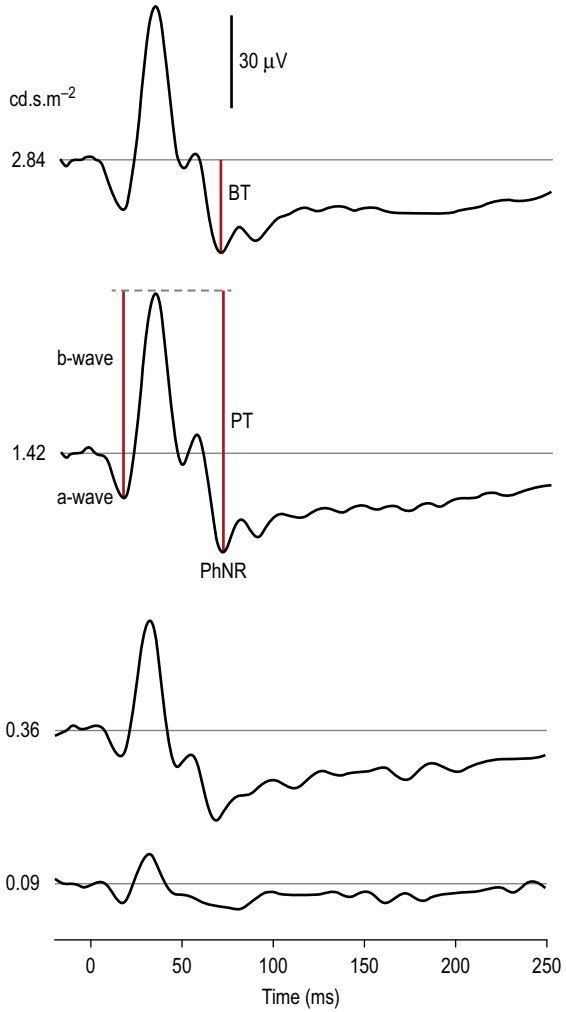


Fig. 5.22 Illustration of the light-adapted ERG of a healthy subject (35 years) in response to a brief red LED flash (660 nm) at each of four flash strengths, on a blue background (460 nm) of 10 cd m^{-2} using the protocol for detection of the photopic negative response (PhNR) amplitude measurements from baseline to PhNR trough (BT) and from b-wave peak to PhNR trough (PT). (From Frishman, 2018.)

around both left and right occipital cortices, each producing a discrete wave pattern of different amplitudes (Fig. 5.24).

Several types of VEP can be induced, including flash, flash-pattern, pattern-onset, pattern-offset and pattern-reversal. Considerable individual variability in the wave pattern is seen with the flash VEP, which is composed of two phases: the evoked potential and the after discharge.

Variations also occur in the amplitude of response, depending on the level of dark adaptation. A pattern-flash VEP is evoked when a black-and-white checkerboard stimulus is presented (Fig. 5.25). The amplitude of this response is considerably better correlated with the visual acuity. However, a significant electric interference in this response is caused by switching the stimulus on and off (the onset/offset response). The flash-evoked potential has three components: N1, P1; N2, P2; and N3, P3; and the pattern-evoked potential is essentially monophasic (see Figs 5.24 and 5.25). The flash VEP is considered to arise from area V2 of the cortex with its retinal origins arising in the entire papillomacular bundle, while the pattern VEP is considered to arise in V1 plus activity in ganglion cell receptive fields corresponding to large checks (M cells) and small checks (P cells).

Presentation of a pattern-reversed black-and-white equiluminant checkerboard can overcome the onset–offset problem if the pattern is reversed at an appropriately short interval, as the effects tend to cancel each other out. Indeed, the stimulus can be set to produce pattern-onset, pattern-offset and pattern-reversal VEPs, each of which has a characteristic set of wave patterns depending on the conditions (Fig. 5.25). Considerable ingenuity has been developed in techniques for studying half-field stimulation, macular vision and the effects of age, right versus left eye, etc. However, clinically, this test has greatest applicability in assessing the function of the optic nerve by measuring the latency of the response and in assessing the integrity of foveal vision by evaluation of the pattern and amplitude of the wave forms. A multifocal (mf)VEP has also been introduced but is prone to variability and artefact from non-retinal or optic nerve defects. However, under strict conditions it can generate reliable data.

Flicker

A spoor or beam of light can flicker so fast that the sensation of flicker is lost. The point at which this happens is known as the critical fusion frequency (CFF) and the brightness of the steady-state light is the same as the mean brightness of the flickering light midway through its cycle. Brighter stimuli have a higher CFF and rods have less ability to achieve fusion than cones (see later).

The fact that fusion occurs at all indicates that there must be some degree of persistence of sensation after the stimulus has ceased, but clearly this is more effective at lower levels of illumination than at higher ones. Fusion is thought to be the result of an after-effect, so that a succeeding stimulus can fall on the retina while it is still responding to the first stimulus. However, it is more probably related to light adaptation in which the response to light is accelerated at higher intensities. When flicker is measured by an ERG response, the CFF is seen as a smoothing out of the electric response, which is set at a higher level in millivolts. Presumably it would be possible to induce a second ERG response by presenting a superimposed flash on this new level of illumination.

Subjectively, fusion occurs at 60 cycles/s, but the ERG CFF occurs at 25 cycles/s. Therefore, by setting the flicker

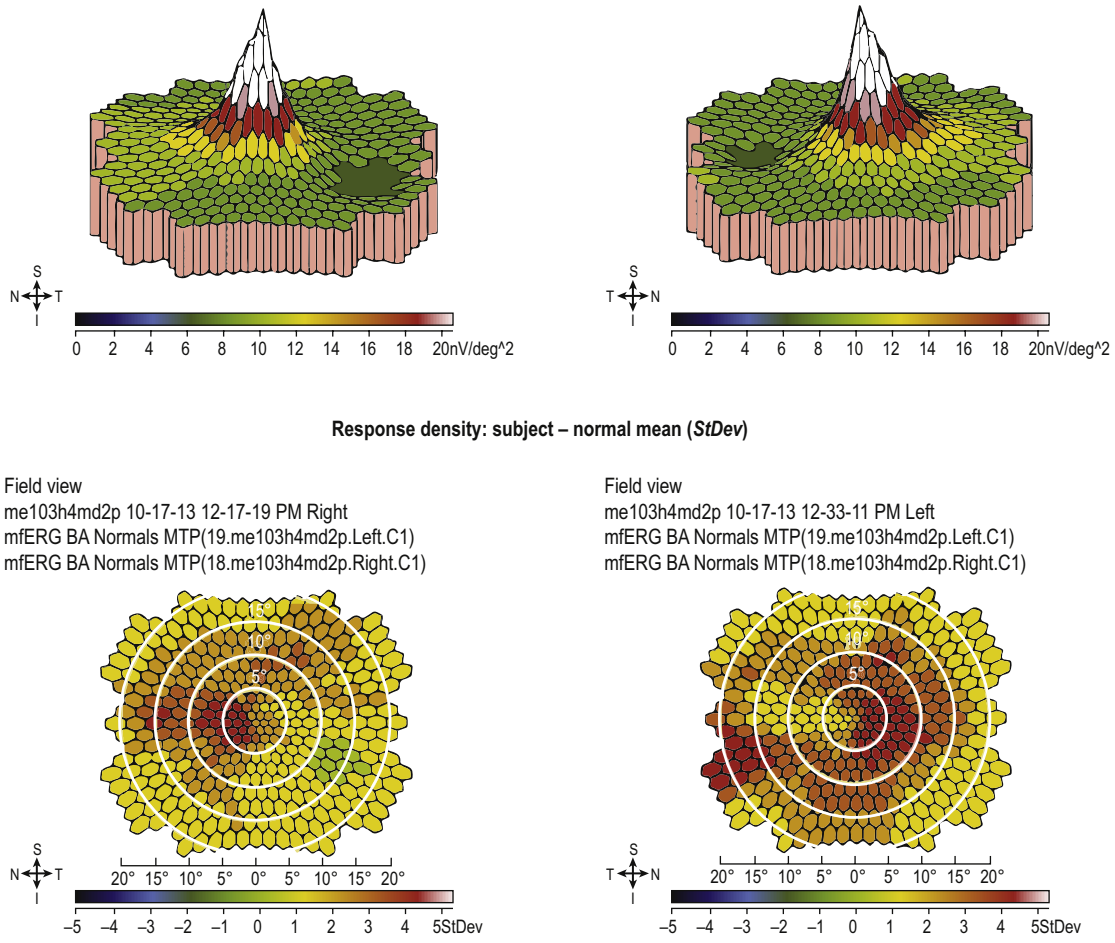


Fig. 5.23 Normal multifocal ERG record showing the characteristic 'hill of vision' and the hexagonal arrays representing each retinal region analysed. (Courtesy Dr Fred Chen, Lions Eye Institute, University of Western Australia.)

cycle at 25 cycles/s, through the flicker ERG, a measure of cone function in isolation can be obtained (see later). Interestingly, direct electrophysiological studies on optic nerve fibres have shown that ganglion cells with a high spike discharge rate also have a high threshold for CFF and vice versa. Interestingly, data obtained using the mfVEP have shown that the magnocellular (M) pathway correlated negatively with flicker threshold but not the parvocellular (P) pathway confirming that the mfVEP provides information on the peripheral vision rather than central vision.

A full update on electrophysiological protocols is available at https://www.ncbi.nlm.nih.gov/pubmed/29397523. In addition, there is an international storage facility for electrophysiological data available under the name ElVisWeb, a free online service.

Colour vision

Colorimetry and colour discrimination

Colorimetry is a measure of visual function at the photoreceptor level. In contrast, colour discrimination is a cortical function related to perception and the later stages of visual processing. Colorimetry or the measurement of colour is based on techniques of colour matching, which have a long history going back to the days of Newton, Helmholtz and particularly Maxwell in the mid-nineteenth century.

For further information on the work of Maxwell in colour vision, see additional content available online at: https://expertconsult.inkling.com

JAMES MAXWELL, SCOTTISH MATHEMATICIAN AND PHYSICIST

Described as 'the man who changed the world', James Clerk Maxwell is known as the most influential physicist of the nineteenth century, ranking with Isaac Newton and Albert Einstein. He developed one of the major unifying theories of physics, namely that of electromagnetism, demonstrating that light energy was one component of this unifying theory of electricity, magnetism and optics. Maxwell was born in Edinburgh and studied at Edinburgh University and after further training at Cambridge became professor of Natural Philosophy at the University of Aberdeen, before moving to University College London where he developed most of his seminal work. It was while in London that he completed much of his work on colour and produced the first 'light-fast' colour photograph. He also completed his work on electromagnetism.

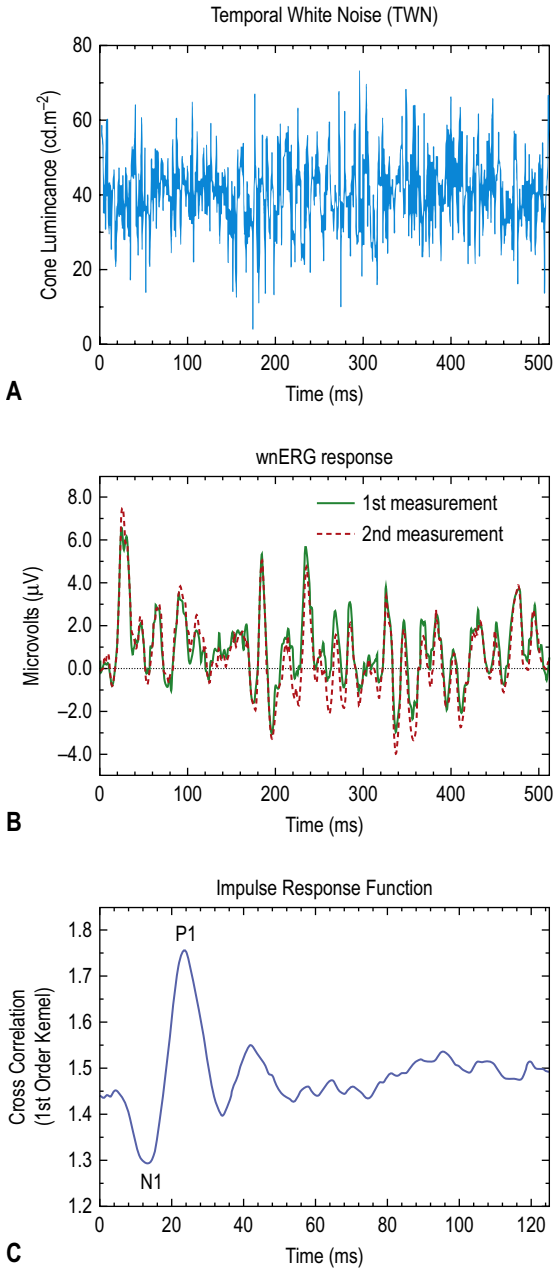


Fig. 5.24 Calculation of the linear Impulse Response Function (IRF) of the ERG *(panel C)* through the cross-correlation of the temporal white noise *(wn)* stimulus *(panel A)* and the measured wnERG *(panel B)*. (**A**) An exemplar TWN stimulus (0–512 Hz) that modulates above and below a mean adaptation level (40 cd.m^{-2}). (**B**) The measured full-field (Ganzfeld) wnERG responses (two measurements, *green solid* and *red dashed lines*) shown in panel A are the average of 40 sweeps each and are reproducible across repeated measurements. (**C**) The cross correlation of the TWN stimulus *(panel A)* and the resultant wnERG response *(panel B)* return the linear IRF. Definitions of the N1 and the P1 components are displayed. (From Zele et al., 2017.)

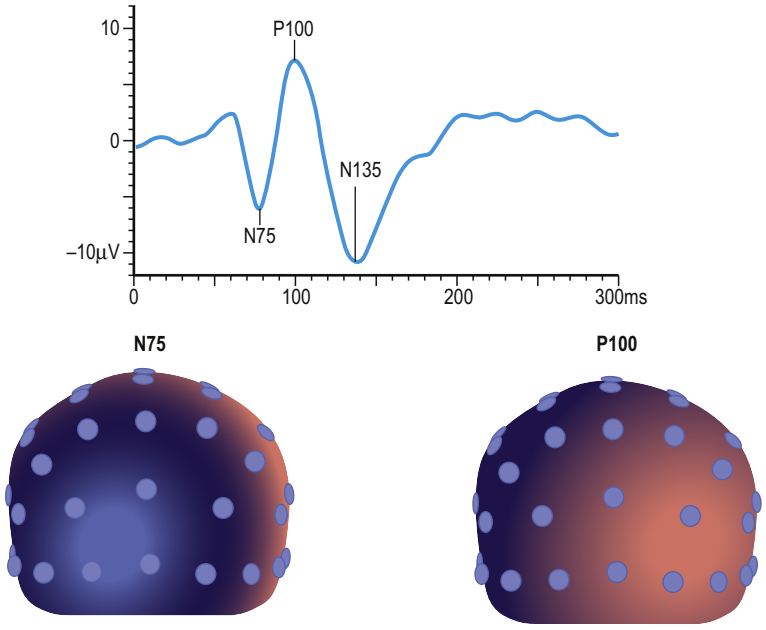


Fig. 5.25 Diagram demonstrating the visual evoked response showing the potential map for the location of the electrical output from the scalp electrodes. (From Kevin Whittingstall: http://fizz.phys.dal.ca/~medbiophys/kevinp.htm.)

Maxwell is credited with asserting that all vision is colour vision; in a sense this is true since rod photoreception is simply setting the threshold for detecting light while cone vision (colour vision) does the job of 'seeing'. Standards for measuring colour have been established for certain reference conditions of illumination based on the assumption that only a small area of the central fovea is illuminated by the test stimulus; a standard V (λ) or visible wavelength curve was adopted by the Commission Internationale de l'Eclairage (CIE) in 1964. A colour (C) is specified in terms of the three primary colours by the equation:

$$\mathbf{C} = \mathbf{r} \cdot (\mathbf{R}) + \mathbf{g} \cdot (\mathbf{G}) + \mathbf{b} \cdot (\mathbf{B})$$

and is measured with colour-matching instruments such as the flicker photometer or with spectrophotometers fitted, for instance, with arrays of wavelength-selective photodiode detectors.

Different colours have different luminosity

Spectral sensitivity curves

In the dark-adapted state, light of different wavelengths appears variably bright with a peak luminosity at about 500 nm in the blue–green region, i.e. for lights of equal energy, blue–green appears brightest in the dark (Fig. 5.26). Under photopic conditions, however, peak brightness occurs around 555 nm in the yellow–green. Brightness curves of different wavelengths like this are determined under photopic conditions by flicker photometry. This phenomenon, in which short wavelengths become brighter compared with long wavelengths as luminance is reduced, is known as the Purkinje shift, and begins under mesopic conditions when cone function is still active. Experimentally, the flicker ERG is useful in studying the Purkinje shift because rod and cone responses can be distinguished by setting the flicker rate above 25 cycles/s, which rod photoreceptors cannot detect (see earlier). Purkinje shifts are not readily detectable in guinea-pigs (mostly rod) or squirrels (mostly cone) but can be detected in cats, which are rod-dominated.

Although cone and rod responses can be differentiated using the flicker ERG, it is also possible to differentiate long (L) and middle (M) wavelength responses using appropriate flicker frequencies. In addition, using mixed luminance (rod) and chromatic (cone) flicker frequencies around 12 Hz and appropriate colour backgrounds for the stimuli, it has been possible to separate out the luminance and colour electrophysiological responses detectable at a retinal (ERG) level rather than at a post-signal processing (cortical/perceptual) level.

Photochromatic interval

The Purkinje shift underlies the photochromatic interval, which is a measure of the difference in 'brightness' between the absolute threshold at which light of any wavelength is just detected and the brightness at which it appears coloured. Clearly, this interval is vanishingly small at the red end of the visible spectrum and is maximal at about 570 nm.

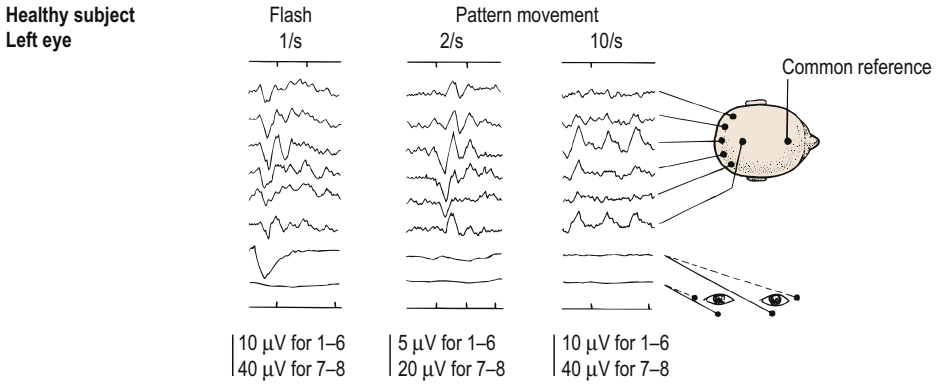


Fig. 5.26 Flash and pattern-averaged VEP responses from each of six electrodes placed over the occipital region of the skull. Pattern reversals were performed at two different rates *(middle and right panels)*. The two *lower* tracings represent simultaneous ERG recordings. (Courtesy Dr Ikeda.)

Cone thresholds

As we have seen above, thresholds are an artificial concept that depend on a critical number of 'hits' on photoreceptors by quanta of light. In practice, under defined conditions, light thresholds are measurable and have been well characterized for rods. Cone thresholds can also be measured, for instance by using only the early part of the dark adaptation curve or by using very small bright flash stimuli, which only impinge on the central fovea. By choosing suitable conditions of light adaptation (e.g. by adapting with blue light to desensitize the rods), it is possible to measure cone thresholds with different wavelengths of light. Using wavefront adaptive optics technology, which minimizes chromatic and other aberrations, the absolute threshold for cone vision has been determined at 203 ± 38 photons at the cornea and is considerably greater than previous estimates. Using these methods cone-specific spectral curves can be produced.

Colour detection requires more than one type of photoreceptor

Photoreceptors respond to stimuli by changes in the frequency of electric discharge. Indeed, this is true for all neuronal impulses. Since photoreceptors respond to both luminosity and wavelength, a retina that has a single type of photoreceptor (such as a rod) will not be able to distinguish one stimulus from the other under different circumstances. Wavelength discrimination therefore requires a panel of photoreceptors (minimum of two) with peak responsiveness at specific wavelengths independent of their responses to changing luminosity. In theory, the more variety in receptor type with specific spectral sensitivities, the greater the ability to discriminate wavelength, as for any single wavelength stimulus it is the pattern of discharges from the entire panel of receptors that determines this discriminatory ability. Probably through evolutionary constraints, two types of cone photoreceptor provide sufficient discriminatory ability for survival of the species and the red–green medium (M) to long (L) wavelength separation is an additional component restricted to primates (see section on retinal connections p. 311).

Trichromatic theory of colour vision

A specific colour or hue is therefore detected by the summation of responses from a mixture of receptors and the contribution from each of the three primary photoreceptor types can be deduced from spectral mixing curves (Box 5.10). Indeed, such colour mixing phenomena are the result of 'confusion', i.e. our inability to differentiate sufficiently narrow wavelengths. This is a reflection of the physiological limits on wavelength discrimination set by our having only three cone photoreceptors. (Certain species of fish have four cone photoreceptors.)

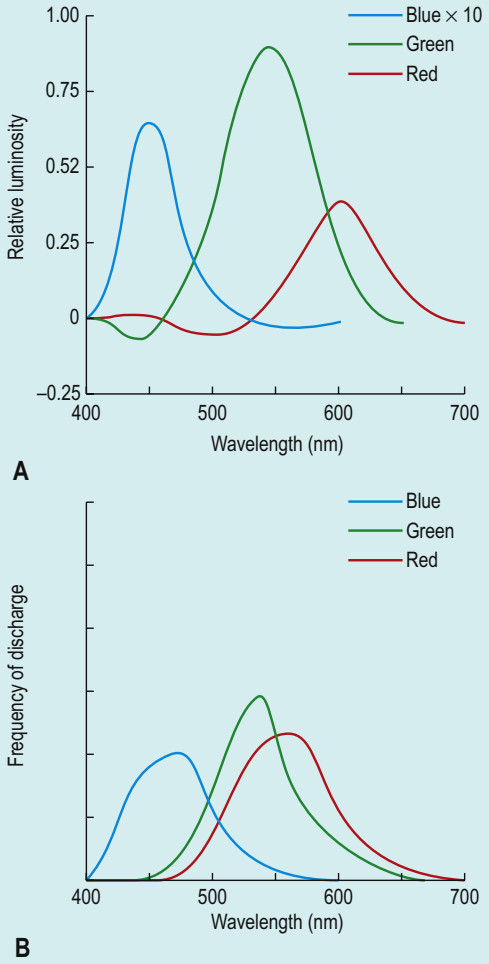
However, hue discrimination is more than a retinal-processing phenomenon. Experiments testing the ability of humans to detect the four unique hues (red, green, yellow and blue) using three chromatic mechanisms in P cells tuned to detect L/M and S-L/M show that hue discrimination requires higher-order colour perceptual mechanisms.

Psychophysical evidence for three cone photoreceptors

An intrinsic property of sensory receptors is to adapt and this is well demonstrated by rod (dark adaptation) and cones (light adaptation). Experiments using cone thresholds and light adaptation techniques have confirmed that there are three types of cone that respond differently to the same

Box 5.10 **Spectral mixing curves**

Mixing the three primary colours will produce any secondary colour or hue. However, a specific quantity of the primary colour is required to produce each hue and this amount is determined by spectral mixing curves.



(**A**) Graph showing relative amount of each primary colour required to match any specific wavelength. (**B**) Actual frequency of discharge in retinal neurones for each primary colour at any specific wavelength.

wavelength of light. These are reflected in the spectral sensitivity curves for the three photoreceptors, which form the basis of the chromaticity chart (see pp. 278 and 308). Light adaptation is accompanied by several molecular events: (1) changes in intracellular calcium level regulating guanylate

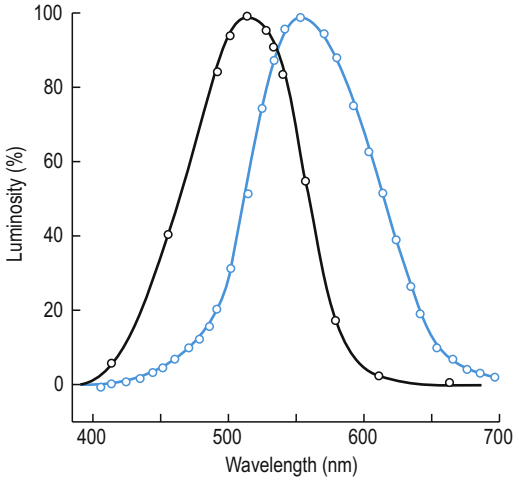


Fig. 5.27 The Purkinje shift showing differences in luminosity observed at different wavelengths.

cyclase activity in some species, duration of opsin activation and channel opening; (2) phosphodiesterase activity; and (3) a slower mechanism, possibly involving interaction between dopamine release and melatonin (see Chapter 4, p. 254). Studies of colour-blind individuals have also provided confirmatory evidence of the trichromatic theory.

Similar results can be obtained using a technique known as the liminal brightness increment (l.b.i.), in which the amount of additional light required to produce a detectable difference in the brightness of a target against a changing background luminance is determined. The technique can be highly discriminatory by measuring, for instance, the liminal brightness increment in a blue central target against a green background. The studies confirmed that there are indeed single receptors that peak in the red and green regions, but the blue spectral sensitivity curve is more complex and there are probably three components for the blue mechanism. Whether the influence of rod mechanisms on these tests can be completely eliminated is not clear. Techniques combining spectral reflectance densitometry and adaptive optics in healthy volunteers have demonstrated the unique arrangement of S, M and L photoreceptors. Interestingly, despite normal colour vision, the variation in density of L and M cones, in particular, varies considerably between individuals (Fig. 5.27). It has been suggested that this may represent a compromise between the competing requirements for spatial versus colour vision and this is reflected in the patchy distribution of the L/M cones at the fovea and their different distribution in the peripheral retina.

Molecular evidence for three receptors

Just as rhodopsin represents the molecular receptor for light energy at the level of the photon, there are cone pigments that are sensitive to photons of light generated within

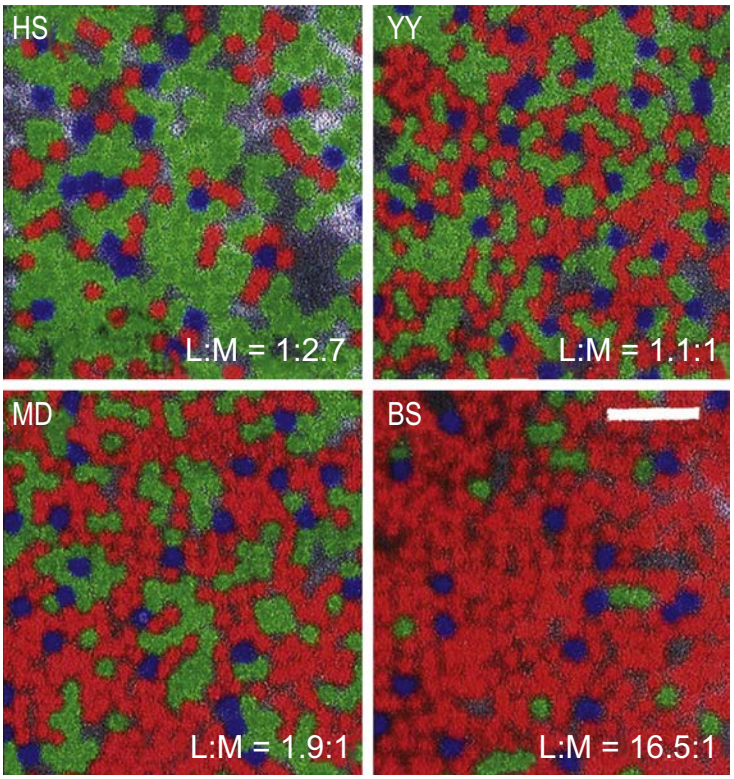


Fig. 5.28 Images of long-wavelength (L, *red*), medium-wavelength (M, *green*) and short-wavelength (S, *blue*) cones as seen using adaptive optics in four healthy individuals. There is a constant ratio of blue to red and green cones but considerable variation exists between red and green cones between individuals. (From Webvision: Color vision, by Peter Gouras: http://retina.umh.es/webvision/color.html.)

specific wavebands. The amino acid sequences of these proteins are known and there are surprisingly few differences in the three cone opsins, especially in the transmembrane regions of the proteins that are important for retinal binding (see Chapter 4, p. 264).

See https://expertconsult.inkling.com for more information on the phototransduction cascade and channel opening (eFig 4.25)

These differences must, however, account for peak wavelength sensitivity of each of the three opsins, thus revealing the extraordinarily fine spectral tuning that occurs at a molecular level. For instance, a threonine at position 65 correlates with the 'red' opsin, while isoleucine is present at the same position in the 'green' opsin. Although the mechanism of light-induced activation of rhodopsin and cone opsins is in principle based on binding of the same chromophore (11-cis retinal) and conversion of 11-*cis* retinal to all-*trans* retinal (see Chapter 4, p. 264), the differences in detail are highly physiologically significant. For instance, although the light response by cones is 100 times less sensitive, the cone opsin responses are several-fold faster, capturing something in the order of 500 photons of light per second. This is related to the availability of Ca^{2+} ions, regulated by a guanylate cyclase-activating protein. In addition, about 10% of cone opsin may be in the apo-state (i.e. lacking any binding to retinol) but still retains a sufficient level of activity to weakly activate transducin, in what is termed 'dark' noise (see Box 5.6).

Despite the similarities in amino acid sequence the crystal structure of cone opsins has not been resolved, unlike the structure of rhodopsin (see Chapter 4, p. 178). However, using an 11-cis-6mr-retinal analogue that has uniquely sensitive binding properties, it has been shown that the amino acids, Tyr262 in blue opsin and Trp281 in red/green opsins, act as selectivity filters for wavelength discrimination (Katayama et al., 2019).

Studies using microspectrophotometric techniques and infrared photography have shown that the distribution of the photopigments in the short-wave, middle-wave and long-wave sensitive cones (Fig. 5.27) is not interdependent as might have been expected on the basis of receptive field analyses (see later) but is random or clumped at least for the long- and middle-wavelength receptors. In the case of short-wave receptors, some degree of organization has been observed from immunohistochemistry data, which show that there are different P ganglion cells for spatial, chromatic and other functions (see section on retinal connections p. 311). In the foveal region, blue-sensitive cones are by far the most infrequent while, in humans at least, psychophysical studies suggest that

long- to middle-waveband cones exist in a ratio of 2:1 but is highly variable (Fig. 5.28).

Do retinal photoreceptor opsins and ipRGC melanopsin share connections?

The role of the melanopsin photopigment in ipRGC induced pupillary function and circadian/photoperiodic signalling are now fully established (see Chapter 4, p. 256). Cone responses may contribute to the overall response particularly in the early slow pupillary response (Hayter and Brown, 2018). Conversely, light activation of melanopsin has been shown to induce a visual sensation in the form of a blurred unpleasant image, which can be recorded in V1 (Spitschan et al., 2017). This sensation may account for the unpleasant sensation associated with bright lights.

Convergence, Young–Helmholtz and Hering

Responses between photoreceptors and neural cells

Considerable processing of information occurs in the retina between the photoreceptor and the ganglion cell. In the magnocellular pathway, which deals with light and motion detection, M ganglion cells have large receptive fields and many rod photoreceptors feed information indirectly onto a single ganglion cell (see earlier). In contrast, the parvocellular pathway deals with spatial resolution and colour vision, P cells have small receptive fields, and a single cone cell may have sole input to a single bipolar cell (see Chapter 1, p. 45). At a retinal level, convergence in the case of colour vision is non-existent (although there is convergence for colour in the cortex).

In spite of this, sensory perception of colour does not correlate directly with stimulation of a wavelength-specific cone. As we have seen, a red-specific bipolar cell responds simply by producing a change in the firing rate in its nerve terminal, which of itself cannot be distinguished from a similar change in a green-specific bipolar cell, i.e. the bipolar cell cannot distinguish wavelengths although its receptor can. In addition, as there is considerable overlap in the spectral sensitivity of the receptors (see Fig. 5.26 and Box 5.10), some degree of confusion must exist in the initial response, which is smoothed out during processing.

Ganglion cell responses and opponent colour theory

Smoothing out this confusion is achieved by the colour-opponent mechanism based on the receptive field organization of the ganglion cells (see Chapter 4, pp. 270–272). In this scheme, there are three colour-opponent arrangements: a reciprocal ON-centre red/OFF-centre green, in which the bipolar cell receives stimulatory or inhibitory input from a single red or green cone; a similar ON-centre blue/OFF-centre yellow; and a third ON-centre white/OFF-centre black in which the bipolar cell receives input from all three cones and in which colour mixing occurs. This has the effect of greatly refining the spectral sensitivity of the ganglion cell responses; it allows the perception of hues and unsaturated colours and accommodates the Young–Helmholtz trichromatic theory of colour vision with Hering's colour-opponent scheme. It also goes some way towards explaining the various colour anomalies found in humans.

Studies have shown that blue-ON/yellow-OFF responses arise from a distinctive bistratified ganglion cell type derived from a dual excitatory cone bipolar cell input: an ON-bipolar cell receiving input only from S cones and an OFF-bipolar cell contacting L and M cone inputs. The mechanism of red–green opponency is still unknown. However, a recently proposed theory that incorporates a contribution of white from rods stimulated under light conditions to the cone input allows a unified concept of how vision is integrated through simultaneous rod and cone input to allow subtle visual perception, such as hue discrimination, motion detection, orientation selectivity and others. Indeed, the significant interactions between rod and cones at the photoreceptor synapse and their respective ON/OFF bipolar cells has been shown in a quantitative analysis (Fig. 5.29). This last concept has been further modified with the discovery that melanopsin-containing ganglion cells not only sense irradiance (see Chapter 4, p. 256), but can also sense colour if the threshold is set sufficiently high. Thus, while the trichromatic theory still holds true at the fovea, with three sets of cones detecting L+ (L+M), (L-M)+M, and (S-(L+M)) ON–OFF colour opponent receptive fields, in the periphery there is a fourth receptor, generating a tetrachromatic mode of colour vision detection. Whether this modifies perception of colours as tested in colour matching experiments is not known and the most accepted notion at present is that this fourth receptor influences visibility rather than specific wavelength detection (Fig. 5.30).

Colour constancy

We use colour to detect and recognize objects. Most of the light we detect is light reflected from objects and their colour depends mainly on the surface properties of the object and not on the illuminating light. The wavelength of the reflected light clearly varies with the lighting conditions. In spite of this, the colour of an object remains the same, a phenomenon known as colour constancy, which is a function of higher visual processing in the cortex.

Factors determining the perception of colour (colour sensitivity) as well as other perceptions (lightness of an object, contrast sensitivity, sensitivity of depth perception)

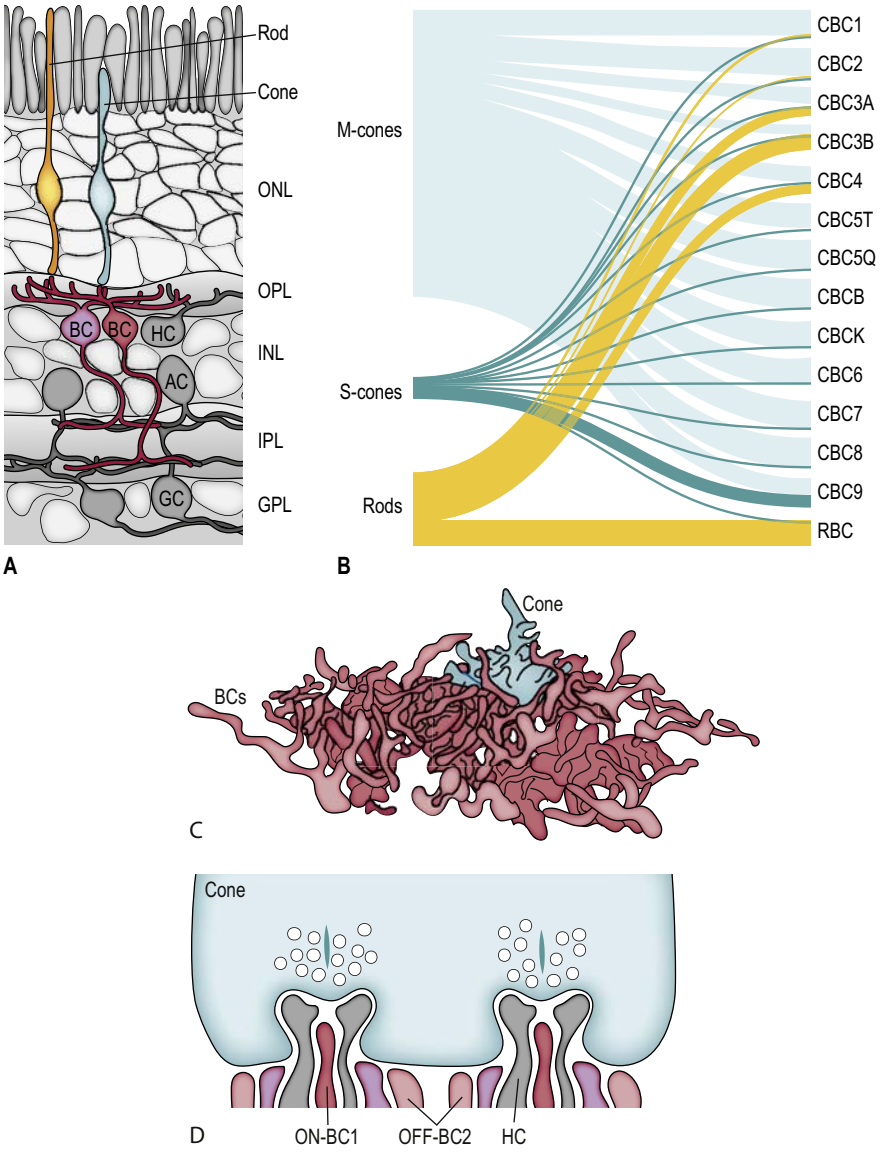


Fig. 5.29 **(A)** vertical view of the mouse retina with five classes of retinal neurons organized by stratification. **(B)** quantitative connectivity map between cones, rods and bipolar cells (BCs). Photoreceptor types are shown on the left, and BC types are shown on the right. Connectivity is scaled to show proportion of BC input received from each photoreceptor type. **(C)** cone axon terminal (cyan) and contacted postsynaptic cone BC dendrites (red, orange), demonstrating the divergence at the photoreceptor synapse. **(D)** two ribbon synapses from a single mouse cone axon terminal (cyan) with postsynaptic partners: invaginating ON-cone bipolar cells (CBCs) (red) and horizontal cells (grey). OFF-CBCs (yellow, orange indicate dendrites form different cells) forming basal contacts. Note that only two cone ribbon synapses are shown, from the total set of 10. *AC*, amacrine cell; *GC*, ganglion cell; *GCL*, ganglion cell layer; *HC*, horizontal cell; *INL*, inner nuclear layer; *IPL*, inner plexiform layer; *ONL*, outer nuclear layer; *OPL*, outer plexiform layer. (From Rogerson et al., 2017.)

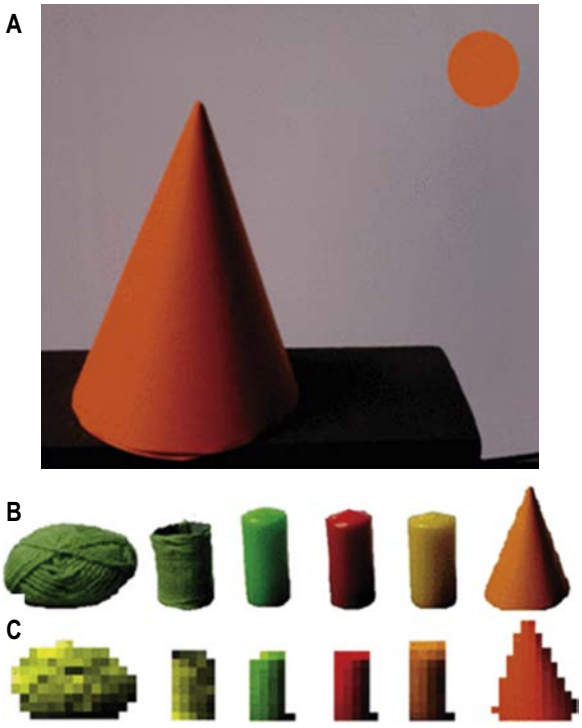
are influenced by several factors including attention, fixation and eye movement. For instance, we tend to take the brighter parts of an object much more into account when we make colour or brightness matches.

Further information on colour constancy and colour perception is available at: https://expertconsult.inkling.com

Remarkably, colour constancy remains stable during a normal life span despite changes in peripheral vision and lens

COLOUR CONSTANCY AND COLOUR PERCEPTION

Both the illumination and the light reflecting from an object, as well as the composition of the scene, determine the amount of light reaching the eye. However, only the reflectance is constant and is therefore the most important factor for vision. When we look at an object, we interrogate the scene both with eye movements and attention and where we look strongly influences our judgement of object properties such as lightness and texture. Our natural tendency is to fixate objects that are brightest and we fix even more closely on the brightest parts of an object and use this level of lightness to determine its colour (as in colour matching tests). As shown in eFig. 5.7, Toscani and colleagues tested an individual's colour matching abilities using the experimental set-up shown: the orange paper cone (**A**) is variably 'bright' from one side to the other and was one of a series of real objects used to match spectro-radiometrically with a corresponding image on a cathode ray tube monitor (Toscani et al., 2013). Several real objects were used, such as a green wool ball, a green wool cylinder (same wool), green candle, red candle, yellow candle, and orange paper cone as shown in (**B**), while (**C**) shows the spectroradiometric data transformed to RGB (red–green–blue) values. The reason for this is that the brightest or lightest, i.e. most immunodominant, part of an object gives us the greatest information regarding the object's reflectance and, because it is constant, we rely more heavily on this for visual judgements. A recent review of colour vision, perception and colour constancy has been published by Conway (2009).



eFig. 5.7 Matching set-up for generating visual information (see text for explanation). (From Toscani et al., 2013.)

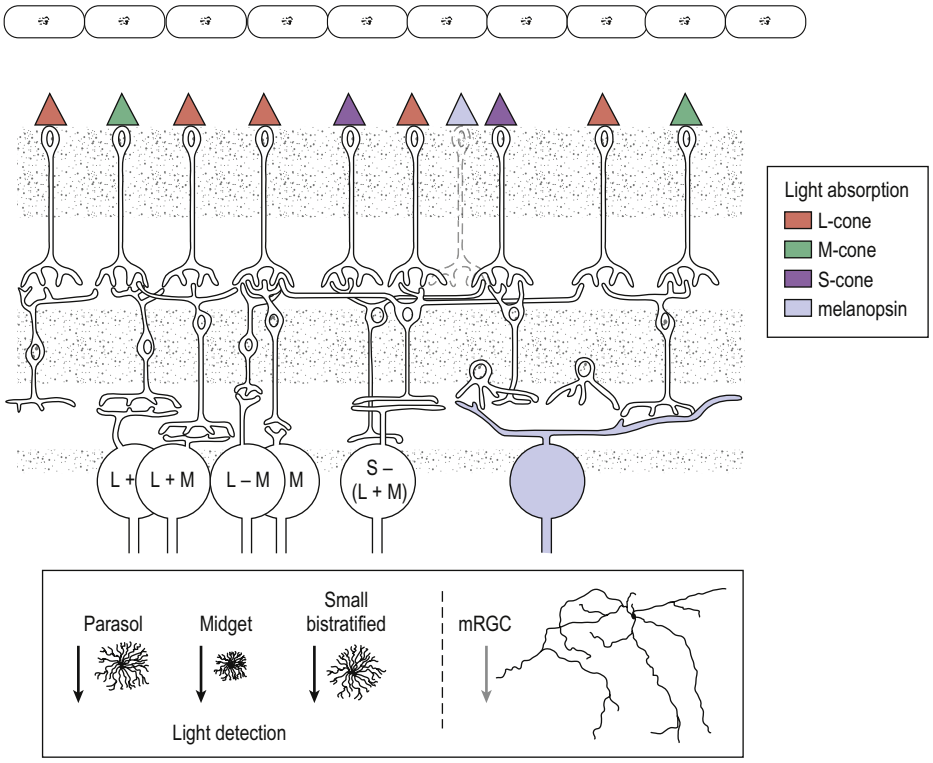


Fig. 5.30 Schematic diagram of the retina showing the three light-sensing cones: in photopic conditions, rod vision is minimal (as Maxwell claimed: 'most vision is colour vision'). Detection of light is mediated by multiple types of ganglion cell serving the three main channels of composite light *(L+M, L-M, and S-(L+M))*. In addition, there is a small population of newly discovered ganglion cells containing a fourth photopigment, melanopsin, and most recently melanopsin-containing cones have also been discovered, but the nature of their contribution to vision is as yet unknown. (From Horiguchi et al., 2013.)

transparency to yellow light. Cortical compensatory mechanisms are considered to allow this. However, a full understanding of the mechanism of colour constancy has been shown to be incomplete by the optical illusion inherent in perception of 'the dress' (Fig. 5.31).

Colour blindness

Some of the defined colour vision defects can be explained in simple terms of loss of one or other specific type of receptor. However, in practice the situation is often more complex, involving not the loss of one particular receptor but the production of combination genes as the result, for instance, of aberrant crossing over in meiosis (see Chapter 3, p. 138); these produce proteins that are intermediate in their spectral sensitivities, thus reducing the range of responsiveness of the protein.

Monochromatism

Rod monochromatism occurs in about 1 in 30,000 of the population; such individuals have true achromatic vision, low visual acuity (0.1–0.3), find high-intensity lights uncomfortable, display nystagmus and may have some signs of macular dystrophy. These patients have morphologically normal cones in their outer retina but their functional status is unclear. It has been suggested that they have a single type of blue cone.

Cone monochromatism is very rare (1 in 100,000). These individuals have normal visual acuity but cannot discriminate coloured lights of equal luminosity. Apparently, cone monochromats possess all three types of cones, indicating that the defect occurs in cortical processing, probably in area V4.

Dichromatism

Dichromatism occurs when the affected individual matches all colours with mixtures of two primaries. Therefore, the range of secondary colours is restricted. Protanopes are missing the red wavelength, deuteranopes the green, and tritanopes the blue. Mixing of the two colours will produce a sensation of white at certain specificities, which for protanopes is 495 nm and for deuteranopes is 500 nm. The dichromat cannot distinguish the large range of non-spectral hues from spectral hues



Fig. 5.31 Optical illusion: Dress colour debate goes global. A debate about the colour of a dress has become an internet sensation. Alana MacInnes, of Uist, and Caitlin McNeill, from Colonsay, the Hebrides, Scotland, sought views on a social media site (Tumblr) about whether it was gold and white or blue. (BBC News, 27 Feb 2015.)

which the trichromat can, leading to a much narrower range of colour detection by the isocolour charts.

Anomalous trichromatism

Red–green colour deficiency occurs in around 10% of males and is X-linked (see Chapter 3, p. 156). Due to their close relationship on the chromosome, unequal recombination events can readily occur, thus removing the green (deuteranopia) or the red (protanopia) genes. Defects in S cone opsin are much less frequent.

Red–green colour deficiency is rarely absolute and in effect such individuals are 'anomalous trichromats' in that they use different proportions of the three primaries to match colour. This is due to the fact that with one single L gene and several M isogenes (mostly single gene polymorphism at Ser180Ala on the red opsin gene), all of which have considerable homology, there is considerable scope for generation of hybrid genes. Thus, protans use more red, deutans more green and tritans more blue. The colour-anomalous individual differs from the trichromat and the dichromat in that he or she will not accept those matches that the other two agree on. This is the common form of 'colour blindness', occurring in 10% of the male population.

Achromatopsia

Colour blindness may also be the result of defects in cortical processing (area V4). Congenital (rare) or acquired lesions in the lingual or fusiform gyrus are associated with cerebral achromatopsia (also accompanied by prosopagnosia: a failure to recognize familiar faces, i.e. from memory). Interestingly, the anterior temporal lobes are the sites for face recognition and are connected to visual cortex area V4, which among other things is the site for colour detection. Similarly, cortical lesions in the superior temporal sulcus (V5) can produce defects in the ability to detect motion (motion blindness or akinetopsia; see pp. 344-346). Isolated defects in form vision have not so far been detected, possibly because they involve more than one cortical area, e.g. V3 and V4, plus connections to other cortical regions involved in psychophysical attributes, and texture analysis.

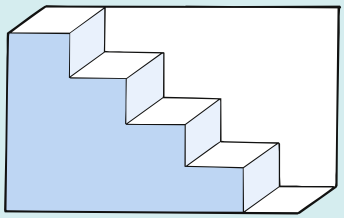
Visual perception

Visual perception is the end product of the processing or reinterpretation by the cortex of sensory responses made by the retina to visual stimuli. However, a strict separation of cortical and retinal events does not occur as some degree of processing takes place in the retina and, conversely, certain processes such as instantaneous parallax (see later) occur so quickly that it is difficult to believe they occur exclusively at a cortical level. In addition, perception should not be considered solely as the end product of the processing of sensory information. Instead, it is part of the 'action–perception cycle', in which perception modifies activity, which then modifies perception in a continuous cyclic pattern, the boundaries of which become indistinct (Wexler and Boxtel, 2005). Motor activity (head and eye positioning, etc.), therefore, is central to perceptual processes such as stereopsis (see later) and detection of colour (see earlier).

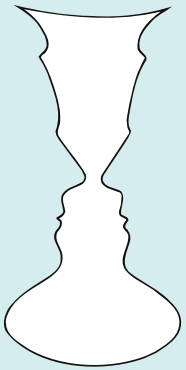
As previously mentioned, there are many different types of visual stimuli, each of which may produce one or more different psychophysical perceptual responses. Sensory perception occurring at an elemental level encompasses visual stimuli such as luminosity, flicker, colour and form, because it involves simple processing of features such as points and lines or wavelengths. However, even such an apparently fundamental function as visual acuity determination involves higher levels of processing because it is more than simply a point-to-point projection of the retina on the cortex.

Box 5.11 **Visual illusions**

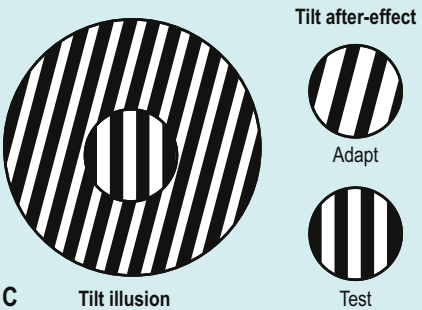
Visual illusions such as the Schrödinger staircase (**A**) and Rubin's vase (**B**) occur in a cyclical manner in which each of the perceptions regularly alternates. Such illusions have a periodicity and are in this sense time-dependent. In addition, they can be modified spatially, as in the tilt illusion (**C**) in which a vertically oriented tilted image occupying a surround regions of vertical stripes will generate the illusion that the orthogonally vertical surround stripes are tilted in the opposite direction. This illusion can be overcome by adaptation using an image after-effect induced by gazing for 30 seconds or more at stripes tilted anti-parallel to the original image and then quickly fixating on the tilted image once more. The periodicity of the fluctuations is also alterable with drugs.



A



B



(**A**) Schrödinger's structure; (**B**) Rubin's vase; (**C**) the tilt illusion.

Evidence for higher integrative activity at the cortical level comes from illusions such as the Schrödinger staircase and Rubin's vase as well as more specific illusions that involve both space and time such as the tilt illusions (Box 5.11). A clear example of the role of cortical function in visual perception is the phenomenon of colour constancy in which large variations in the chromaticity of an object, induced for instance by changing the wavelength of the

illuminating light of the object, do not alter the perceived colour of the object: a yellow banana remains yellow even when illuminated by a green light (see earlier).

Our understanding of these processes has been greatly advanced by the careful analysis of both the stimuli and the responses and has allowed specific functions to be attributed to discrete regions of the cortex.

Monocular versus binocular vision

Positioning objects in space

Most of the primary visual sensory responses are monocular and are not changed by binocular viewing. Images of objects are projected onto definite positions in space (spatial perception) and each retina has its own delimited visual field. However, the position of an object in space is not an absolute entity but is related to the position of the observer and of other objects. The relative position of objects can be determined only if the retinal sensors are composed of discrete units that have precise 'markers' for localization. This, indeed, is the basis of the visual field.

In spite of this, objects appear fixed in position even when the observer changes position; simple ray diagrams demonstrate that a new set of retinal receptors must be stimulated every time the relative position of the object to the observer changes, but the observer does not experience the sensation of motion. This 'image stabilization' is achieved by compensatory psychophysical events at the cortical level. However, extensive non-linearities in ganglion cell function (e.g. ON/OFF responses) have shown that a significant proportion of this neural processing takes place at the retinal level through selective inhibition of ganglion cell firing (see earlier) and allows the observer to distinguish between 'real world motion' and apparent motion due to changes in positioning of retinal stimulation.

The existence of such mechanisms can be deduced in part from experiments showing that accurate localization of objects does not occur with all forms of eye or head movement. For instance, if the eye of an alert individual is forcibly moved using a surgical instrument such as a squint hook, the image is falsely projected to an incorrect position as if the eye has not moved. Spatial perception on normal eye movement must therefore be integrated with, if not controlled by, higher centres within the brain, such as the frontal cortical eye fields, which influence motor discharge in the ocular muscles. The corollary, of course, is that the proprioceptive stretch receptors in the ocular muscles do not have a role in determining eye position as was previously thought, but that their probable role is simply to coordinate muscle tension in opposing muscles at a local 'axon reflex' level. This, however, may not hold true for all situations.

For instance, a problem arises in analysis of images that are perceived during slow visual tracking of a moving object. Despite the fact that no compensation is made for movements of the eyes during tracking, the changing position of the object is accurately observed and followed. This indicates that the higher centres are receiving a continuous flow of information from centres controlling ocular muscle movement, which is assimilated into the total information concerning object positioning; the actual adjustment of speed of eye movement to permit accurate tracking is achieved through visual input, which is ignored by the perceptual process. This means that having initiated a tracking movement, sequential images are interpreted on angular velocity assumptions determined by this initial response and ignored by the higher integrative centres. These assumptions may, of course, be inaccurate, especially if the velocity of the moving target changes. Thus, any induced errors require repetitive readjustments of the tracking response.

Conclusions derived from experiments such as these are greatly influenced by the design of the experiment. It has been observed that the perception of heading (i.e. the direction taken by the observer under conditions of radial retinal image flow or optic flow) in a situation where a moving target is also fixated can be achieved accurately only if extraretinal information concerning the position of the eyes is available, i.e. via proprioceptors from extraocular and head and neck muscles. Under special circumstances, such information, termed structure from motion (SFM) information, can be integrated with purely retinal image information. In addition, the resultant perception can vary significantly depending on whether the object or the observer is moving, even if the relative disparity in motion between them is the same. For curvilinear heading, the rotational velocity of the eye-in-the-world needs to be as precise as 0.2 degrees/s (Perrone, 2018). This sort of information has direct relevance to clinical problems, such as oscillopsia, where image stabilization is lost and the patient experiences 'retinal slip' (see next section). In such patients it is not clear whether the defect lies in the well-established vestibulo-ocular reflexes or whether loss of a putative cervico-ocular reflex, via neck proprioceptors, is contributory.

Thus, spatial constancy, i.e. maintaining a 'stationary' percept of moving objects despite their repeated and sudden changes in retinal stimulation, is a fundamental visual perception. Spatial constancy also has a memory component. When an illuminated object is viewed in the dark, the eyes will, after an initial 'searching' response, adopt a position close to fixation (approximately within 2 degrees of the target) when the illumination is switched off (Fig. 5.32). This has been attributed to some element of positional sense from an extraretinal source, such as head–eye position in space and locomotion (egocentric versus allocentric signals). Similarly, the constant drift of the eyes towards fixation in the dark has been attributed to a similar mechanism (see next section). 'Place' cells occur in the hypothalamus, thalamus and hippocampus of the brain

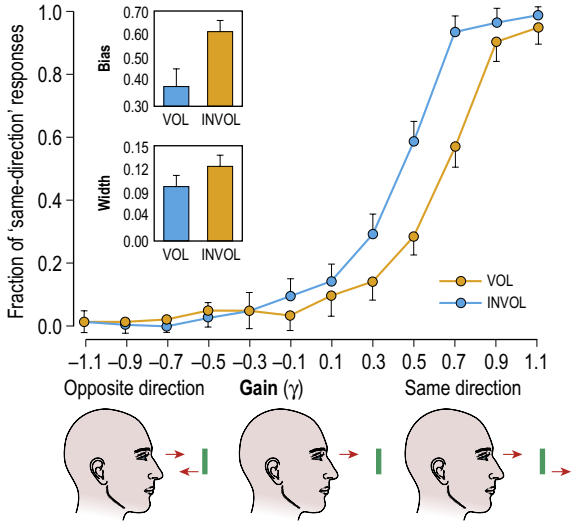


Fig. 5.32 Results of the voluntary *(VOL)* and involuntary *(INVOL)* eye movement studies. The curves show responses averaged over subjects and over all simulated distances (there was no significant effect of distance). Mean bias and width were calculated by fitting the data of individual subjects with the logistic curves and averaging the parameters thus obtained. Bars indicate between-subjects standard errors. (From Wexler, 2003, with permission from Blackwell Publishing.)

and correct visual input to active locomotor and possibly navigational (optic flow, see later) input.

Measuring by eye

We frequently use our vision to measure things by eye, e.g. to line up objects in a row or to determine the distance between two points. Simplistically, it might be thought that, for example, the distance between two objects should correlate with a defined distance between stimulated receptors in the retina, or that the length of a line will be determined by stimulation of a fixed number of retinal receptors. However, the psychophysical basis of these abilities is not as simple as it might appear. For instance, two lines of the same length running in different directions may stimulate a different number of receptors as the direction of lines will be distorted by the curvature of the eye and its position on eye movement. In addition, errors will be introduced on head movement since the head rolls further than the eye in the socket (some of this disparity is compensated for by input from the semi-circular canals). Attempts have been made to adjust for image distortion due to eye curvature and movement using after-images, which obviously must reflect closely the actual retinal stimulation versus the 'real' direction of the object, but even this is not sufficient to account for the fact that we perceive the object in its true position and not in its distorted position. This is nicely demonstrated by Aubert's phenomenon (Box 5.12) in which the speed of a moving object appears slower when the object is tracked by eye movements than when its motion is detected with the eyes remaining in one position, i.e. the object is not visually tracked.

A compensatory mechanism based in the higher centres must therefore be in place; it must be highly developed because the accuracy of measuring by eye is extremely high: for detecting the orientation of horizontal lines the error rate is 0.2% and for vertical lines it is 1.0%, as determined with a perceived direction test. The greater accuracy for estimating vertical and horizontal lines over oblique lines is not the result of preferred direction of eye movements or of the numbers of retinal receptors stimulated but is a function of cortical activity.

Considerable evidence has now accrued to show that the visual cortex contains specific cells that are responsive to the orientation of lines. Indeed, the ability to 'parse visual scenes for the orientation of purely spatial cues' has been shown to be a fundamental property of even the simple insect brain. Vertical and horizontal orientation detectors occur as simple cortical cells (see later) and are distinct from motion detectors. Orientation detectors are considered essential to the analysis of form.

The patterns of objects also greatly influence our ability to make visual measurements. For instance, two angles can be accurately compared if we can fixate each of them directly and if the sides are parallel. In addition, there are numerous examples of optical illusions where objects to be compared appear to be different in length or area/size if one of them has been altered by the addition of other visual cues that are interpreted by the higher visual centres in one particular direction (see Box 5.11).

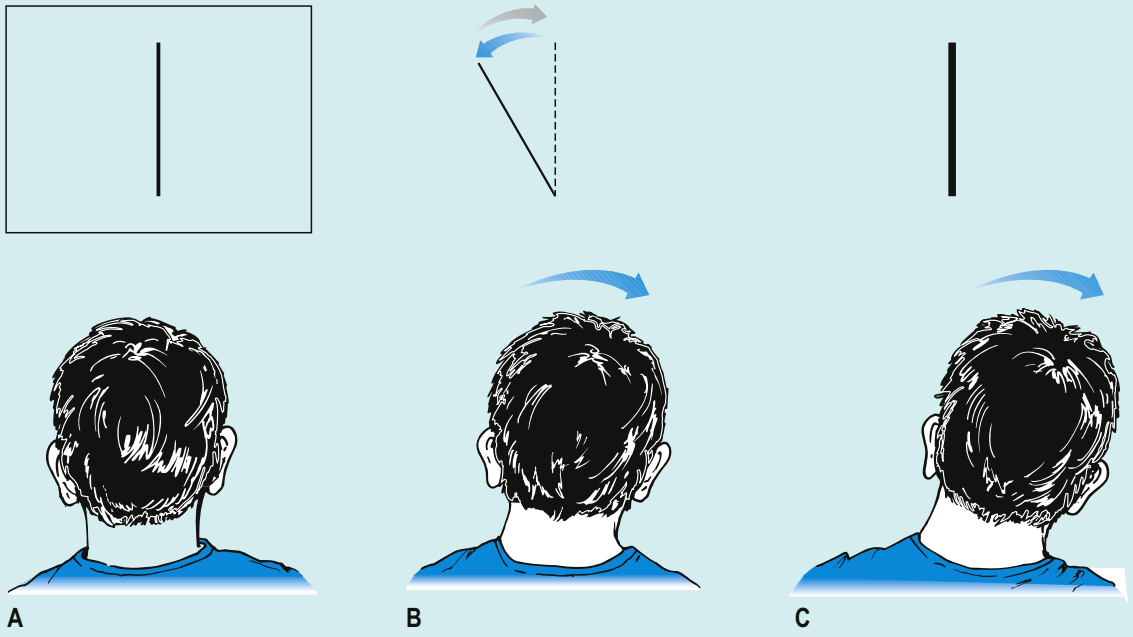
Therefore, both patterns and directions are important in visually estimating object dimensions: the retina uses the horizontal and vertical meridians as x and y axes to provide coordinates and thereby to pinpoint objects in space. As these axes are subject to displacement by, for instance, eye movement, it is important that a psychophysical compensatory mechanism exists that will reinterpret the position of the coordinates to project the real position and direction of the object in space.

Are two eyes better than one?

As most objects are three-dimensional, it is clearly important that depth perception is achievable by the visual system. The major advantage obtained by binocular viewing is that it permits depth perception or stereopsis. This occurs in the cortex and depends on the fusion of images from

Box 5.12 **Aubert's phenomenon**

A vertical bright light viewed in a completely dark room will tilt to the left if the head is slowly tilted to the right (**B**). If the head is tilted suddenly or if the line is viewed in the light, the line appears upright in its normal position (**C**). Thus, information on retinal position is fed via the semi-circular canals to the object-positioning centre.



(**A**) Resting position; (**B**) slow head tilt to right, object appears to move to left; (**C**) fast head tilt to right, object stays upright.

each eye. However, depth perception is a complex event and information from many sources is used to achieve it.

One mechanism for depth perception could be related to the convergence of the eyes. Fixation of an object with both eyes requires a variable degree of convergence depending on the distance of the object, and information derived from this can be used to determine object distance. Indeed, it has been suggested that variations in convergence can lead to three-dimensional illusions, although these may be caused by other mechanisms.

Is it possible to perceive depth with one eye only? Determination of object position using x and y axes as described earlier provides two-dimensional information only. Geometrically, it should not be possible to obtain three-dimensional information using one eye. However, certain cues, mostly built upon previous experience, indicate that some sense of depth is possible, e.g. by comparing the relative size of objects (for example, a person and a house), the blue colour of distant mountains (although they should be yellow: they are darker blue in comparison with the background light blue of the sky), overlapping edges of objects, effects of light and dark shading, effects of texture and parallax on movement of the observer's head. It has also been suggested that the sensory feedback from the ciliary muscles on accommodation might provide some information centrally regarding depth (similar to the effects of convergence when both eyes are used), but this is unlikely.

When both eyes are used to observe an object in the straight ahead (primary) position, the image is still perceived as one, even though the image of that object must appear positionally but symmetrically different to each eye individually; indeed the image can be treated as if it were projected from a single centrally placed eye (the 'cyclopean' eye). When a second object is presented in the primary position, but closer to the observer than the original object, the second object is seen double when the original object

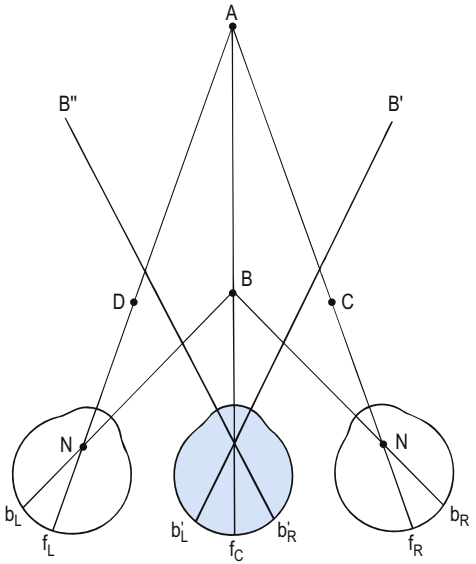


Fig. 5.33 The 'cyclopean eye' *(shaded)* is the abstract notion representing fusion of the two images from left and right. Image *A* is fused (f_L + f_R at f_C) but image *B* cannot be simultaneously viewed without diplopia (heteronymous) (b_L + b_R at b'_L and b'_R) due to the projected images at *B′* and *B″*, respectively. Homonymous diplopia would occur if *B* was distant from *A*. Using either eye alone for alignment of the object would entail point *C* aligning with *A* (right eye) and point *D* aligning with *A* for the left eye. *N*, Nodal point. (Courtesy H. Dawson.)

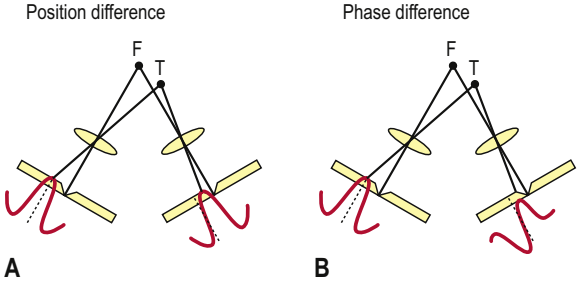


Fig. 5.34 Depth perception based on binocular disparities. The fovea of each eye fixates point *F*; because object *T* is closer to the observer than *F*, the image of *T* falls at a different retinal location in the two eyes. The *dotted line* marks the equivalent retinal location in the two eyes. Neurones with receptive fields in both eyes could detect this disparity in two ways. (**A**) Position difference: the right eye receptive field is an exact copy of the left eye receptive field, but in a different retinal location. (**B**) Phase difference: the envelope enclosing the right receptive field profile sits in the same position as for the left receptive field but, within the envelope, the right receptive field has a different structure, responding best to white light on the right-hand side. When tested with a bright bar, both of these mechanisms produce a maximal response to a stimulus with a disparity equal to that of *T*. (From Cumming, 1997, with permission from Elsevier.)

is fixated (Fig. 5.33). Diplopia in this position is described as heteronymous; when the second object is distant from the first, the induced diplopia is described as homonymous. However, this form of diplopia is rarely appreciated because we normally do not attempt to fixate more than one object in the primary position when viewing with both eyes. When we aim at a target using a second object to line up fixation, as with the sights of a rifle, we normally do so with one eye only (see Fig. 5.34).

A similar form of diplopia can be induced by a divergent squint or by placing a base-out prism in front of one eye; both of these conditions have the effect of re-siting the projection of the image to a false position from its normal cyclopean position. False projection in a squinting eye can lead to the development of a false macula, which in fact is a cortical event because the anatomy of the retina remains the same. The development of a false macula reflects the plasticity of the cortex and indicates that projection of the eye through the nodal point is an innate mechanism. A similar 'pseudofovea' may develop in the presence of hemianopia.

The fact that we rarely experience double vision, even when rapidly and simultaneously fixating many objects in a scene, indicates that the images from both eyes are merged or fused; this is not simply a reduplication of information from both eyes but an actual psychophysical event that is used to provide depth information or stereopsis.

Stereopsis and depth perception

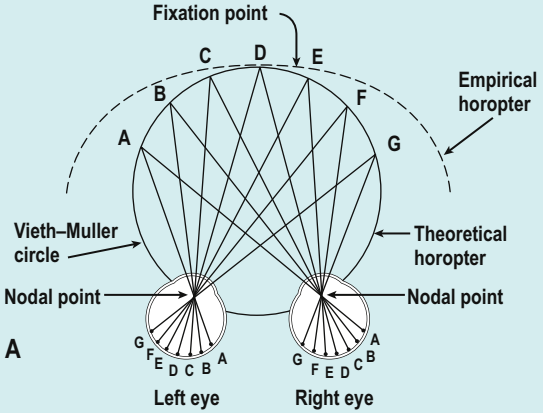
The fusion of images to create the perception of depth requires certain conditions: first, the images from each eye must have corresponding points on both retinas. If all points on a sizeable object were exactly corresponding, however, this would merely lead to a reduplication of information and it is likely that only one of these points would register (this would be analogous to allelic dominance in chromosomal gene duplication in which only one gene is expressed transcriptionally while its partner is not; see Chapter 3, p. 146). Single-cell measurements have shown, however, that impulses from both retinas induce electrical activity in cortical neurones.

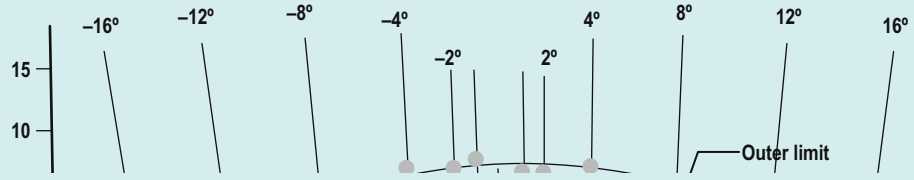
The second requirement for image fusion is that a certain proportion of points are non-corresponding and represent the differences between the two images on the retina, which may be smaller than the width of a single cone, and are known as binocular disparities. There is a disparity limit to the degree of fusion (Panum's limit, see later) and when this is reached, interocular image suppression prevents

Box 5.13 **Corresponding points of fixation on the retina constitute the horopter field of vision**

(**A**) The horopter circle. The theoretical horopter circle is shown as the *dark inner sphere*, while the actual horopter for an emmetropic individual is shown by the *dashed line*. The horopter has been known for centuries. However, in 1818 Vieth-Muller calculated that the horizontal line of binocular fixation was a circle that passed through the centre (nodal point) of the lenses.

(**B**) The binocular single vision horopter field. In practice the empirical horopter is larger than the theoretical horopter since there are many more areas both in front of and behind the circle where single vision can be obtained: these are the inner and outer limits of binocular single vision and correspond to Panum's area.





diplopia dependent on the amount of edge blur. Eventually diplopia supervenes.

Further information on measuring binocular disparities is available at: https://expertconsult.inkling.com

It is the integration of information from corresponding and disparate points that induces the perception of depth. Both position and phase disparity in the corresponding receptive fields are important in the detection of depth cues (Fig. 5.33).

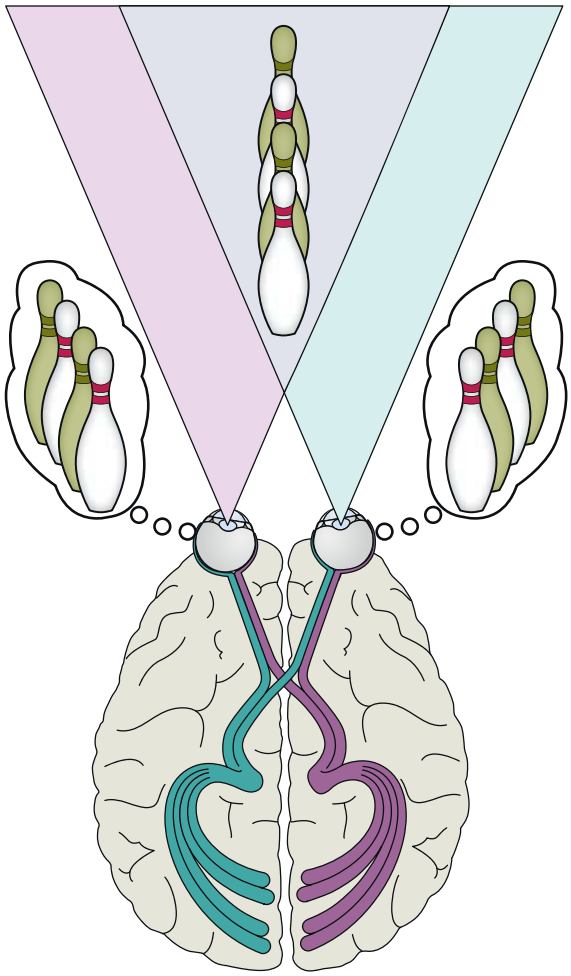
The horopter

Psychophysical and theoretical investigation of the projected points in space where an object is seen as a single image have a long history going back to the ancient studies of Ptolemy and the work of Ibn al-Haytham in the eleventh century. The corresponding regions in the retina include both the horizontal and vertical meridia and equidistant points from each of these meridia. Single vision can be achieved only when the images of the object are projected from each eye to the same point in space.

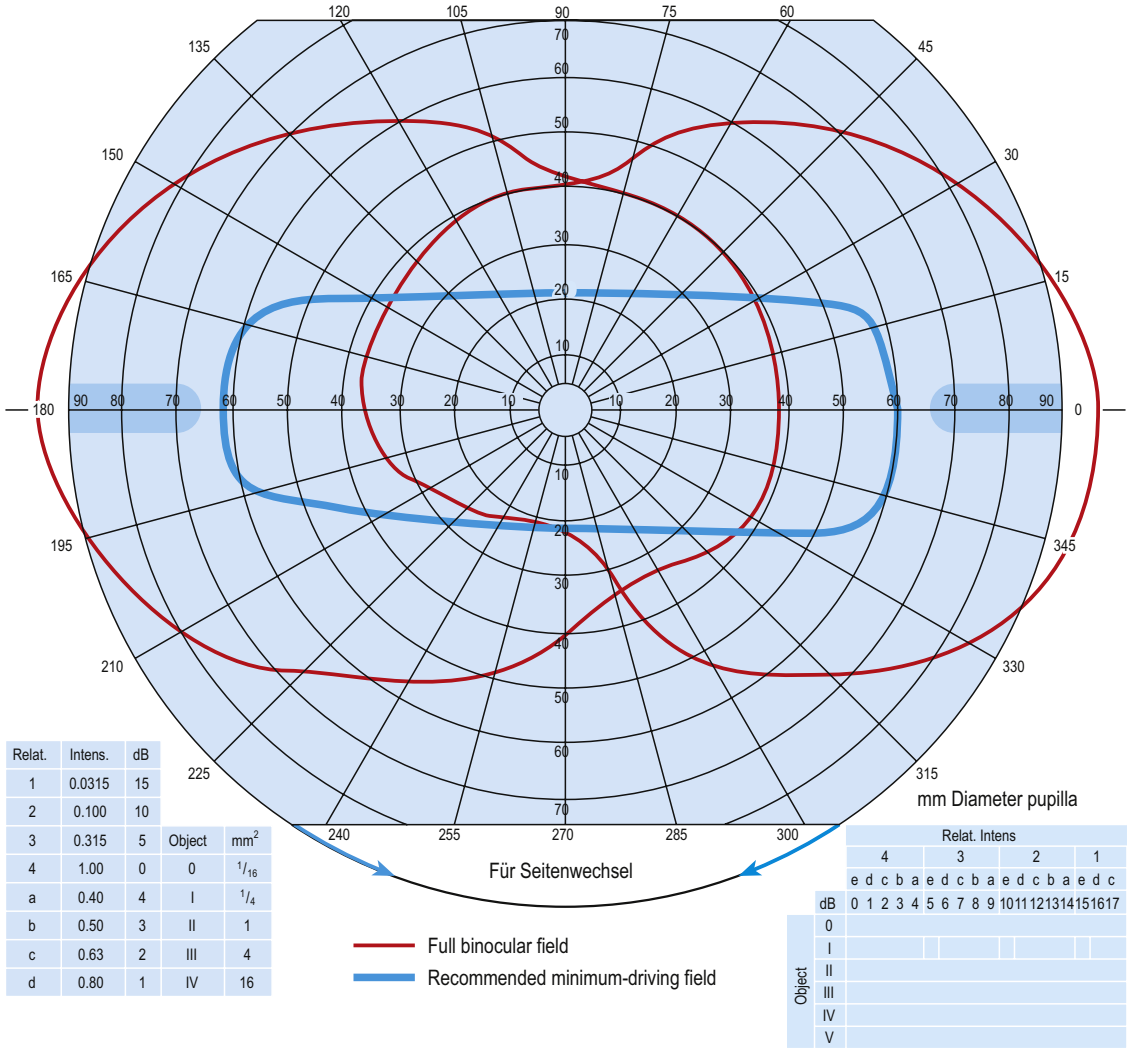
Corresponding points can be charted as a horopter, in which specific points on the retina project to definite single points in space, within the field of binocular single vision (Box 5.13) (the word horopter drives from the Greek: *horos* meaning limit and *optos* of seeing where seeing means seeing a clear single image). The vertical horopter has a backwards tilt that passes through the fixation point and a point near the feet of the observer and is the result of a shear in binocular retinal correspondence. Thus, the vertical horopter takes the form of a cylinder that may be reversed (see Box 5.13). The true (empirical/actual) horopter is strictly

MEASURING BINOCULAR DISPARITIES

The images registered through each eye are different (eFig.5.8) (see open access link to http://www.vision3d.com/stereo.html) and the brain processes these images, making note particularly of the differences (disparities). Simultaneous registration of the images is thought to be necessary for the images to be fused, although there is evidence that each image is registered differently in time, i.e. each eye takes a 'shot' of the image with a very brief time interval between the registrations. Stereopsis can be measured with a variety of devices, one of the oldest being the stereoscope, which is essentially an instrument constructed from a pair of base-out prisms, whose position can be manipulated. The stereoscope has a long history but continues to be modernized, one of the latest being the View-Master, which is now fully within the digital age and includes a video-based version (see http://gajitz.com/view-master-grows-up-modern-stereoscope-video-viewer/). Continual development of stereoscopic viewers has now brought these into use in virtual reality technology, which can be used on a smart phone, e.g. the Virtual View Master.



eFig. 5.8 User-friendly bowling alley images demonstrating image disparity, which underlies stereopsis (see open access link to vision3d.com)



Relat.	Intens.	dB
1	0.0315	15
2	0.100	10
3	0.315	5
4	1.00	0
a	0.40	4
b	0.50	3
c	0.63	2
d	0.80	1

Object	mm²
0	1/16
I	1/4
II	1
III	4
IV	16

Object	Relat. Intens															
	4					3				2				1		
	e	d	c	b	a	e	d	c	b	a	e	d	c	b	a	e	d	c
dB	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
0																
I																
II																
III																
IV																
V																

Fig. 5.35 The field of vision for each eye separately, the field of binocular vision (BSV) and the minimum visual field required for driving licence purposes in the UK are shown. (Courtesy H. Dawson.)

limited to an area of about 3 degrees from fixation, as determined experimentally. A special form of horopter is one based solely on corresponding points, defined as a circle of projected points in space passing through the fixation point and the nodal point of the eye (theoretical horopter). This is truly applicable only in the horizontal meridian, as the vertical meridians are not exactly parallel. Horizontally placed horopters can form a stack of slices, producing a longitudinal horopter named so as to reflect the vertical lines of longitude on the globe of the earth. In practical terms, the longitudinal horopter is not a circle but has a well-defined shape, approximately representing the field of binocular single vision (BSV) (see Box 5.13).

The field of BSV is an important parameter, not simply from a physiological standpoint but also for socioeconomic reasons. The normal visual field of each eye is approximately elliptical, with a considerable degree of overlap (Fig. 5.35), and the overlapping fields of each eye represent the field of BSV in which full stereopsis is assumed to occur in the context of horopter-related corresponding points on the retina. In the UK, a certain minimum field of BSV is required to qualify for a driver's licence and is defined as a BSV field of 20 degrees above and below the horizontal meridian and 60 degrees to either side of the vertical meridian (Fig. 5.35). Measurement of visual fields can be performed by many techniques:

currently, static automated visual fields are the normal practice, although kinetic and flicker-based fields are also highly informative. In addition, microperimetry is a more detailed method of visual field testing, which, when combined with scanning-laser ophthalmoscopy, permits fine retinal mapping and analysis of discrete regions of retinal function or dysfunction.

Further details on microperimetry is available online at https://expertconsult.inkling.com

Thus, it has considerable value for detecting retina-derived as opposed to visual pathway associated field defects (see Chapter 1, pp. 99–104). In clinical practice, quite different visual field tests are performed when assessing either of these central causes of visual field defects.

The horizontal horopter is also defined for specific fixation points and therefore certain degrees of convergence; clearly this will change with the distance from the observer. At about 2 m from the observer, the horopter is approximately a straight line, while it is concave to the face within this distance and convex beyond (see Box 5.13).

Measuring stereopsis

The measurement of stereopsis involves at least two parameters: the degree of convergence required to fuse images from two slightly dissimilar objects, and the limits of dissimilarity between two objects at which the two images can be fused (stereoacuity). The former is relatively easy to measure with an instrument incorporating two base-out prisms, known as a stereoscope.

More information on measuring binocular disparities is available online at https://expertconsult.inkling.com

In the stereoscope, two slightly dissimilar but symmetrical images are presented to each eye and the angle of convergence at which the sensation of depth is achieved is recorded. In practice, true stereopsis is not measured by this method because light and shade (i.e. monocular cues) provide considerable amounts of image disparity to the same object viewed by each eye in turn. Other tests that remove the monocular cues but use a camouflaged object include the random-dot stereogram, the random-dot E-test and the Frisby test, in which elements that are non-resolvable monocularly are presented in a random pattern at different disparities and the ability to perceive depth and form in the objects is assessed.

Stereoacuity can be measured as instantaneous parallax, which is the difference in binocular parallax of both objects (Fig. 5.36). The limits of stereoacuity are in the region of 4′ of arc (range 1.6–24′), which is equivalent to an image disparity less than the diameter of a cone. Instantaneous parallax is lost at a distance of about 450 m but varies with the measurement technique. Conversely, within a certain

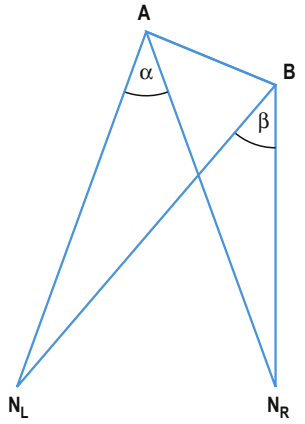


Fig. 5.36 Images *A* and *B* cannot be fused; instead they induce 'instantaneous parallax'.

range of distances, stereoacuity is improved the further away the object is from the observer, although theoretically this should not be so. In this case, the improvement is attributed to greater differences in monocular cues such as relative object size, to which the eye is more sensitive, than instantaneous parallax. Stereopsis also improves with duration of the stimulus, which is not the result of small searching or 'image-refreshing' movements of the eye but reflects the minimum time required for neural processing of the stimuli. Indeed, similar but disparate images can be perceived in three dimensions if they are presented to each eye in sequence, but the time interval has to be small (less than 5′ of arc). The contribution of interocular image suppression has to be considered here (see earlier).

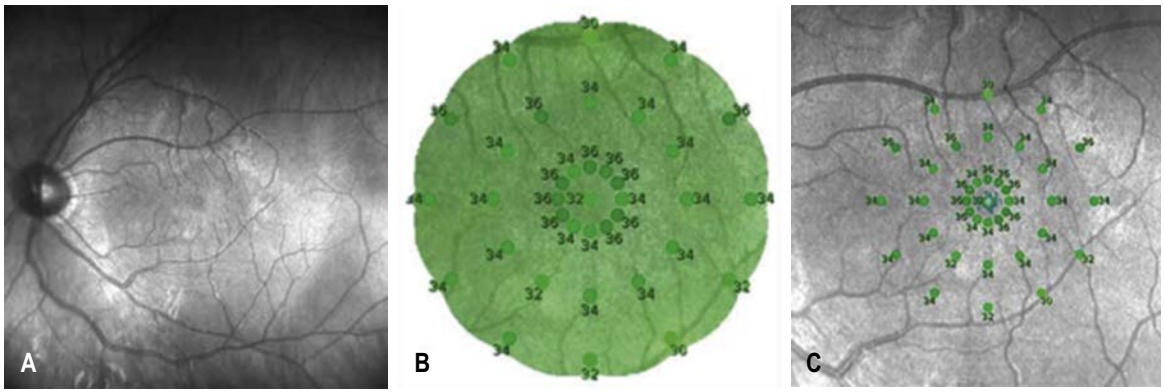
Image disparity and stereopsis

True stereopsis is dependent on disparities between the two images received by each eye, and therefore a certain number of points must fall on disparate points on the retina. It is also essential that these disparate points are fused like the corresponding points. The position of corresponding versus disparate points can be assessed by determining the actual differences in the stereoscopic projection of a point in space from the separate projections of the point made by each eye (Nouin's technique).

There is a limit to the fusional capacity of projected images, which is a circumscribed area known as Panum's area (see Box 5.13). It has been shown that it is possible to fuse greater disparities in the horizontal meridian than in the vertical, and therefore Panum's area forms an ellipse. The size of this area varies between individuals and the threshold disparity that can be fused is greatest at the horopter. The extent of Panum's area is reduced by small

MICROPERIMETRY

Microperimetry assesses retinal sensitivity to visual stimuli over very discrete areas and uses eye-tracking devices to eliminate fixation errors resulting from fine eye movement. Regional retinal sensitivity is mapped to a fundus image (eFig. 5.9) and can very accurately locate areas of defect. Thus, it has been used in numerous diseases in which retinal function is compromised including glaucoma and macular degeneration as well as numerous retinal degenerative diseases. The conventional target stimulus used has been in the form of a small cross subtending an angle of 1 degree but recently, for the central region at least, a circular target, of 1 degree radius, has been found to provide more definitive data (Nizawa, 2017).



eFig. 5.9 Fundus image obtained with SLO from left eye of a normal subject (**A**); sensitivity exam from left eye of a normal subject (**B**); both sensitivity exam and fundus image obtained with SLO from left eye of a normal subject (**C**). All images were obtained from the same patient with MAIA microperimeter (Centervue). (From Molina-Martin et al., 2018.)

'normal' disjunctive eye movements, which can be compensated for by using stabilized retinal images. Double images can be induced outside Panum's area and can be used effectively to estimate depth.

Fusion of disparate images to produce stereopsis tends to invalidate the notion of the cyclopean eye in which single vision is produced by fusion of corresponding points. However, the cyclopean eye is of value in providing a baseline on which an estimate of the degree of neural processing involved in the fusion of disparate images can be made. Studies of neural circuitry in stereopsis have thus shown that it is possible to perceive depth without monocular cues, for instance by using random-dot stereograms and other more novel tests.

In these tests, 9 × 10 picture elements composed of dots, some of which correspond while others are symmetrically disparate, are presented in duplicate to each eye. These studies also reveal that discrete contours or edges are not essential for three-dimensional vision, although the contribution from 'texture analysis' is not clear (see later). Random-dot stereograms are also not quantitative. Importantly, random dot stereograms identified the two neural processes involved in stereopsis: (1) the brain has to decide which two (R and L) viewed elements in the image correspond (correspondence), involving some initial neural processing; and (2) R and L image element fusion is then performed to determine whether there is true 'depth' to the image (Parker, 2016). Stereopsis is essentially the ability to detect patterns which correspond between the right and left eyes and is an important survival property for instance in seeing through camouflage (Parker, 2007).

The level at which processing for stereopsis occurs has been questioned on many occasions because it is such an instantaneous response and is difficult to separate from a retinal 'sensation'. Stereopsis is also sensitive to certain optical effects such as aniseikonia; horizontally, it is affected by as little as a 0.25% change in image size and vertical magnification disparities are simply transferred to the horizontal meridian of the fellow eye. The Pulfrich phenomenon is an optical illusion based on similar processing events (Box 5.14). However, true disparity selective cortical neurones have been detected in V1, although the extent to which these neurones are simply 'rivalrous' (see next section) or stereopsis-inducing may depend on the degree of further cortical processing that appears to take place in the middle temporal (MT, V5) area of the visual cortex, an area associated with motion detection (see later). Furthermore, disparity matching appears to be a two-dimensional and not merely a one-dimensional process, involving fusion of images in a vertical as well as horizontal disparity.

Retinal rivalry and ocular dominance

Retinal rivalry is in essence a term that describes simultaneous perception by each eye individually without fusion of the images. This can be demonstrated as, for instance, when the letters F and L are viewed by each eye separately to produce the letter E. This phenomenon also has a periodicity to it, which is involuntary. It is related to but different from binocular rivalry as illustrated by the Schrödinger staircase or Rubin's vase (see Box 5.11). Both are described as 'bistable phenomena' and are processed at retinal and cortical levels, respectively. fMRI studies confirm that activation of the sensation of monocular rivalry involves recruitment of whole-brain networks in addition to visual area V3 (see later), particularly for complex objects such as faces and houses (Fig. 5.37). Retinal rivalry can also be affected by saccades and extraretinal eye movement signals.

Ocular dominance refers to the preferential use of one eye when performing monocular activities. This can be demonstrated electrophysiologically and is not necessarily related to handedness. It can also be studied in terms of the stimulus strength presented to each eye and the change in the perception of the stimulus. Certain involuntary events take place that bear on retinal rivalry and ocular dominance. For instance, if the same image is presented to each eye at different levels of brightness, then the image in one eye may be suppressed (ocular dominance) or the binocular image might appear less bright than the same image when viewed monocularly (e.g. with a uniocular cataract).

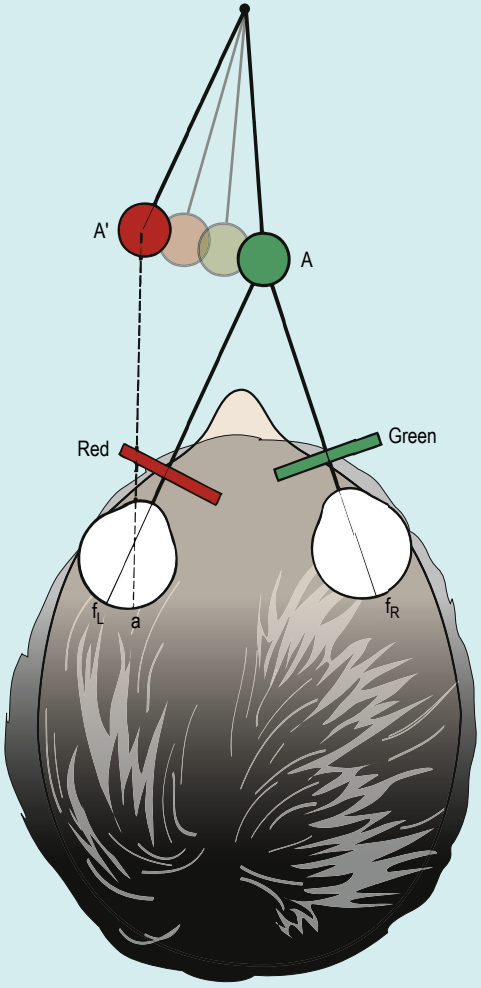
However, whether there is evidence for true cortical dominance for preferential use of one eye has not been established. In a perfectly equally sighted individual, input from one retina will mirror exactly that from the other retina. Ocular dominance can be induced experimentally and is dependent on a critical period of neural plasticity, which has been attributed to molecular mechanisms involving neuregulin /ErbB4 as well as silent synapse signalling via AMPA receptors.

Colour processing

The perception of colour is a complex cortical event that is dependent on input from several sources. The use of mondrians (coloured patterns produced with variably illuminated narrow-waveband light) has shown that the predominance of a given waveband reflected from a surface does not alone determine its colour, but that its colour also depends on the wavelength composition of the light reflected from its surround. Mondrians have been used to demonstrate the phenomenon of colour constancy but their relative artificiality has been challenged: for instance, demonstration of other phenomena, such as the AMBEGUJAS phenomenon, in which perceived colours can change dramatically depending on the three-dimensional surface of wavelength reflectance. Instead, well-defined real colour scenes have been devised

Box 5.14 **The Pulfrich phenomenon**

The illusion of depth can be demonstrated by viewing a swinging, luminous pendulum through both eyes, one of which is covered by a red filter and the other by a green filter. It is thought to be the result of disparate images occurring during the movement of the pendulum stimulating corresponding points at fractionally different times.



that contain cues for the intrinsic surface colours and the recovery of the light source. These studies show that colour constancy is a real phenomenon but is not as absolute as previously thought and depends on input from local and global contrast. This depends on several factors, including spatial configuration and scale, and context. Texture, as in the AMBEGUJAS phenomenon, appears to the most important (Fig. 5.38).

It is clear, therefore, that the reflectance of light is central to the perception of colour. Although the amount of light reflected from a surface might vary, the brain constructs an image that fits the reflectance, which is a constant physical attribute of the object. The brain assesses the lightness or darkness of a surface compared with the surround, for each of the three predominant wavelengths in turn, and this permits it to assign a colour to the surface.

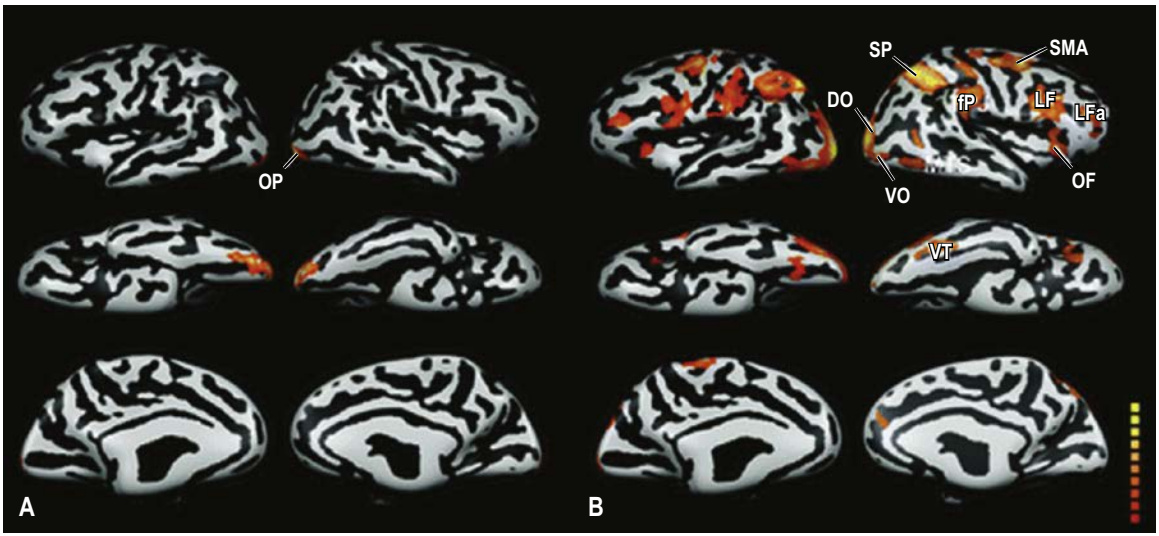


Fig. 5.37 Monocular rivalry as demonstrated by fMRI using a 'passive' stimulation protocol and complex objects such as faces and houses. (**A**) Shows cortical activation for the non-rivalrous control stimulus as a limited region of activation in the occipital pole and includes areas V1 and V2. The lateral, ventral and medial views of the inflated brain are shown (left and right hemisphere). (**B**) Shows cortical activation for passive viewing of monocular rivalry with grating stimuli. Cortical sites included dorso-occipital *(DO)*, ventral-occipital *(VO)*, ventro-temporal *(VT)*, medial temporal sulcus *(MTS)*, superior parietal *(SP)*, temporal-parietal junction *(TPJ)*, supplementary motor area *(SMA)*, lateral prefrontal *(LF)* and anterior lateral prefrontal *(LFa)* and orbital frontal *(OF)*. (From Mendola and Buckthought, 2013.)

Colour is achieved, therefore, by a comparison of the reflected intensities of lights from one surface with those of surrounding surfaces for lights of different wavebands, followed by a comparison of the comparisons.

Shape detection

The detection of form and shape also presents a problem when we consider exactly what we mean by shape. As described previously, specialized cells exist that act as edge detectors for vertical and horizontal lines. However, the shapes we perceive are much more complex than can be simply broken down to a series of discrete lines on x and y axes. For example, most shapes in the natural world are curvilinear and solid, and require significant processing in the cortex. This has led psychophysicists to develop mathematical algorithms and a 'shape index' to describe these shapes and thus provide insight into how the brain might compute the information. These effects can be demonstrated using specially oriented stimuli and show that global orientation detection is not simply the result of input to the primary visual cortex (V1). Second-order orientation detection may therefore also exist as 'collector units' for first-order V1 stimuli. Such collector units may also be affected by brightness and texture.

Contribution of texture analysis and motion detection to depth perception

Depth perception is, of course, not only about locating objects in space but also about perceiving solid shapes. In fact, some of the early work on depth perception involved studies of visual function in aircraft pilots, particularly on take-off and landing. They revealed that motion detection and texture analysis were more relevant for the detection of solid structures, probably the most important aspect of visual perception (Figs 5.39 and 5.40). Many other sources of information combine to produce this effect, including binocular viewing, parallax, illumination and shading, and edge detection. For instance, a random set of dots may assume shape if a group of dots within the set moved in relation to the remaining dots—an image would thus 'pop out' of the page. Similarly, texture and lustre are attributes of an object that relate to discontinuities in colour and/or luminance coming from the edges of the object and significantly affect perception of its shape. Thus, there are multiple extrastriate areas that are activated in stereopsis and depth perception involving many other visual areas (Fig. 5.41).

It has also been shown that the detection of a shape also depends on previous experience/memory and whether the

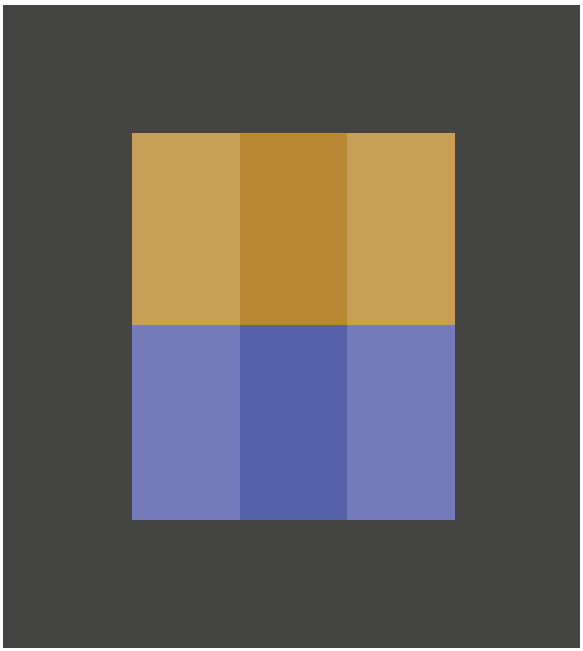


Fig. 5.38 AMBEGUJAS is a recently described visual perceptual phenomenon in which shapes can be discerned from shading of colours. In this image a 3D illusion is created by the darker-shaded colour strip passing through the centre of the blue and orange rectangles, giving the impression of a box. It is, however, ambiguous as to both shape and colour and frequently returns to a flat surface with a grey central strip. (Bergstrom SS, 2011. The AMBEGUJAS phenomenon and colour constancy. Perception 40: 30–38. http://www.ncbi.nlm.nih.gov/pubmed/15460510, Fig 4, Pion Ltd, London. http://www.pion.co.uk; http://www.perceptionweb.com.)

observer is 'expecting' to see the shape. It is clear therefore that this complex response is built up from multiple inputs and that the search for a shape-detecting centre may prove elusive.

Division of labour in the visual system: parcellation and the human connectome

Topographic mapping of brain function continues to advance especially as applied particularly to vision, specific visual functions and the connections between visual areas and other areas that influence visual perception. Zeki's early conceptual framework for what he describes as the division of labour in the visual cortex has become less well defined as the numerous interconnections between different regions of the brain are identified for discrete visual tasks such as depth perception. However, his concept

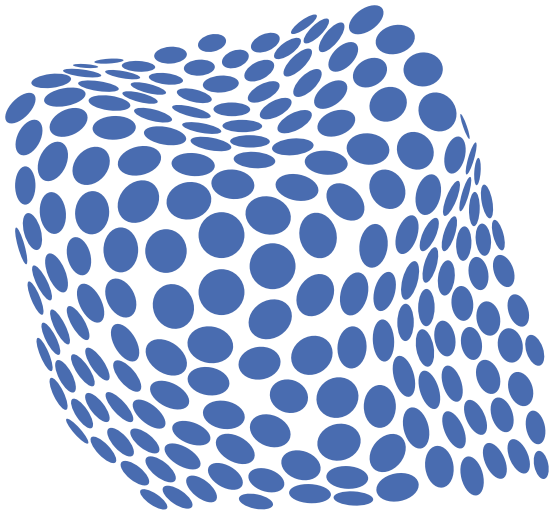


Fig. 5.39 A pattern of optical texture that is perceptually interpreted as a smoothly curved three-dimensional surface. (From Todd et al., 2005, with permission from Elsevier.)

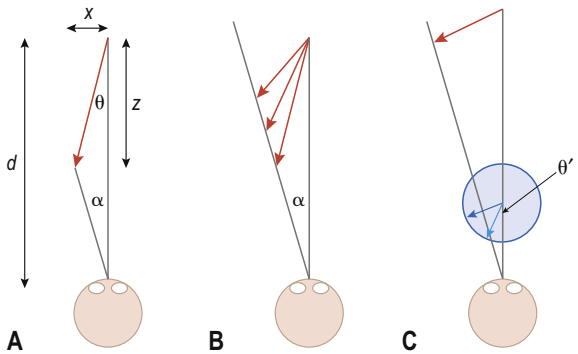


Fig. 5.40 Schematic of stimulus geometry. (**A**) Observer and typical target trajectory *(solid arrow)* at an angle θ to straight ahead are illustrated. Visual direction of the moving target at the end of its motion is given by angle α. (**B**) Many different trajectories correspond to a single visual direction. (**C**) Location of the pointer *(grey circle)*. If observers used θ to set the pointer, they would respond as shown by the *black arrow*. The *grey arrow* shows a typical response if observers used α. (From Harris and Drga, 2005.)

of both bottom-up and top-down processing occurring almost continuously mediated through defined fibre pathways is a central dogma: thus, the dorsal (occipitoparietal) and ventral (occipitoventral) pathways parcel out bottom-up and top-down processing, respectively (Fig. 5.42). The human connectome project (HCP) aimed at mapping the human neuronal connections in health and disease was

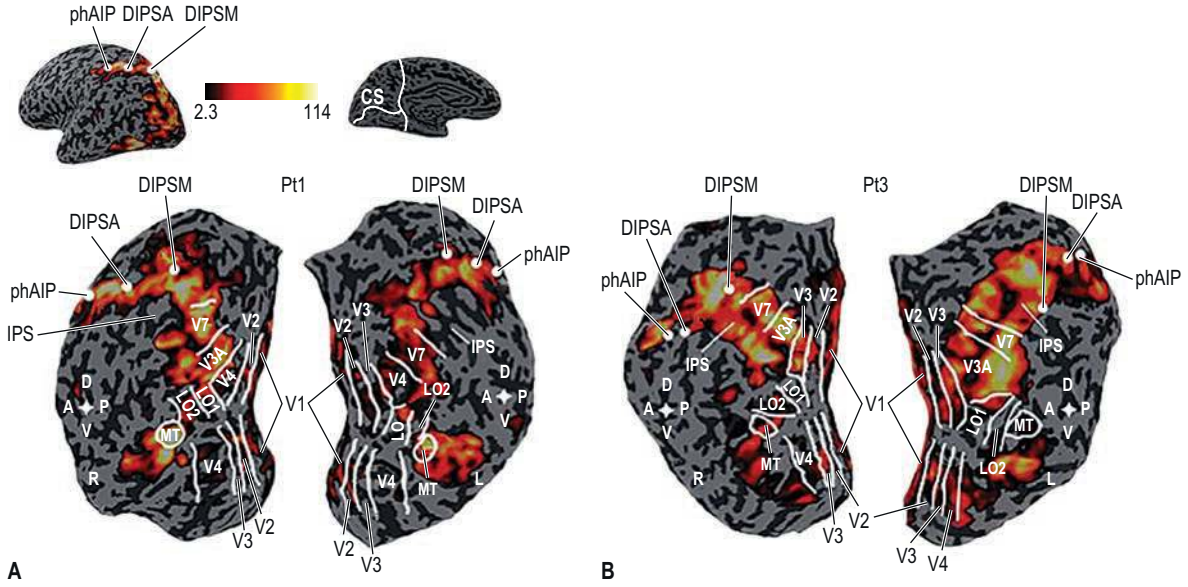


Fig. 5.41 Activation of human cortical responses by stereoscopic depth tasks is widespread, including much of occipital visual cortex and parts of parietal cortex. The figures show heat-map colour scale of z-scored (depth-measurements) activations (two participants in (**A**) and (**B**), respectively) in response to two fields of random dots with a sinusoidal depth profile either side of central fixation point, compared against the response to a flat plane. Subjects were asked to perform a task to judge depth of sinusoidal modulation. The images show anatomical landmarks on the computationally flattened maps of cortex: *R* and *L*, right and left cortical hemisphere; *D, A, V* and *P,* dorsal, anterior, ventral and posterior axial directions; *CS* and *IPS*, calcarine and intraparietal sulci. The identified functional areas are: *V1–V7, hV4, LO1, LO2* and *MT,* visual cortical areas; parietal areas: dorsal intraparietal sulcus anterior *(DIPSA)*, dorsal intraparietal sulcus medial *(DIPSM)* and putative human anterior intraparietal *(phAIP)*. (From Parker et al., 2016.)

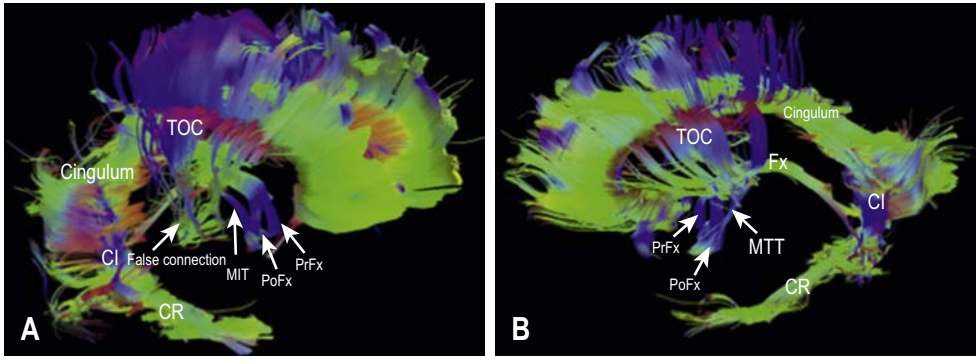


Fig. 5.42 Three-dimensional reconstruction of the Papez circuit also known as the medial limbic circuit which is considered to represent the region controlling emotional experiences. The images are individual results. (**A**) Right side; (**B**) left side. The visualized structures included the fornix, mammillothalamic tract, thalamocingulate connections, and full length of the cingulum, as well as its connections with the frontal and parietal lobes. *CI*, Cingulate isthmus; *CR*, cingulate radiation; *Fx*, fornix; *MTT*, mammillothalamic tract; *PoFx*, post-commissural fornix; *PrFx*, pre-commissural fornix; *TCC*, thalamocingulate connections. (From Wei et al., 2017.)

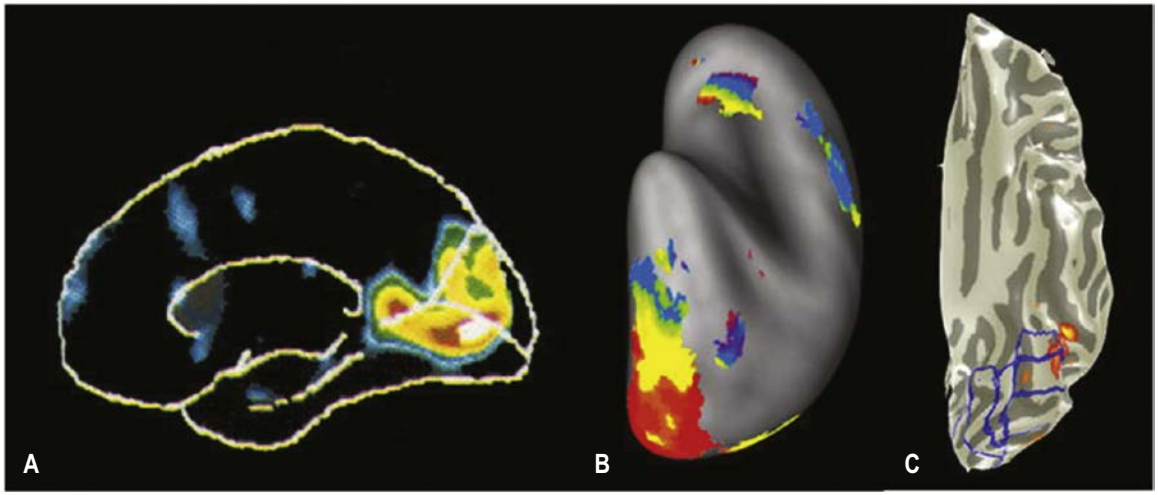


Fig. 5.43 Neuroimaging technology has greatly improved to allow high levels of spatial resolution from positron emission tomography (PET) (**A**), to fMRI on 3D surfaces (**B**), to fMRI data (**C**) (allowing words to be differentiated from phase-scrambled words) coordinated with visual field maps (shown in the *blue outlines*) in individual subjects (right, ventral view) using retinotopic mapping procedures. The word-related activation is near the foveal representation of ventral occipital maps (VO-1 and VO-2); V2 and V3 are also outlined. (From Wandell, 2011.)

launched in 2010 and has made some important observations relating to major psychological disorders; in addition, two groups in the consortia are investigating how cortical connections are disrupted in visual disorders (http://www.humanconnectomeproject.org/).

The construction of the visual image is therefore viewed as arising in anatomically discrete areas of cortex subserving different functions, which are reconstituted through their extensive connections to produce the perception of a final integrated image, for instance during complex psychophysical process such as those involved in ocular dominance (see Fig. 5.36). Thus, an image should not be regarded as being 'impressed' on the retina like the film in a camera, which is then codified to make it 'understood' by the cortex; rather we should appreciate that the processes of sensing (seeing) and cognition (understanding) are not separate but totally integrated. It is likely that, as additional sensory information is added to the image, a higher level of cognition and thus perception can be achieved.

Imaging studies

Much of the early information came from work in animals but advances in imaging the human brain have in many ways now taken precedence. 'Parcellation' in the human visual cortex has been shown by positron emission tomography (PET), magnetoencephalography and functional magnetic resonance imaging (fMRI) (see Fig. 5.43). Several developments in MRI from initial diffusion tensor imaging (DTI) (imaging of molecules as they move) to newer techniques such as diffuse functional MRI, diffusion-weighted imaging and diffusion spectrum imaging have provided the tools for the HCP. Importantly, advances in spatial resolution of fMRI images such that active regions in specific gyri can be determined without interference from image signal of irrelevant neighbouring sites has allowed localization of ventral occipitotemporal pathways involved in, for instance, seeing words (Fig. 5.43). Similar fMRI studies of V4 for coloured stimuli and V5 for moving targets have been done.

In humans, therefore, it can be said that there is a 'colour centre' outside the visual cortex; i.e. that there is functional specialization as in the monkey. Another intriguing observation is that the effect of coloured stimuli is lateralized in humans and not necessarily related to handedness or ocular dominance. Similarly, three different types of motion detection have been located to discrete areas of the visual cortex.

The magnocellular and parvocellular pathways subserve different functions

The striate cortex (area V1) contains the entire map of the retina in a highly ordered and predictable distribution. It receives this information via the LGN, where the neuronal organization is also highly ordered (see Chapter 1, pp. 94–95). At first sight it would seem, therefore, that retinal images should be truly represented in the visual cortex on

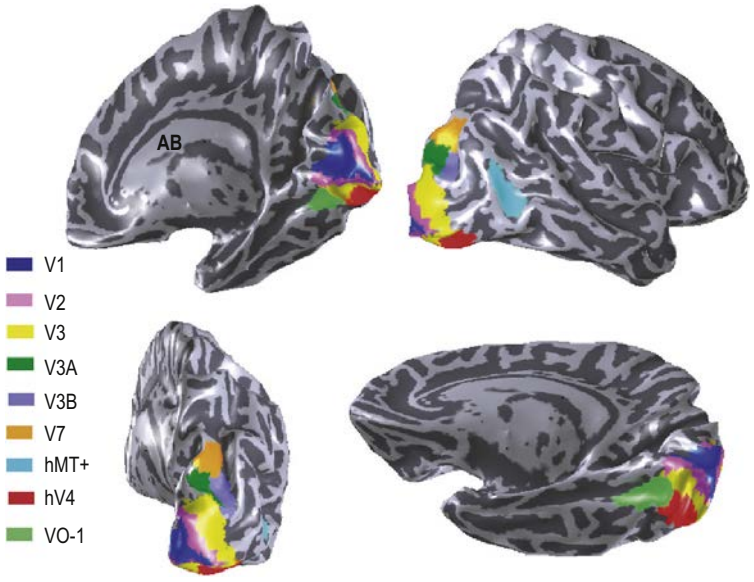


Fig. 5.44 The locations of nine hemifield maps in the human visual cortex. The maps are shown for one typical subject (AB). (From Wandell et al., 2005, with permission from the Royal Society of London.)

the basis of a point-to-point topographical representation. In a very limited sense this is true and may even apply to functional differentiations associated with certain neuronal cell types. Thus, the parvocellular (slow) fibres carry information concerning foveal and parafoveal activity such as spatial discrimination and colour, while the magnocellular (fast) fibres act as transmitters of light detection. This explains in part the high sensitivity that we have for light and motion detection, which are served via the peripheral retina, while contrast and colour detection are slower processes.

However, psychophysical phenomena such as colour and spatial constancy indicate that simple representation of images on the visual cortex in a retinotopic fashion is insufficient to explain the resultant perception. Even the briefest consideration of wavelength discrimination, which is an intrinsically colourless event, despite the fact that there are three discrete receptors, would reveal this truth; thus, a pillar box appears red in most conditions of illumination even though the actual wavelength of the reflected light from the pillar box will vary greatly depending on the light source (but see previous discussion of colour constancy). The perception of colour, and indeed of any visual stimulus, is the result of input from many other cortical sources in addition to the primary visual cortex (see Fig. 5.40). A further unexplained problem in studying cortical and indeed lower levels of activity is the constant high level of neuronal 'noise'. It appears that even in the absence of specific stimulation, there is a significant level of endogenous neuronal activity, which is now presumed to modify output (i.e. perception in the case of vision). This is considered to be one form of top-down modulation of output.

The striate cortex and the prestriate cortex shuffle information between them in the build-up to a perceived image

The striate cortex (area V1) is connected to the prestriate cortex (areas V3–V8) directly and also via area V2 (Fig. 5.44). Each of these areas has one or more specific functions. For example, all cells in area V5 respond to motion in the visual system and are directionally selective (i.e. each cell responds to motion in only one direction); none of these cells, however, is specific for colour. This function is subserved by cells in area V4, in which some of the cells act as wavelength discriminators, but some of these cells also respond to orientation of lines and are involved in shape (form) detection. Other studies have, however, shown that contour information can also be derived from motion detection and that this activity takes place in the primary visual cortex. Cells in V3 and V3A are also selective for form but are indifferent to changes in wavelength.

It is clear, therefore, that colour, orientation, motion, stereoacuity, texture, etc., are all processed separately in areas V3–V8. As areas V3–V8 receive their information from V1, V1 (and V2) must also be functionally specialized. In addition, there must be extensive exchange of information via synaptic connections (the connectome) (Fig. 5.45) and there are small populations of cells that respond to more

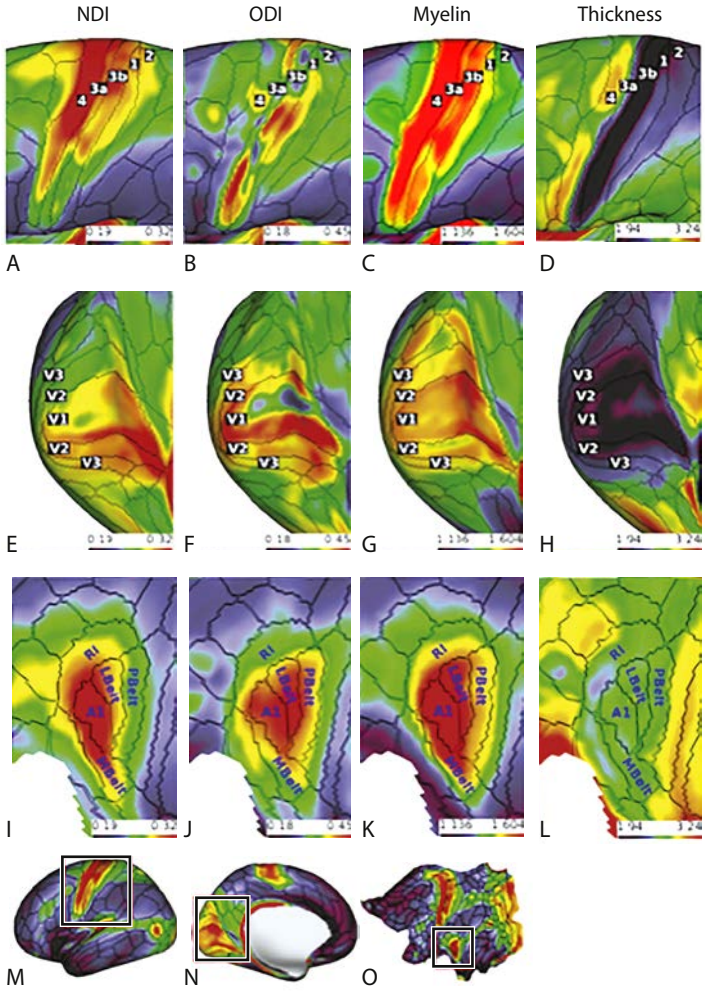


Fig. 5.45 The images show a variety of patterns in which neuritis is distributed in the cerebral cortex. Zoomed views of subject 505 display the neurite density *(NDI)*, the orientation dispersion *(ODI)*, *myelin*, and the *thickness* maps representative areas of early sensory cortices. (**A, B, C** and **D**), The motor somatosensory cortex, (**E, F, G, H**), the visual cortex, and (**I, J, K** and **L**) the auditory cortex. (**M, N, O**) Full-hemisphere views showing myelin maps with boxes outlining expanded regions are shown in images **A**–**L**. *1*, Area 1; *2*, area 2; *3a*, area 3a; *3b*, area 3b; *4*, area 4; *A1*, primary auditory cortex; *LBelt*, lateral belt complex; *MBelt*, medial belt complex; *PBelt*, parabelt complex; *RI*, retro insular cortex; *V1*, primary visual cortex; *V2*, secondary visual area; *V3*, third visual area. (Data at https://balsa.wustl.edu/XX6P. From Fukutomi et al., 2018.)

than one stimulus, such as texture and motion, when these are tested separately.

V2 is likewise organized into areas with thin stripes (for colour detection) and thick stripes (for motion detection) separated by interstripes. Form-selective detectors are present in both the thick and the thin stripes. This form of organization of the visual, and indeed the entire, grey matter into discrete columns of cells responding to specific stimuli has long been known (>50 years) but its functional significance is not clear.

Parcelling out the processing in V1

The concept of functional/anatomical segregation of visual stimuli into components such as colour, motion and orientation detection, depth perception, and other features is now well established (but see later). The cortical sites of other visual tasks, such as texture analysis and shape recognition, are not so easily located. Still others, such as face recognition, involve regions outside the visual and prestriate cortex, such as the anterior temporal gyrus, and sites that store memory such as the hippocampus.

Despite our lack of knowledge, it is still remarkable that such a level of segregation occurs from the retinal ganglion cell input, through the LGN to V1, V2 and V3–V8 in the cortex (Fig 5.46). Segregation may have developed as a result of the different requirements for generating form, colour and motion (e.g. colour compares input from one part of the visual field to another) but topography for colour may be less important. In contrast, precise topographic localization is important for form analysis and motion detection; in the latter, however, this is assessed only transiently.

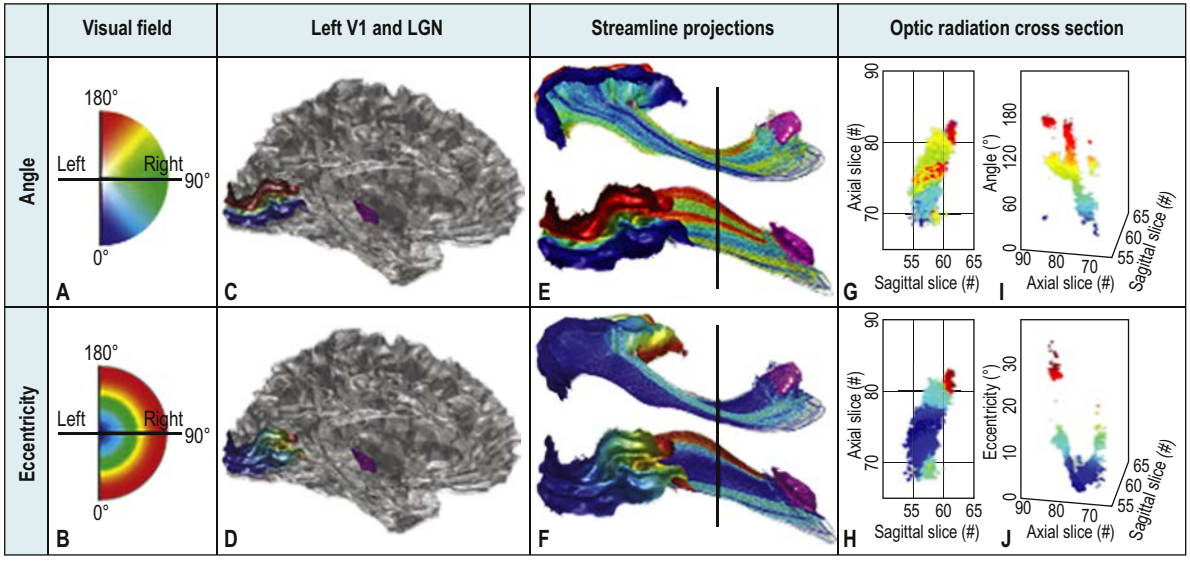


Fig. 5.46 The retinotopic (corresponding retinal and subcortical and cortical sites of visual sensation in the visual field) organization is demonstrated here in a healthy volunteer. (**A** and **B**) show the angle and eccentricity coordinates of the visual field. (**C** and **D**) show LGN in *purple* and the mappings of angle and eccentricity coordinates on V1. (**E** and **F**) show streamlines between LGN and V1. Streamlines are coloured using the corresponding angle and eccentricity values on V1. For clarity, visualizations from inferior and lateral point of views are provided. (**G** and **H**) show the angle and eccentricity values on the cross sections of the optic radiation that are marked with *black lines* in (**E** and **F**). (**I** and **J**) show the 3D scatter plots of angle and eccentricity. Both coordinates vary smoothly. However, the variation of angle is mathematically more complicated than the U-shaped eccentricity. All angle and eccentricity values are to scale and shown using the corresponding colour maps on the left. (From Aydogan et al., 2018.)

Our perception of the external environment may therefore depend on a system of circuits rather like combined and serial parallel processing in computers where there is multistage integration, as Zeki describes it, with feed forwards and feedback control (bottom-up and top-down processing, Zeki, 1990). In line with this concept, perception and comprehension of the visual world occur simultaneously and continuous processing of information, both past and present, is ongoing.

Is the visual cortex organized for hierarchical neural processing or for functional specialization?

The above outline of the organization of visual information at the cortical level may be oversimplistic. In reality, organization of information reception and integration has been considered in two ways, both of which are probably contributory: (1) hierarchical processing of information through the different regions (e.g. sequentially from retinal to LGN to V1, then to V2, and then simultaneously or differentially to V3, V4, V5, etc.); (2) functional specialization in the form of precise topographical localization of aspects of vision to discrete areas of the visual cortex. One of the important recent observations from fMRI work is that the concept of precise retinotopic mapping being restricted to V1 (i.e. representation of the fovea and peripheral retinal regions to precise sites on the striate cortex) may not be accurate. Areas previously considered non-retinotopic such as V4 for colour and V5 for motion also have retinotopic representation although less exact than in V1. Not only is there region-specific cortical representation for instance of colour, but eccentricity maps (distance from fovea) and polar angular maps (angle from the horizontal meridian) cross each other in their cortical representations, allowing a form of mapping of visual space on the cortex (Figs 5.47 and 5.48).

The perception of colour, motion, form, texture and stereopsis therefore all have varying levels of the following attributes: (1) functional specialization in discrete cortical areas; (2) retinotopic representation within each of those areas; and (3) processing in 'streams', e.g. in a colour stream where neural processing for wavelength discrimination takes place at several levels from the retina, to the LGN to V1 and onwards to the specific cortical region in V4. Similar organization underpins motion detection but stereopsis and texture appreciation impinge on several extrastriate areas (see Fig. 5.41). Stereopsis and depth perception are particularly interesting since this requires binocular neurones that will merge signals from right and left eyes: V1 is

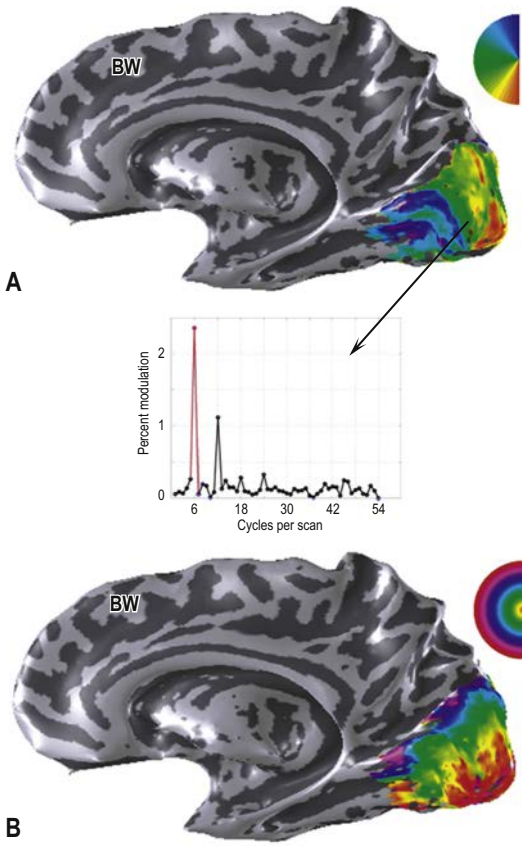


Fig. 5.47 Angular and eccentricity maps near the calcarine cortex. Maps were measured using (**A**) rotating wedges and (**B**) expanding rings comprising contrast-reversing dartboard patterns. The stimuli extended over the central 20 degrees of the visual field and completed six cycles during each experimental scan. The colour overlay indicates the visual field angle (**A**) or eccentricity (**B**) that produces the most powerful response at each cortical location. For clarity, only responses near the calcarine cortex are shown. The graph plots the response amplitude as a function of temporal frequency as measured in a 3 mm radius disc located in the calcarine (see *arrow*). The response is significantly greater at the stimulus repetition frequency (six cycles per scan, shown in *red*) than other temporal frequencies. The secondary peaks at integer multiples of the stimulus frequency are expected and are also significant. The graph is included in the image to provide the reader with an assessment of the reliability of the responses. The stimulus-driven responses shown here are substantially above the statistical threshold ($P < 0.001$, uncorrected). (From Wandell et al., 2005, with permission from the Royal Society of London.)

clearly important, but extrastriate input, e.g. from V2 and V4 as well as other areas, is also important. Broadly, dorsal visual pathways control cross-correlation between signals and ventral pathways address the problem of making multiple signal matches.

There is further sophistication in perceptual pathways. Regions anterodorsal and anteroventral to the striate cortex are not only specialized to detect aspects of vision such as form or colour but also to detect specific objects such as faces, tools, words and even places (Fig. 5.49). Despite this, there is still a strong hierarchical organization for information processing. For instance, the receptive fields for the same stimulus are smallest in V1 and increase progressively through V2, to V3A and V4. To add further complexity to the final percept, there is evidence that top-down processing occurs, for instance activity induced in visual areas V1 and V2 by increased expectation or attention to a region even in the absence of a specific visual object. Even input from emotional stimuli or stereotypical events can modify the visual areas as seen on fMRI.

As fMRI advances with ever better technology, including diffusion tensor MRI, detailed information about the organization of areas such as the intraparietal suture (IPS) have been revealed.

Further information on the intraparietal suture and visual attention is available online at https://expertconsult.inkling.com

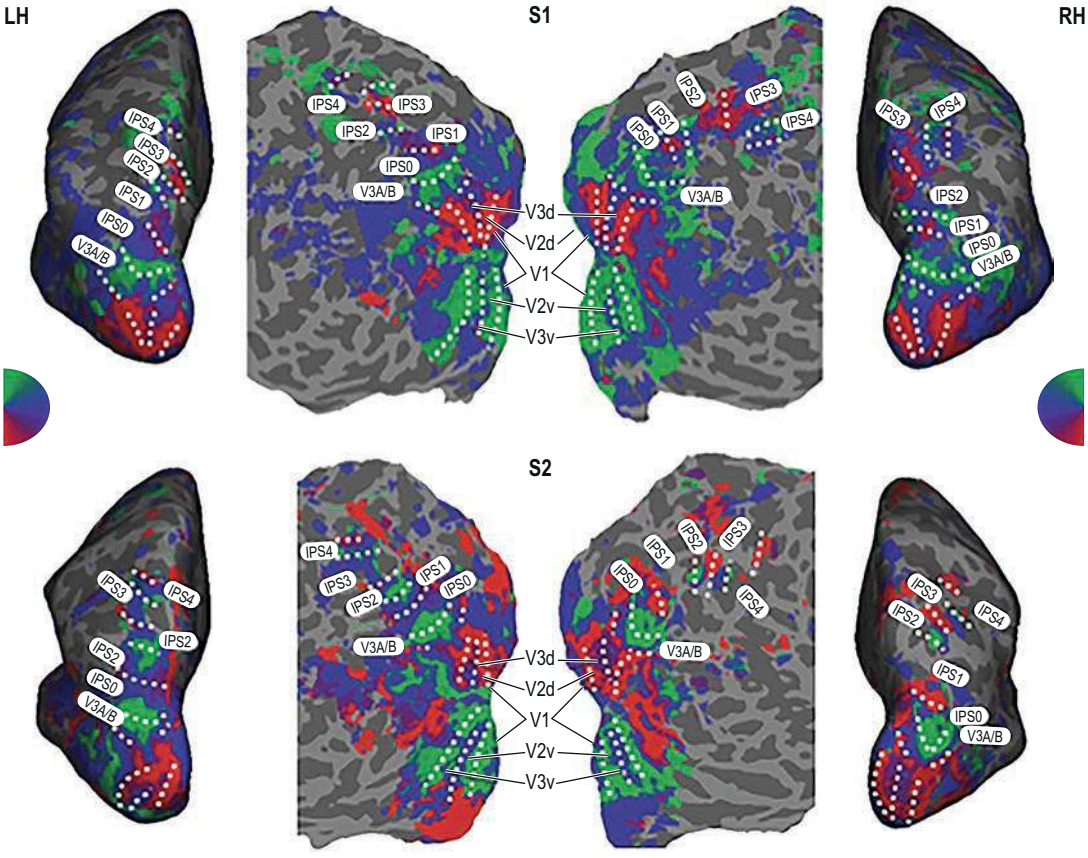
Functional connectivity analyses also showed increased interaction between the IPS and prefrontal regions during manipulation, as well as interhemispheric interactions. Two control tasks demonstrated that covert attention shifts and non-spatial manipulation (arithmetic) engaged patterns of IPS activation and connectivity that were distinct from working memory (WM) manipulation. These findings add to our understanding of the role of IPS in spatial WM maintenance and manipulation. This type of study reveals the visuotopic arrangement of the fibres as they pass through the classically described 'optic radiation' (see Chapter 1, pp. 95–97 and Fig. 5.45) and will be of considerable value in identifying very specific neurological defects as well as informing us on basic neurophysiology.

Area V4 remains a bit of a puzzle. While many functions have been attributed to it, precisely how this region modulates other visual areas such as colour or feature processing are unclear. It has been proposed that the role of V4 is to facilitate neuronal traffic into sets of domain-based networks (such as colour and motion) in two directions, i.e. it probably acts as a hub for the convergence of top-down and bottom-up information (Roe at al., 2012).

Machine learning and artificial intelligence are now being applied to this new knowledge for instance in the development of face recognition technology using Deep Convolutional Neural Networks mirrored on the temporal

THE INTRAPARIETAL SUTURE AND VISUAL ATTENTION

Tasks that involve attention in addition to the maintenance and re-configuration of information in working memory (WM) are centred on the IPS (eFig. 5.10). WM tasks are closely integrated with what is termed the 'contralateral visual space' (i.e. fMRI mappable areas on the contralateral visual cortex) while both sides of the IPS are used for processing this information. In a recent study, Bray and colleagues performed visuotopic mapping in a group of volunteers and showed that a briefly flashed target preferentially engaged the contralateral IPS but when the target was 'mentally' rotated around a circle (i.e. manipulating spatial information) both IPS were activated, and activation was most marked in region IPS1 in most but not all individuals (Bray et al., 2013).



eFig. 5.10 The intraparietal sulcus (IPS) shown in its various regions (IPS0-4) in various views of inflated images of the cortex. Visual cortex areas 1–4 are also shown. *Dashed white lines* represent borders between adjacent areas. (From Bray et al., 2013.)

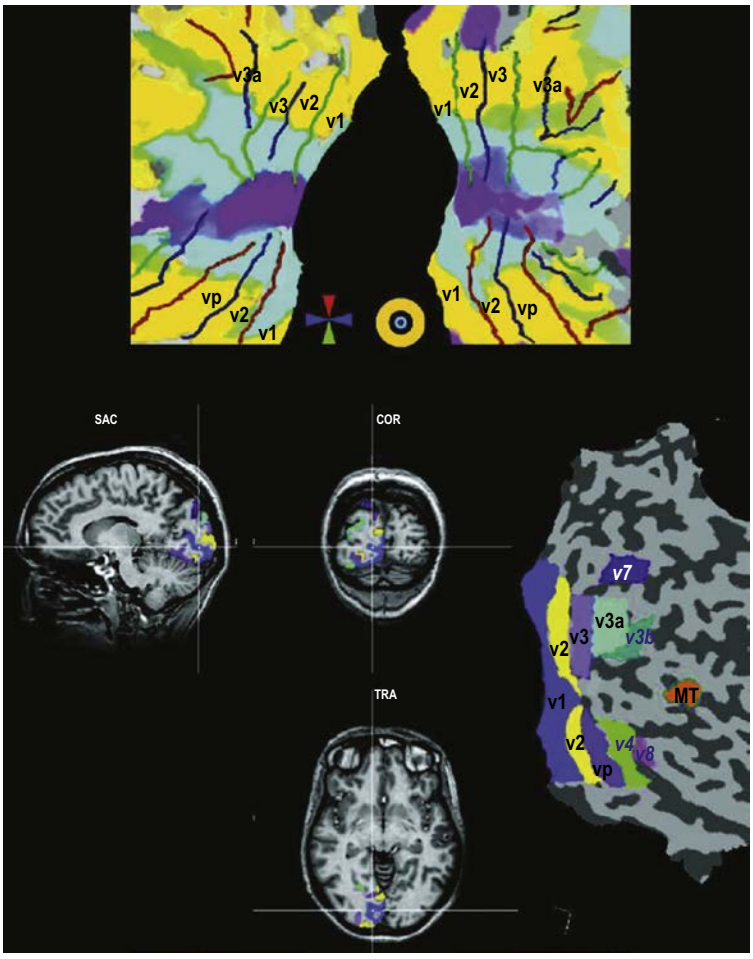


Fig. 5.48 Early and mid-level visual areas. *Top*: Superposition of eccentricity and polar angle maps. *Yellow, blue and pink* bands indicate eccentricity maps; lines indicate centres of upper, lower and horizontal representations (see icons). Note that meridian lines cross all eccentricities orthogonally. *Bottom*: Visual areas on a flattened representation and on the brain volume. Visual area names under consensus are shown in *black*, and areas under debate are marked in *blue italics*. (From Grill-Spector and Malach, 2004, with permission from the publisher of Annual Reviews.)

ventral neural network where much of this activity is sited in the brain (Grill-Spector, 2017).

Physiology of ocular movement

Many of the aspects of the visual response described previously would not be possible without the coordinated movement of the eyes; indeed, eye motion is a fundamental feature of ocular and visual physiology since eyes in the alert state are never at rest. Eye movements are paired even when they move in different directions, as in convergence responses. Neural control of paired eye movements occurs at several levels, as for any neuromuscular event, i.e. at a reflex/subcortical level and via cortical control. The anatomy of the ocular muscles and the innervations of the ocular muscles via the cranial nerves and brainstem nucleus have been reviewed in Chapter 1 (see pp. 67–81).

Types of movement

Uniocular eye movements

Each eye can be moved in the direction of action of the ocular muscles (Box 5.15), which are usually described around a centre of rotation of the globe placed about 14 mm behind the cornea when the eye is in the primary position. Rotation of the eye is either in a vertical (z axis) or a horizontal (x axis) plane, otherwise known as Listing's plane (Listing's law). Torsional motion of the eye occurs around a median or vertical plane through the midline of the skull; this movement can also be described in most circumstances in reference to the retinal horizon (the x, y plane) thus allowing for a modification of Listing's Law to be applied to any initial

position of the eyes, e.g. when the eyes undergo torsional movement as they converge. Rolling movements of the eye occur along an anteroposterior axis and intermediate movements between any of these axes are possible.

Binocular eye movements

The extent of movement of one eye is equal and symmetric to the other (Hering's rule); in conjugate movements the eyes move in parallel and in disjunctive movements (convergence and divergence) they move in opposite directions. In the fusion-free or physiological position of rest (not the primary position of gaze, as this requires fixation on a target) the eyes are slightly divergent.

Conjugate movements require reciprocal innervation of the muscles, which can therefore be described as conjugate pairs of muscles for each direction of gaze (see Box 5.15). This is limited in that the excursion of each muscle is usually greater than that of the pair. The effect is to produce a field of binocular single vision, a parameter of great practical significance

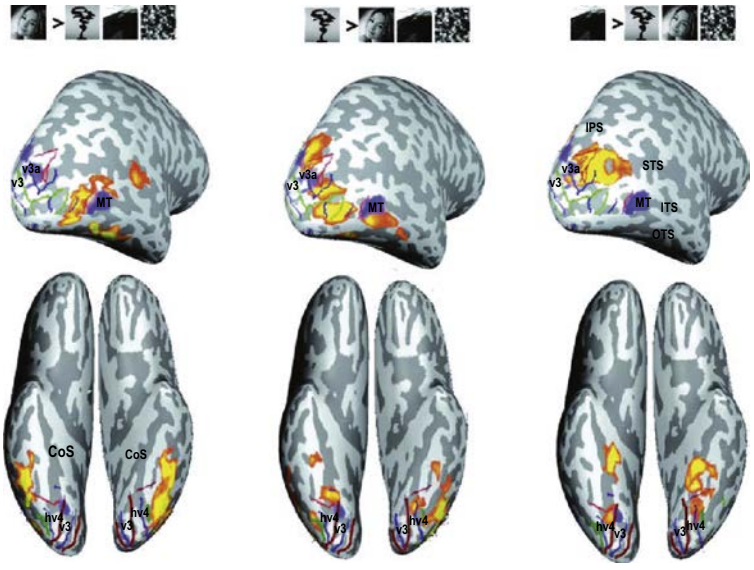
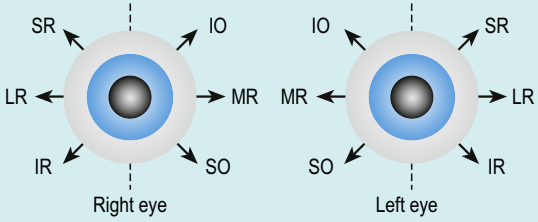


Fig. 5.49 Face-, object-, and place-selective regions in the human brain displayed on an inflated surface representation of the same subject as in Fig. 5.48. Icons indicate the comparisons done in the statistical tests. *Left*: areas responding more strongly to faces than objects, places, or textures. *Centre*: areas responding more strongly to objects than faces, places, or textures. *Right*: areas responding more strongly to places (scenes) than faces, objects, or textures. *Yellow* and *orange* indicate statistical significance: $P < 10^{-12} < P < 10^{-6}$. *Coloured lines* indicate borders of retinotopic visual areas. *Blue* indicates area hMT1, defined as a region in the posterior bank of the inferotemporal sulcus that responds more strongly to moving versus stationary low-contrast gratings (with $P < 10^{-6}$). (From Grill-Spector and Malach, 2004, with permission from the publisher of Annual Reviews.)

Box 5.15 **Evaluation of extraocular muscle function**

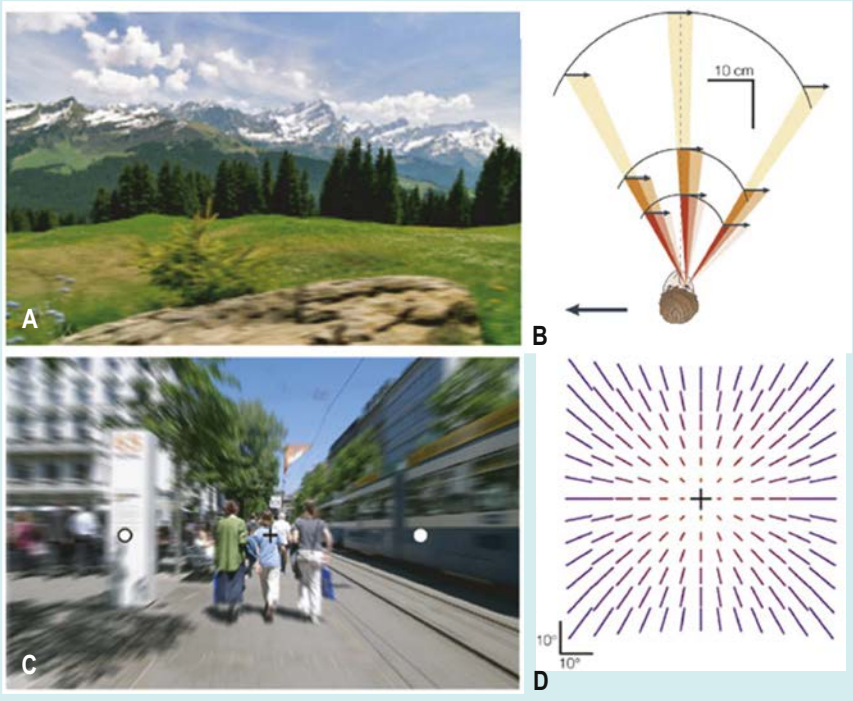
The direction of action of the extraocular muscles is complex and should be considered in three dimensions. The diagram indicates the action of the muscles when they are to be tested clinically for their function.



For instance, the action of the SO muscle is to move the eye down (depression) and out (divergence). However, depression of the eye can also be induced by the IR. Therefore, in order to test the depressor (downward) function of the SO, the simplest way is to test this function when the eye is in a position where the other depressor (the IR) cannot act, i.e. when the eye is adducted. Adducting the eye in effect shortens the IR (the other depressor of the eye, thereby compromising its role as a depressor—muscles are less efficient if their muscle belly is shortened). The SO is thus the only depressor in the adducted position and vice versa for IR. *IO*, Inferior oblique; *IR*, inferior rectus; *LR*, lateral rectus; *MR*, medial rectus; *SO*, superior oblique; *SR*, superior rectus.

Box 5.16 **The image we see is destabilized when we are moving because of optic flow**

Optic flow is a form of visual streaming that occurs as we are moving continuously in one direction. It occurs because the image of the same object(s) is constantly changing with regards to which area of the retina it stimulates. An object of interest is fixed by our gaze and is usually tracked as we go forwards but the eye movement used for this purpose interferes with the flow of information (optic flow) generated by objects in the background, which inform us where we are going and allow us to navigate properly. Accordingly, the object of interest loses focus and becomes blurred (see figure). Both the visuomotor and vestibulomotor systems are at play here and attempt to compensate for the instabilities of retinal images, which typically vary as a function of retinal location and differ for each eye.



(**A**) Blurring of a natural scene image on the retina during simulated leftward self-motion. Near objects blur more than far objects. Without a compensatory eye movement, only objects at optical infinity (for example, mountains on the horizon) remain stable on the retina. (**B**) Idealized patterns of corresponding planar optic flow (in panel **A**). *Small black arrows* indicate the relative displacement of three groups of targets, located at different isovergence distances, during the first 100 milliseconds of constant velocity (50 cm s^{-1}) translation to the left (*large black arrow*). Shaded areas indicate the angular displacements of the targets at different distances and horizontal eccentricities, as seen from the right eye. (**C**) *White circles* indicate two horizontal spatial locations of image blurring on the retina; the *cross* indicates forward movement direction. (**D**) Idealized patterns of corresponding planar optic flow (in panel **C**). Shades from *red to blue* illustrate transition from the fovea to the retinal periphery. Note that these components represent the flow velocity perpendicular to the optic axis.

with respect to standards of normal visual function for the purposes of vehicle license regulations (see Fig. 5.35).

Convergence movements require the combined action of both medial recti; the extent of movement is limited by the near point (5–10 cm from the eyes; this is not affected by age unlike the near point of accommodation) and the far point of convergence (determined in the position of rest as the projected intersection). The converging power of the eye is measured by the metre angle, which depends on the interpupillary distance and can be assessed using graded prisms. From this, the amplitude of convergence can be calculated, which is the difference between the converging power of the eye for the near and far points of convergence. The fusional drive can add a component to the converging power of the

eye: the fusion supplement. This has practical significance, for instance, when estimating the effects of accommodating intraocular lenses on depth of field.

Further information on the dimensions of the idealized eye is available online at: https://expertconsult.inkling.com

Conjugate movements of the eye may be in the form of short sharp movements (saccades) or continuous tracking movements (smooth pursuit). Even when under apparent steady fixation, there are small conjugate movements (microsaccades) as well as other involuntary eye movements such as drift and tremor (see p. 346). Voluntary gaze or 'search' movements (i.e. directed towards non-defined targets) are under higher cortical control (see later). Experimental studies have shown that visual input is required for saccadic movements and is linked to image latency. Tracking movements, however, require visual input plus object speed to be no greater than 30–40 degrees per second and to match that of the eye movement.

Saccades.
- Rapid voluntary relocation of fixation
- Under supranuclear contralateral control
- Latency of 100 ms
- Velocity of 800–1000 degrees/s.

Pursuit.
- Slower tracking movements
- Under supranuclear ipsilateral cortical control
- Latency of 150 ms
- Velocity of 30–50 degrees/s.

During saccades, there is selective suppression of motion detection over other stimuli, suggesting that saccades suppress only the magnocellular pathway. It is also important to consider the concept of optic flow, in which an object moving in relation to a static observer generates a pattern of relative motion in the retinal image. The control of eye movements under these conditions may be difficult to analyse, particularly if the observer is tracking a slowly moving object against a faster moving background.

The situation becomes even more complex if the observer is also moving. Here the visuomotor and vestibulomotor interact with each other to stabilize the image (Box 5.16).

Optic flow is based on what are described as 'short latency' eye movements generated by the sensation of a changing image or scene, experienced by the observer while moving so that correct navigational or directional movement can be made, e.g. 'heading'.

Control of eye movement

The eye muscles in the primary position of gaze are in a state of tonic activity. Each muscle, however, is activated when the eye moves in its field of action and is inhibited in the opposite direction. The final pathway for neuronal control of eye movement occurs via the cranial nerves (see Chapter 1, pp. 67–81), which are the motor neurone equivalent of the spinal nerves subserving reflex responses. As for any muscle, however, the ocular muscles are under both reflex and 'higher centre' control, with the frontal cortex regulating voluntary activity and the occipital cortex and superior colliculus serving as coordinating centres. In addition, there are numerous interneurones and connections with other pathways at the cortical level, e.g. via the paramedian pontine reticular formation (PPRF), and at the reflex level, e.g. the vestibulo-ocular reflex and the cervico-ocular reflex (see earlier section in relation to optic flow). Suppression of frontal cortical activity, which regulates vestibulo-ocular function and optic flow, has been shown in fMRI studies in patients with symptoms of visual vertigo. The generation of horizontal and vertical saccades (gaze) and the fine-tuning of eye movements involve the integrated supranuclear network within the midbrain (PPRF and rostral interstitial nucleus of medial longitudinal fasciculus (riMLF)) and brainstem (vestibulo-ocular and cervico-ocular reflexes), which will be discussed in more detail later.

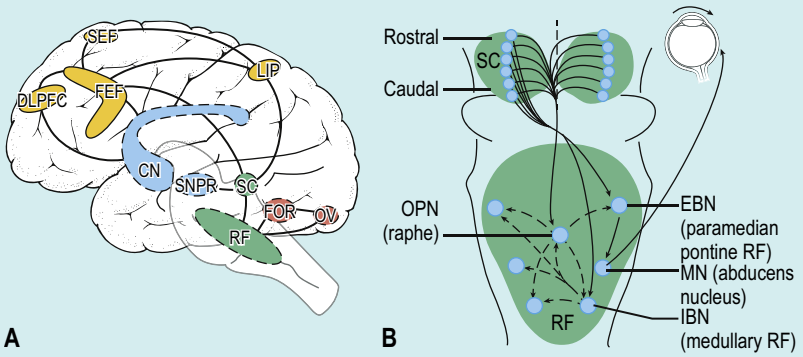
The fixation reflex

The ability to fixate a bright light is a basic reflex that is evident within a few days of birth, but the binocular reflex involving conjugate eye movements and a sustained response takes several months to be fully developed. Foveal fixation is the endpoint of the searching movement of the muscles and may be considered the point of peak activity in the nerve/muscle response. The nerve response can therefore be said to be tuned to foveal fixation. In addition, the very small fine eye movements (microsaccades) that occur with sustained foveal fixation are the result of reflex attempts by the oculomotor centre to achieve the best perceived image, as this falls off rapidly unless a new set of cones is stimulated (Box 5.17).

The fixation reflex can be demonstrated easily by testing for optokinetic nystagmus, where either the stationary subject views a moving scene or a moving subject views a stationary scene. The nystagmus has a slow phase when the eyes follow the target and a fast flick when they readjust to the new target position. The optokinetic nystagmus (OKN) response in humans requires an intact cortex, although there may be a subcortical pathway via the superior colliculus especially for the involuntary searching component of the response (see Box 5.17). Lesions of the cerebral cortex, for example in the temporal lobe, are associated with defects in the OKN response. The nystagmus is preserved in parietal lobe lesions and this test is therefore clinically useful in localizing lesions.

Box 5.17 **Saccades are generated in different areas of the brain**

Saccades describe a fundamental feature of our eyes, i.e. they are never at rest (even when we are sleeping). Even when we fix our gaze on an object, our eyes still continually make small movements termed microsaccades. There is also a tendency for the eyes to 'drift', especially if we are not fixing strongly on an object, and ocular tremor adds a third intrinsic movement to the eyes. These movements are essential to continued perception of an image since without repeated retinal stimulation the image fades. Saccadic eye movements are generated in several areas of the brain, including the lateral intraparietal area, the frontal eye fields, the supplementary eye fields and the dorsolateral prefrontal cortex (see figure). The neural connections are integrated with input from the basal ganglia (*blue* in the figure) and the substantia nigra generate inhibitory impulses which prevent unwanted eye movements as well as control the onset of saccades. (**A**) The brainstem shown in *green* in with higher magnification in (**B**) contains a complex feedback and feedforward circuitry involving 'excitatory burst' and 'omnipause' neurones in the superior colliculus, which control the overall regulation of the microsaccades through inhibitory and activating impulses. CN, cerulean nucleus; DLPFC, dorsolateral prefrontal cortex; EBN, external brachial nucleus; FEF, frontal eye field; FOR, foramen; IBN, internal brachial nucleus; LIP, lateral intraparietal area; MN, marginal nucleus; OPN, oral pontine nucleus; OV, ovale; RF, reticular formation; SC, superior colliculus; SEF, supplementary eye field; SNPR, superior nuclear pontine raphe



Microsaccades are probably as important in achieving a strong fixation reflex as saccades are in searching or exploratory eye movements.

Oculovestibular reflexes are eye movement responses to positional changes in the relationship between the head and the trunk

The vestibular apparatus has structures that convey static head/trunk positional reflex information (i.e. when the subject is not in motion)—the utricle and saccule—and kinetic positional information under conditions of head/trunk acceleration and deceleration—the semi-circular canals. In the utricle and saccule, stimulation of the receptor may occur simply on changing position of the head with respect to gravity, but the ampullae in the semi-circular canals are stimulated via inertial forces in the endolymph surrounding the hair cells (viscous drag). The semi-circular canals are arranged so that they act in synergistic pairs on each side of the head in the *x*, *y* and *z* axes. Vertical and torsional movements involve all four vertical canals. The vestibulo-ocular apparatus is thus of great importance to navigational movements, heading and optic flow (see earlier section). Three-dimensional analysis of ocular movements allows the possible location of defects to, for instance, a single semicircular canal and for curvilinear/optic flow/navigational function has been estimated to be as precise as 0.2 degrees per second (Perrone, 2018). The techniques are based on mathematical models containing information on rotation vectors, reference frames, coordinate systems and Listing's law, and use magnetic search coils in preference to video-based systems.

Listing's law states that, when the head is fixed, the primary position of the eye is such that there is a restricted degree of orientation that can be reached by a single rotation about an axis in Listing's plane (see above) (nine positions of gaze; Box 5.15 and Figs 5.50 and 5.51). Listing's law applies during fixation, saccades, smooth pursuit and vergence movements but not during sleep or during vestibulo-ocular reflexes.

The oculovestibular reflex can be demonstrated by the ability of a rotating observer to maintain fixation on a stationary target by reflex movement of the eyes at the same

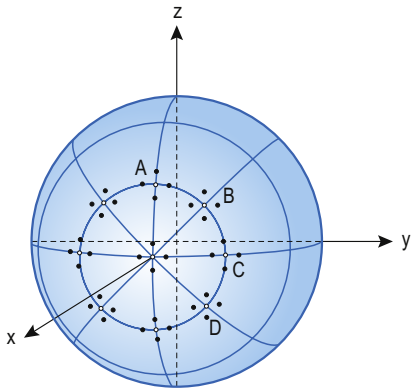


Fig. 5.50 Primary position and Listing's plane. There is a unique orientation of the eye, called 'primary position' or 'primary gaze direction' (direction parallel to the *x*-axis), such that pure vertical and pure horizontal movements that move the eye or gaze line from primary to secondary positions do not change ocular torsion (eye rotations along the respective meridians through *A* or through *C*). Similarly, any movement that rotates the eye/gaze line from primary to tertiary positions on oblique meridian planes does not change torsion (e.g. movements along the meridians through *B* to *D*). The axes of single rotations that move the eye from primary to secondary or tertiary positions lie all in one plane, called Listing's plane (the plane containing the *y* and *z* axes). Tertiary positions cannot be reached from secondary positions by any combination of horizontal and vertical ocular rotations (a torsional component is also needed; see half-angle rule). (From Angelaki and Hess, 2004, with permission from Blackwell Publishing.)

angular rotation (up to 300 degrees/s) as the observer in the opposite direction. In this way, there is stabilization of the retinal image. The reflex can also occur in the dark but is less accurate in its predicted excursion.

Rolling eye movements are due to oculovestibular and oculocervical reflexes

Compensatory eye movements during tilting of the head towards the shoulders initially involve the semi-circular canals but, if the movement is sustained, static (utricle and saccule) responses participate. However, this compensatory movement of the eyes also involves information about neck position from proprioceptors in the neck (oculocervical reflex). Lateral movement of the head about a vertical axis will induce predominantly oculocervical reflexes and movement of the head in the median plane with the eyes fixated produces predominantly oculovestibular reflexes (doll's head movement). Doll's head movements are an important clinical sign to test for intact brainstem reflexes in cases of cortical damage and loss of supranuclear control. Furthermore, video-oculography is becoming widespread

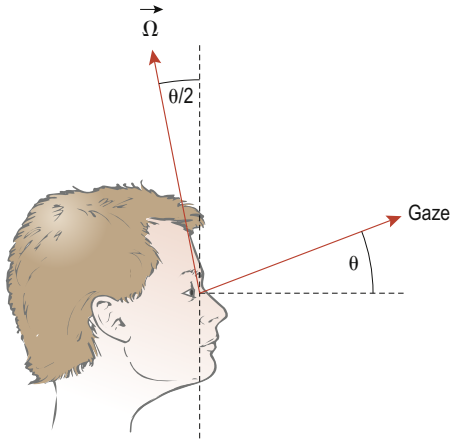


Fig. 5.51 According to Listing's law, the axis of rotation of the eye (Ω) is neither head-fixed nor eye-fixed, but rotates in the same direction as gaze through half the gaze angle (θ/2; the half-angle rule). Thus, at eccentric eye positions, during a horizontal saccade or pursuit eye movement, the axis of rotation of the eye is not purely horizontal *(head-vertical dashed line)* but also has a torsional component *(head-horizontal dashed line)*. (From Angelaki and Hess, 2004, with permission from Blackwell Publishing.)

in the instant assessment of brain injury, particularly concussion as part of sport trauma.

The midbrain is a coordinating centre for reflex eye movement and connects input from multiple sources

Voluntary eye movements (saccades) are initiated in the contralateral motor strip of the frontal cortex (see Chapter 1, p. 99) and pass down to the midbrain via the anterior limb of the internal capsule to synapse in the horizontal gaze centre within the paramedian pontine reticular formation (PPRF) (see Box 5.17; Fig. 5.52). Neurones then pass to the ipsilateral VI nerve and interneurones cross to the opposite medial longitudinal fasciculus to subserve the contralateral III nerve. Within the PPRF are excitatory burst neurones (EBN, see Box 5.17), which have a high but transient rate of discharge (1000 Hz/s) and, when fired, generate the saccade. Normally the burst cells are continuously inhibited by (omni) pause neurones (OPN), until this inhibition is released by discharge from neurones from the frontal eye fields. Burst-pause cells, and other types of cells such as pacemaking cells, are characteristic of different types of neurones in other areas of the brain such as the cerebellum.

Once the saccade has been generated, eye position and fixation are maintained via the tonic neural integrators, also situated within the PPRF. The PPRF also receives inputs from vestibular nuclei, cerebellum, basal ganglia

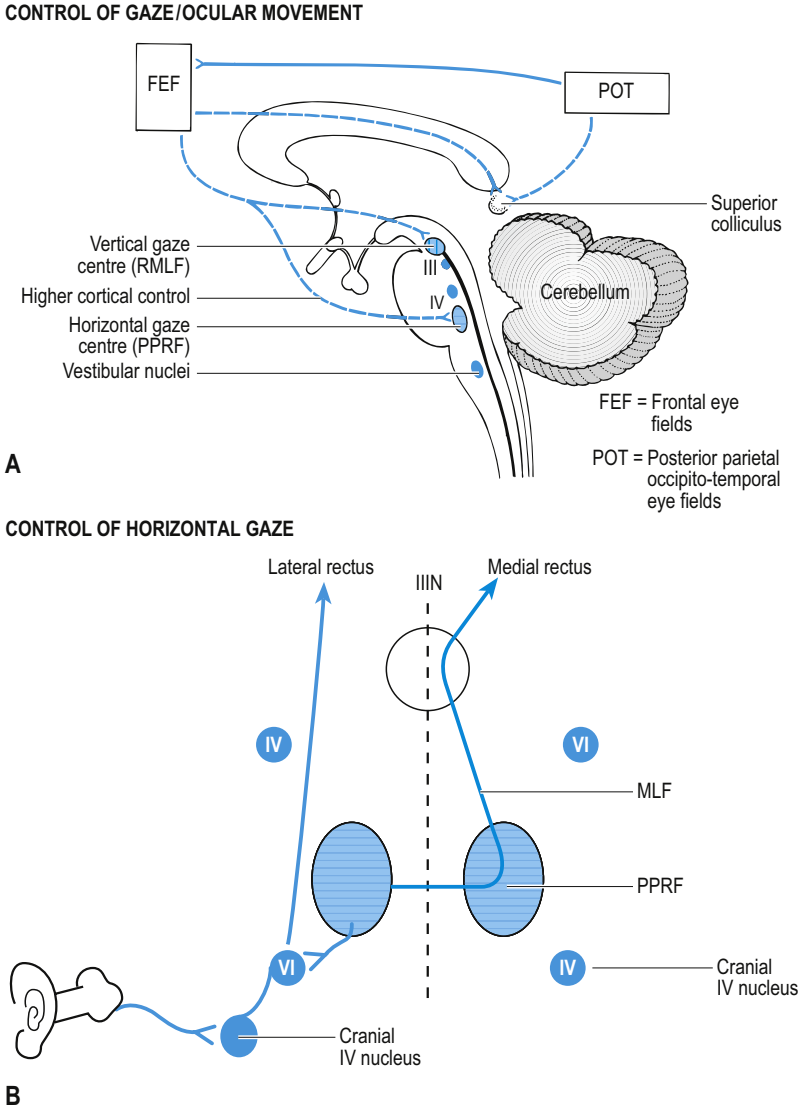


Fig. 5.52 Outline diagrams for integrated control of ocular movements: (**A**) 'higher centre' regulation; (**B**) brainstem, nuclear control. *MLF*, Medial longitudinal fasciculus; *PPRF*, paramedian pontine reticular formation; *RMLF*, reticular medial longitudinal fasciculus.

and cervical proprioceptors, giving rise to fine accurate control of gaze (see later).

The vertical gaze centre is located in the reticular medial longitudinal fasciculus (RMLF), opposite the superior colliculus and above the level of the III nerve nucleus. Unlike the horizontal gaze centre, the vertical gaze centre has no identifiable cortical control, neurones from which cross to the III and IV nerves nuclei to subserve vertical gaze. The medial longitudinal fasciculus, as already mentioned, carries fibres of conjugate horizontal eye movement (involving the VI and III nerves) and also signals for holding vertical eye position, vertical smooth pursuit and vertical vestibulo-ocular reflexes.

Loss of supranuclear control by lesions affecting the midbrain and brainstem can give rise to a variety of clinical features, the commonest being involvement of the reticular medial longitudinal fasciculus (RMLF) in multiple sclerosis, giving rise to abnormal horizontal saccades (internuclear ophthalmoplegia) (Box 5.18).

The superior colliculus is involved in both perception and eye movement control

A small number of fast (M) fibres relay from the retina to the superior colliculus and then to the pulvinar and finally the cortex. These fibres bypass the LGN and are described as the

Box 5.18 **fMRI and detection of discrete brainstem lesions**

fMRIs of patients with gaze palsies and similar oculomotor lesions are revealing discrete lesions in the brainstem. For instance, lesions in the PPRF or the VI nerve nucleus cause lateral gaze palsies, while internuclear ophthalmoplegia is caused by a lesion of the MLF, and the one-and-a-half syndrome is caused by lesions at both sites. Tiny infarcts in the region of the pons correlating with the gaze palsies have been described (reviewed by Bae et al., 2013).

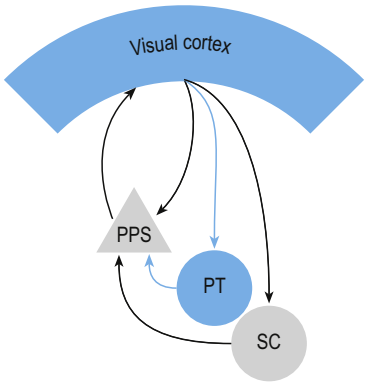


Fig. 5.53 The extrageniculostriate pathway. Neuronal information to the superior colliculus *(SC)* is relayed to the posterior pulvinar system *(PPS)* and on to the visual cortex. From there it relays back to the *PPS*, the *SC* and the pretectal nucleus *(PT)*, which completes the reverberating loop with the *PPS*.

extrageniculostriate pathway. The fibres have crossed chiasmal representation, like optic tract fibres, and synapse in the superior colliculus in retinotopic structured layers, as occurs in fibres to the LGN. The organization, however, is less clearly demarcated, being in broad superficial and deep categories of fibres.

The function of these fibres is not entirely clear. There is evidence of neuronal delay to and from the visual cortex via the posterior pulvinar system of the thalamus and to the pretectal region where the pupillary fibres relay (Fig. 5.53). The cells also have a receptive field organization and a preference for motion detection (a movement field), particularly the rapid reflex locking-on eye movement that occurs in the initiation of tracking a moving target or in automatic scanning during reading. In certain patients with occipital cortex lesions, 'blindsight' (the patient can detect motion or adjudge orientation without perceptually 'seeing' the object) may be present through preservation of this extrageniculostriate pathway.

Cells in the superficial layers of the pulvinar respond to visual input, while those in the deeper layers respond to motion stimuli, although the cells in both layers are in register with each other. The pulvinar also receives many other subcortical inputs and acts as an 'early processing centre', receiving feedback and feedforward information from the cortex and the retina.

Probabilistic tractography (tensor fMRI) studies in humans using a range of 'fearful' stimuli have confirmed the prominence of the retinotopic afferents to the subcortical SC–pulvinar–amygdala pathway in processing fearful stimuli in humans (Koller et al., 2018) although contributions from cortical pathways may also have a role. The SC–pulvinar–amygdala pathway seems to be the danger signal pathway in humans (Fig. 5.54).

Lesions in the pulvinar might thus affect such diverse functions as pattern recognition, eye movement and cerebellar integration in visual responses (see later). The colliculus–pulvinar–cortex relay system, therefore, is a priming system for ocular movement and for reducing errors in the localization response by linking visual and saccadic activity, possibly in response to danger. Recent findings in hemianopic patients with lesions in the pulvinar support this concept.

Cortical centres regulate complex eye movements

Voluntary saccadic gaze movements are initiated by centres in the frontal cortex. Most of the fibres cross the midline in the anterior limb of the internal capsule to end in the gaze centres for motor neurone control of eye movement, but some pass to the ipsilateral superior colliculus where they inhibit automatic gaze responses (see earlier). In addition, cortical efferents remove tonic inhibitory impulses from collicular output to the ocular muscles, which are present between saccades, as if to 'free them up' for full-excursion eye movements.

Cortical voluntary saccades are 'tuned' as for automatic saccades, controlled by the superior colliculus; in the cortex, tuning is broad and appears to depend on recruitment of a precise number of neurones rather than a selected number of highly tuned neurones responsive to motion and visual activity. Single-cell recording studies have also shown that there are cells that exhibit presaccadic activity, while others respond only to the visual stimuli. A third group appears to show complex responses and may be involved in the integration of the response through a direct connection with the PPRF. The frontal eye fields also have an oculomotor loop to the substantia nigra in the basal ganglia, which contains high levels of dopamine. Loss of cells in the substantia nigra and a consequent decrease in dopamine concentration is a characteristic feature of Parkinson disease. This can be

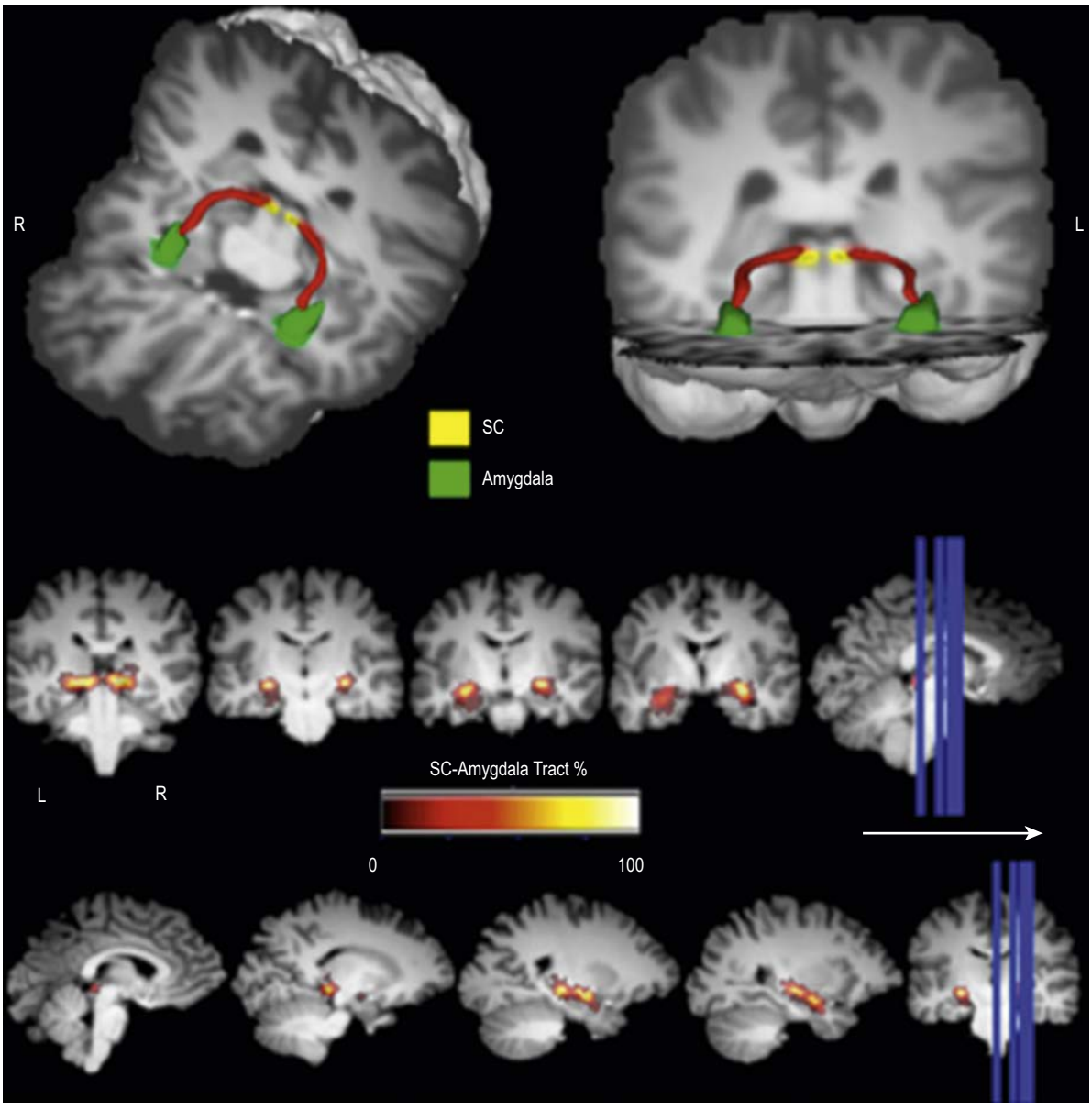


Fig. 5.54 Tractography maps between the amygdala *(A)* and the superior colliculus *(SC)* in areas which register emotions such as fear. *Top*: image shows a 3D reconstruction of the connections. The *middle* (coronal) and *bottom* (sagittal) slices show different views of the connections (pooled data from several individuals). (From Koller, 2019.)

tested using the anti-saccadic task test, in which the urge to fix on objects in the peripheral field is voluntarily suppressed through frontal eye field and superior collicular activity. Patients with this disorder characteristically have difficulties with voluntary gaze movements and the anti-saccadic task.

Tracking, smooth pursuit eye movements are under cortical control through relay of object position information from the occipital cortex to the posterior parietal cortex (motor cortex of smooth pursuit, posterior parietal occipitotemporal regions) and thence to the PPRF. Here the information is integrated with retinal information on object

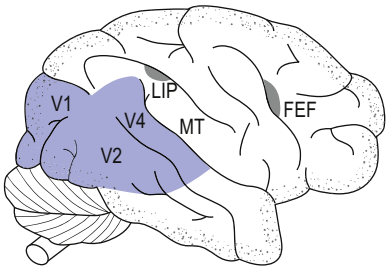


Fig. 5.55 Schematic diagram of primate brain. Areas of the visual cortex activated by visual attention are V1, V2 and V4. The lateral intraparietal areas *(LIP)* and the frontal eye field *(FEF)* determine attention responses in conjunction with the control of eye movements, also served by the superior colliculus (not visible from this view). *MT*, Middle temporal area. (From Bisley, 2011.)

velocity via the optic tract and with impulses from the head and neck on observer position, before being forwarded to the conjugate gaze centre for horizontal eye movement. In addition, recent studies have revealed that the frontal eye fields have control over pursuits as well as saccades.

The temporal cortex, as a centre for relay of motion detection information (see earlier), might be expected to be involved in the control of eye movement. Thus, lesions in this area affect saccades to moving targets but not to stationary ones, while smooth pursuit movements are also impaired.

Aside from obvious visual responses to objects and other specific stimuli, the process of entraining visual attention in deciding what is to be seen is extremely important. Clearly, higher cortical centres are at play here. The role of eye movements in initiating visual attention is paramount and requires integration of information from various regions of the brain, particularly the frontal eye fields (Fig. 5.55). Attention can be both overt (rapid fixating, saccadic eye movement) and covert (selective, 'out of the corner of the eye', not involving eye movement) and each can also be voluntary or involuntary (top-down or bottom-up, respectively). Parieto-frontal and superior collicular pathways are important in determining how attention is directed (see Figs. 5.52 and 5.53). In addition, connectome defects in the brain have been identified in disorders such as attention deficit syndrome and, remarkably, in neurodegenerative disorders such as open angle and normal tension glaucoma.

The cerebellum

Afferent input from the extraocular muscles (the stretch fibres and proprioceptors) is carried in the trigeminal nerve to synapse with cells in the granular cell layer (the Purkinje cell layer; see Chapter 1). There are two, and in some species three, different types of proprioceptor: muscle spindles, Golgi tendon organs and palisade endings, each restricted to the orbital, global and marginal layers of the ocular muscle, respectively. However, some direct cerebellar afferent input is also visual via slit-like, narrow, vertical receptive fields.

The bulk of information connecting the cerebellum with the visual system is transferred via two-way traffic with brainstem centres. The oculomotor cerebellar centre, located in lobules VI and VII, produces saccade-type movements for which Purkinje cells are essential. Input is derived via the PPRF, the vestibular nucleus and the mesencephalic reticular formation. The output from the cerebellum predominantly concerns positional sense and some of it is inhibitory/regulatory. Positional information applies not only to the position of the observer with relation to the object but also to the velocity of the eye movement with relation to the target and the position of the head. Most of this is derived from the vestibular apparatus and not from visual or proprioceptive input.

This combined input also contributes to tracking movements. Experimental data have shown that cells in lobules VI and VII respond with bursts of activity during smooth pursuit movement when the eye is not actually fixating on a target (burst-pause cells).

Compensatory eye movements during movements of the head (e.g. during walking or running) are mediated mostly by the vestibulo-ocular reflex and less so by the cervico-ocular reflex through neck proprioceptors. The effect of these reflexes is to stabilize the retinal image by preventing retinal slip, but the control is imperfect. The perceptual system can cope with a certain amount of retinal slip, but if this is too great (more than 5 degrees/s), symptoms of oscillopsia appear. The cerebellum may contribute to control of the vestibulo-ocular reflex via input from the retina to the flocculus.

It has been also been suggested that the cerebellum is the seat of a control mechanism for integration of information on spatial displacement during eye movement. Lobule VIIb/VIIIa appears to be important in visuospatial cognition. Spatial information coding and visual memory processing appeared to activate different sites in VIIb/VIIIa independently during visual spatial cognition and appeared to be highly specific for attention functions.

Currently, it is considered that there are two processes occurring simultaneously in vestibulo-ocular control of posture and 'body sway': namely, a fast-responding and short-lived response to visual cues such as parallax and a slower developing response to self-motion as it occurs during spontaneous displacement.

Ocular movements during natural activity

Most of the information regarding eye movement has come from studies that were designed to evaluate a particular

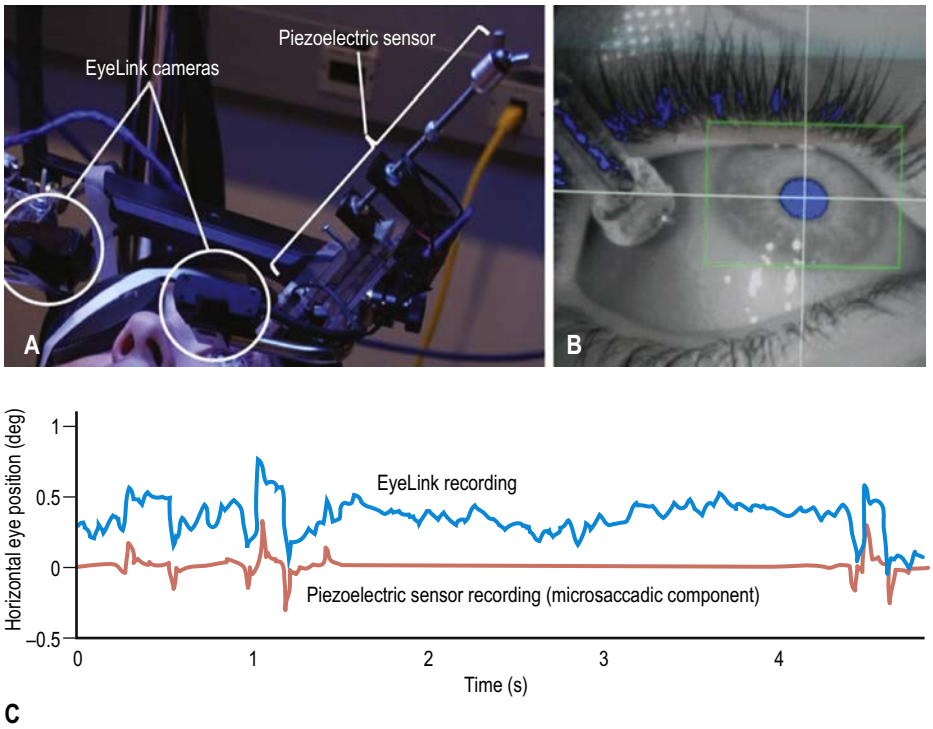


Fig. 5.56 A custom-built set-up for recording simultaneous eye movement. In (**A**) a piezoelectric sensor is mounted to the EyeLink II helmet, while the image in (**B**) is a close-up of the sensor on the eye. The subject's pupil *(blue)* was tracked easily and data collected (**C**), showing microsaccade detection as small changes in eye position in degrees. (From McCamy et al., 2013.)

movement, e.g. saccades or smooth pursuit movements. Investigation of eye movements during normal activities using infrared eye tracking devices has revealed the complex pattern of eye movements involved in performing day-to-day tasks and has shown how extensively higher cortical information guides eye movements. Eye tracking devices have been developed that allow differentiation of microsaccades from even finer movements, termed ocular microtremors (OMTs) (Fig. 5.56). This has shown that microsaccades rather than OMTs underpin image persistence (prevent perceptual fade) during fixation (see p. 341).

Neural versus mechanical control of eye movement?

A continuing issue in studies of gaze control is to understand how much regulation of muscle function is mediated via neural control and how much can be attributed to mechanical effects of the muscles. Gaze control involves rotational three-dimensional movements (which obey Listing's law when the head is fixed but not when the head is moving), plus other movements such as OKN and those generated via the vestibulo-ocular reflex, to achieve image stabilization. Not all movements under control of the vestibulo-ocular reflex fail to obey Listing's law, such as the rotational movement that occurs in response to head movement (see p. 341), but the majority do. Movements that do not obey Listing's law, i.e. gaze shifts when the head is not restrained, obey Donder's law: for each position of gaze there is only one three-dimensional orientation (torsional movement).

Anatomical studies have suggested that the surrounding muscle sheath with the check ligaments can act like a pulley, allowing fine changes in the pulling direction related to the degree of torsional rotation. The data from these studies are also consistent with Listing's law and have formed the basis of some aspects of management of patients after strabismus surgery using adjustable sutures. However, the major, if not the sole, control of muscle behaviour rests with neural elements at both nuclear and supranuclear levels, with considerable regulation coming from higher centres such as the frontal eye fields and the cerebro-cerebellar network (see previously). Moreover, some of the neurophysiological control would be consistent with a pulley mechanism while others such as saccades, which also obey Listing's law, are not. What is clear is that both mechanical and neural

mechanisms regulate motor activity, but what remains to be determined is how the three-dimensional perception of space regulates the final motor command (top-down control) as well as the role of visual attention.

Conclusion

Visual neurophysiology continues to be the cornerstone of psychophysical neurophysiology, but it is under increasing challenge from advances in imaging such as diffusion tensor MRI as well as very recent information on high-resolution mapping of the human brain (The Human Connectome Project http://www.humanconnectomeproject.org/; The Organization for Human Brain Mapping http://www.humanbrainmapping.org). Increasingly, our understanding of what constitutes a visual image becomes more sophisticated. Research in this field is extremely active since there are many areas of uncertainty and, possibly more importantly, many potential applications of this knowledge to clinical medicine.

The primate visual system is a highly complex arrangement for analysing information concerning the external world derived from a wide array of possible signals, all of which are captured by the retinal sensory receptors. It is remarkable to consider that around 30% of all sensory information to the brain comes via the visual system. This information is integrated with input from many other sensory systems and stored information from past experience (memory). The final image and its interpretation (perception) are extensively edited by the brain to ensure normality and constancy wherever possible. However, this image is, of course, unique to each individual, despite the fact that we ascribe common definitions to familiar objects. Continuing research into the psychophysics of vision will fuel our thirst for knowledge into what makes us tick.

Further reading and additional references are available online at: https://expertconsult .inkling.com

References

Anstis, S., 2018. The role of the pupil, corneal reflex, and iris in determining the perceived direction of gaze. Iperception 9, 2041669518765852, https://doi.org/10.1177/2041669518765852.

Aydogan, D.B., Shi, Y., 2018. Tracking and validation techniques for topographically organized tractography. Neuroimage 181, 64–84. https://doi.org/10.1016/j.neuroimage.2018.06.071.

Babai, N., et al., 2016. Functional roles of complexin 3 and complexin 4 at mouse photoreceptor ribbon synapses. J. Neurosci. 36, 6651–6667. https://doi.org/10.1523/JNEUROSCI.4335-15.2016.

Baker, D.H., Lygo, F.A., Meese, T.S., Georgeson, M.A., 2018. Binocular summation revisited: Beyond radical2. Psychol Bull 144, 1186–1199. https://doi.org/10.1037/bul0000163.

Buena-Atienza, E., Nasser, F., Kohl, S., Wissinger, B.A.73, 2018. 128 bp de novo deletion encompassing the OPN1LW/OPN1MW gene cluster in sporadic Blue Cone Monochromacy: a case report. BMC Med. Genet. 19, 107. https://doi.org/10.1186/s12881-018-0623-8.

Field, G.D., Uzzell, V., Chichilnisky, E.J., Rieke, F., 2019. Temporal resolution of single-photon responses in primate rod photoreceptors and limits imposed by cellular noise. J. Neurophysiol. 121, 255–268. https://doi.org/10.1152/jn.00683.2018.

Frishman, L., et al., 2018. (2018) ISCEV extended protocol for the photopic negative response (PhNR) of the full-field electroretinogram. Doc. Ophthalmol. 136 (3), 207–211. Published online 2018 May 31. doi: 10.1007/s10633-018-9638-x PMCID: PMC6061118 PMID: 29855761.

Fukutomi, H., et al., 2018. Neurite imaging reveals microstructural variations in human cerebral cortical gray matter. Neuroimage 182, 488–499. https://doi.org/10.1016/j.neuroimage.2018.02.017.

Gorea, A., Tyler, C.W., 2013. Dips and bumps: on Bloch's law and the Broca-Sulzer phenomenon. Proc Natl Acad Sci U S A 110, E1330. https://doi.org/110.1073/pnas.1221807110.

Katayama, K., et al., 2019. Specificity of the chromophore-binding site in human cone opsins. J. Biol. Chem. https://doi.org/10.1074/jbc.RA119.007587.

Kim, Y.J., Reynaud, A., Hess, R.F., and Mullen, K.T., 2017. A Normative Data Set for the Clinical Assessment of Achromatic and Chromatic Contrast Sensitivity Using a qCSF Approach. Invest Ophthalmol Vis Sci 58, 3628–3636. https://doi.org/10.1167/iovs.17-21645.

Koller, K., Rafal, R.D., Platt, A., Mitchell, N.D., 2018. Orienting toward threat: contributions of a subcortical pathway transmitting retinal afferents to the amygdala via the superior colliculus and pulvinar. Neuropsychologia. https://doi.org/10.1016/j.neuropsychologia.2018.01.027.

Koller, K., Rafal, R.D., Platt, A., Mitchell, N.D., 2019. Orienting toward threat: contributions of a subcortical pathway transmitting retinal afferents to the amygdala via the superior colliculus and pulvinar. Neuropsychologia 128, 78–86. https://doi.org/10.1016/j.neuropsychologia.2018.01.027.

Maguire, J., Parry, N.R., Kremers, J., Murray, I.J., and McKeefry, D., 2017. The morphology of human rod ERGs obtained by silent substitution stimulation. Doc Ophthalmol 134, 11–24. https://doi.org/10.1007/s10633-017-9571-4.

Martinez-Conde, S., Macknik, S.L., 2017. Unchanging visions: the effects and limitations of ocular stillness. Philos. Trans. R. Soc. Lond. B Biol. Sci. 372. https://doi.org/10.1098/rstb.2016.0204.

Masland, R.H., 2017. Vision: two speeds in the retina. Curr. Biol. 27, R303–R305. https://doi.org/10.1016/j.cub.2017.02.056.

Molina-Martin, A., Perez-Cambrodi, R.J., Pinero, D.P., 2018. Current clinical application of microperimetry: a review. Semin. Ophthalmol. 33, 620–628. https://doi.org/10.1080/08820538.2017.1375125.

Mooney, S.W.J., et al., 2018. Curveball: a tool for rapid measurement of contrast sensitivity based on smooth eye movements. J. Vis. 18, 7. https://doi.org/10.1167/18.12.7.
Palchikova, I.G., Smirnov, E.S., Palchikov, E.I., 2018. Quantization noise as a determinant for color thresholds in machine vision. J. Opt. Soc. Am. A Opt. Image Sci. Vis. 35, B214–B222. https://doi.org/10.1364/JOSAA.35.00B214.
Parker, A.J., Smith, J.E., Krug, K., 2016. Neural architectures for stereo vision. Philos. Trans. R. Soc. Lond. B Biol. Sci. 371. https://doi.org/10.1098/rstb.2015.0261.
Perrone, J.A., 2018. Visual-vestibular estimation of the body's curvilinear motion through the world: a computational model. J. Vis. 18, 1. https://doi.org/10.1167/18.4.1.
Pons, C., et al., 2017. Neuronal mechanisms underlying differences in spatial resolution between darks and lights in human vision. J. Vis. 17, 5. https://doi.org/10.1167/17.14.5.
Rogerson, L.E., et al., 2017. Connectomics of synaptic microcircuits: lessons from the outer retina. J. Physiol. 595, 5517–5524. https://doi.org/10.1113/JP273671.
Sabel, B.A., Flammer, J., Merabet, L.B., 2018. Residual vision activation and the brain-eye-vascular triad: Dysregulation, plasticity and restoration in low vision and blindness - a review. Restor. Neurol. Neurosci. 36, 767–791. https://doi.org/10.3233/RNN-180880.
Spitschan, M., et al., 2017. The human visual cortex response to melanopsin-directed stimulation is accompanied by a distinct perceptual experience. Proc. Natl. Acad. Sci. U. S. A. 114, 12291–12296. https://doi.org/10.1073/pnas.1711522114.
Wei, W., 2018. Neural mechanisms of motion processing in the mammalian retina. Annu. Rev. Vis. Sci 4, 165–192. https://doi.org/10.1146/annurev-vision-091517-034048.
Wei, P.H., et al., 2017. In vivo visualization of connections among revised Papez circuit hubs using full q-space diffusion spectrum imaging tractography. Neuroscience 357, 400–410. https://doi.org/10.1016/j.neuroscience.2017.04.003.
Werblin, F.S., 2010. Six different roles for crossover inhibition in the retina: correcting the nonlinearities of synaptic transmission. Vis. Neurosci. 27, 1–8. https://doi.org/10.1017/S0952523810000076.
Wood, S., Loudon, A., 2018. The pars tuberalis: the site of the circannual clock in mammals? Gen. Comp. Endocrinol. 258, 222–235. https://doi.org/10.1016/j.ygcen.2017.06.029.
Zele, A.J., et al., 2017. A temporal white noise analysis for Extracting the impulse response function of the human electroretinogram. Transl. Vis. Sci. Technol. 6, 1. https://doi.org/10.1167/tvst.6.6.1.

Further reading

Angelaki, D.E., Hess, B.J., 2004. Control of eye orientation: where does the brain's role end and the muscle's begin? Eur. J. Neurosci. 19, 1–10. https://doi.org/10.1111/j.1460-9568.2004.03068.x.

Angelaki, D.E., Hess, B.J., 2005. Self-motion-induced eye movements: effects on visual acuity and navigation. Nat. Rev. Neurosci. 6, 966–976.

Anton-Erxleben, K., Carrasco, M., 2013. Attentional enhancement of spatial resolution: linking behavioural and neurophysiological evidence. Nat. Rev. Neurosci. 14, 188–200.

Bae, Y.J., Kim, J.H., Choi, B.S., Jung, C., Kim, E., 2013. Brainstem pathways for horizontal eye movement: pathologic correlation with MR imaging. Radiographics 33, 47–59.

Barlow, H.B., 1972. Single units and sensation: a neuron doctrine for perceptual psychology? Perception 1, 371–394.

Bergstrom, S.S., 2004. The AMBEGUJAS phenomenon and colour constancy. Perception 33, 831–835.

Bergstrom, S.S., Gustafsson, K.A., 2011. Concave and convex phases in ambiguous figures showing colour shifts. Mach's figure and the AMBEGUJAS phenomenon. Perception 40, 30–38.

Bertamini, M., Wagemans, J., 2013. Processing convexity and concavity along a 2-D contour: figure-ground, structural shape, and attention. Psychon. Bull. Rev. 20, 191–207.

Bisley, J.W., 2011. The neural basis of visual attention. J. Physiol. 589, 49–57.

Blake, R., Wilson, H., 2011. Binocular vision. Vision. Res. 51, 754–770.

Bloomfield, S.A., Volgyi, B., 2009. The diverse functional roles and regulation of neuronal gap junctions in the retina. Nat. Rev. Neurosci. 10, 495–506.

Boynton, G.M., 2002. Color vision: how the cortex represents color. Curr. Biol. 12, R838–R840.

Braunstein, R.E., Sparrow, J.R., 2005. A blue-blocking intraocular lens should be used in cataract surgery. ArChapter. Ophthalmol. 123, 547–549.

Bray, S., Almas, R., Arnold, A.E., Iaria, G., Macqueen, G., 2013. Intraparietal sulcus activity and functional connectivity supporting spatial working memory manipulation. Cereb. Cortex [Epub ahead of print].

Bridge, H., 2011. Mapping the visual brain: how and why. Eye 25, 291–296.

Cangiano, L., Dell'Orco, D., 2013. Detecting single photons: a supramolecular matter? FEBS Lett. 587, 1–4.

Chapot, C.A., Euler, T., Schubert, T., 2017. How do horizontal cells 'talk' to cone photoreceptors? Different levels of complexity at the cone-horizontal cell synapse. J. Physiol. 595, 5495–5506. https://doi.org/10.1113/JP274177.

Collin, S.P., 2008. A web-based archive for topographic maps of retinal cell distribution in vertebrates. Clin. Exp. Optom. 91, 85–95.

Constable, P.A., Ngo, D., Quinn, S., Thompson, D., 2017. A. A meta-analysis of clinical electro-oculography values. Doc. Ophthalmol. 135, 219–232. https://doi.org/10.1007/s10633-017-9616-8.

Conway, B.R., 2009. Color vision, cones, and color-coding in the cortex. Neuroscientist 15, 274–290.

Cropper, S.J., Derrington, A.M., 1996. Rapid colour-specific detection of motion in human vision. Nature 379, 72–74.

Cumming, B., 1997. Stereopsis: how the brain sees depth. Curr. Biol. 7, R645–R647. https://doi.org/10.1016/s0960-9822(06)00324-1.

D'Angelo, E., Mazzarello, P., Prestori, F., Mapelli, J., Solinas, S., Lombardo, P., et al, 2011. The cerebellar network: from structure to function and dynamics. Brain Res. Rev. 66, 5–15.

Demb, J.B., Singer, J.H., 2012. Intrinsic properties and functional circuitry of the AII amacrine cell. Vis. Neurosci. 29, 51–60.

Field, G.D., Rieke, F., 2002. Mechanisms regulating variability of the single photon responses of mammalian rod photoreceptors. Neuron 35, 733–747.

Field, G.D., Rieke, F., 2002. Nonlinear signal transfer from mouse rods to bipolar cells and implications for visual sensitivity. Neuron 34, 773–785.

Freeman, E.D., Sterzer, P., Driver, J., 2012. fMRI correlates of subjective reversals in ambiguous structure-from-motion. J. Vis. 12, 35.

Gainotti, G., 2013. Is the right anterior temporal variant of prosopagnosia a form of 'associative prosopagnosia' or a form of 'multimodal person recognition disorder'? Neuropsychol. Rev. 23, 99–110.

Gilbert, C.D., Li, W., 2013. Top-down influences on visual processing. Nat. Rev. Neurosci. 14, 350–363.

Goossens, L., Schruers, K., Peeters, R., Griez, E., Sunaert, S., 2007. Visual presentation of phobic stimuli: amygdala activation via an extrageniculostriate pathway? Psychiatry Res. 155, 113–120.

Grill-Spector, K., Malach, R., 2004. The human visual cortex. Annu. Rev. Neurosci. 27, 649–677. https://doi.org/10.1146/annurev.neuro.27.070203.144220.

Grill-Spector, K., Weiner, K.S., Kay, K., Gomez, J., 2017. The functional neuroanatomy of human face perception. Annu. Rev. Vis. Sci 3, 167–196. https://doi.org/10.1146/annurev-vision-102016-061214.

Grimes, W.N., 2012. Amacrine cell-mediated input to bipolar cells: variations on a common mechanistic theme. Vis. Neurosci. 29, 41–49.

Grossberg, S., Yazdanbakhsh, A., Cao, Y., Swaminathan, G., 2008. How does binocular rivalry emerge from cortical mechanisms of 3-D vision? Vis. Res. 48, 2232–2250.

Harris, J.M., Drga, V.F., 2005 Feb. Using visual direction in three-dimensional motion perception. Nat. Neurosci. 8 (2), 229–233 Epub 2005 Jan 23.

Harris, K.D., Thiele, A., 2011. Cortical state and attention. Nat. Rev. Neurosci. 12, 509–523.

Hartveit, E., Veruki, M.L., 2012. Electrical synapses between AII amacrine cells in the retina: function and modulation. Brain Res. 1487, 160–172.

Hayter, E.A., Brown, T.M., 2018. Additive contributions of melanopsin and both cone types provide broadband sensitivity to mouse pupil control. BMC. Biol. 16, 83. https://doi.org/10.1186/s12915-018-0552-1.

Hecht, E., 1987. Optics, second ed. Addison Wesley, Reading, Massachusetts, USA.

Hirasawa, H., Yamada, M., Kaneko, A., 2012. Acidification of the synaptic cleft of cone photoreceptor terminal controls the amount of transmitter release, thereby forming the receptive field surround in the vertebrate retina. J. Physiol. Sci. 62, 359–375.

Horiguchi, H., Winawer, J., Dougherty, R.F., Wandell, B.A., 2013. Human trichromacy revisited. Proc. Natl. Acad. Sci. U.S.A. 110, E260–E269.

Hurlbert, A.C., Derrington, A.M., 1993. How many neurons does it take to see? Curr. Biol. 3, 510–512. https://doi.org/10.1016/0960-9822(93)90042-m.
Kim, Y.J., Reynaud, A., Hess, R.F., Mullen, K.T., 2017. A normative data set for the clinical assessment of achromatic and chromatic contrast sensitivity using a qCSF approach. Invest. Ophthalmol. Vis. Sci. 58, 3628–3636. https://doi.org/10.1167/iovs.17-21645.
Koch, K.W., Dell'orco, D., 2013. A calcium-relay mechanism in vertebrate phototransduction. ACS Chem. Neurosci. 4, 909–917.
Koenig, D., Hofer, H., 2011. The absolute threshold of cone vision. J. Vis. 11 pii: 21.
Korf, H.W., 2018. Signaling pathways to and from the hypophysial pars tuberalis, an important center for the control of seasonal rhythms. Gen. Comp. Endocrinol. 258, 236–243. https://doi.org/10.1016/j.ygcen.2017.05.011.
Kravitz, D.J., Saleem, K.S., Baker, C.I., Mishkin, M., 2011. A new neural framework for visuospatial processing. Nat. Rev. Neurosci. 12, 217–230.
Kremers, J., Rodrigues, A.R., Silveira, L.C., da Silva Filho, M., 2010. Flicker ERGs representing chromaticity and luminance signals. Invest. Ophthalmol. Vis. Sci. 51, 577–587.
La Morgia, C., Ross-Cisneros, F.N., Sadun, A.A., Hannibal, J., Munarini, A., Mantovani, V., et al, 2010. Melanopsin retinal ganglion cells are resistant to neurodegeneration in mitochondrial optic neuropathies. Brain 133, 2426–2438.
Marmor, M.F., Brigell, M.G., McCulloch, D.L., Westall, C.A., Bach, M., 2011. ISCEV standard for clinical electro-oculography (2010 update). Doc. Ophthalmol. 122, 1–7.
Martinez-Conde, S., Macknik, S.L., Troncoso, X.G., Hubel, D.H., 2009. Microsaccades: a neurophysiological analysis. Trends Neurosci. 32, 463–475.
Martinez-Conde, S., Otero-Millan, J., Macknik, S.L., 2013. The impact of microsaccades on vision: towards a unified theory of saccadic function. Nat. Rev. Neurosci. 14, 83–96.
Masland, R.H., 2011. Cell populations of the retina: the Proctor lecture. Invest. Ophthalmol. Vis. Sci. 52, 4581–4591.
Masland, R.H., 2012c. Another blue neuron in the retina. Nat. Neurosci. 15, 930–931.
Masland, R.H., 2012a. The neuronal organization of the retina. Neuron 76, 266–280.
Masland, R.H., 2012b. The tasks of amacrine cells. Vis. Neurosci. 29, 3–9.
McCamy, M.B., Collins, N., Otero-Millan, J., Al-Kalbani, M., Macknik, S.L., Coakley, D., et al, 2013. Simultaneous recordings of ocular microtremor and microsaccades with a piezoelectric sensor and a video-oculography system. PeerJ 1, e14.
Mendola, J.D., Buckthought, A., 2013. fMRI investigation of monocular pattern rivalry. J. Cogn. Neurosci. 25, 62.
Noudoost, B., Moore, T., 2011. Control of visual cortical signals by prefrontal dopamine. Nature 474, 372–375.
Op de Beeck, H.P., Haushofer, J., Kanwisher, N.G., 2008. Interpreting fMRI data: maps, modules and dimensions. Nat. Rev. Neurosci. 9, 123–135.
Palczewski, K., 2012. Chemistry and biology of vision. J. Biol. Chem. 287, 1612–1619.
Park, J.C., McAnany, J.J., 2015. Effect of stimulus size and luminance on the rod-, cone-, and melanopsin-mediated pupillary light reflex. J. Vis. 15. https://doi.org/10.1167/15.3.13.
Parker, A.J., 2007. Binocular depth perception and the cerebral cortex. Nat. Rev. Neurosci. 8, 379–391.
Peelen, M.V., Downing, P.E., 2007. The neural basis of visual body perception. Nat. Rev. Neurosci. 8, 636–648.
Pelli, D.G., Bex, P., 2013. Measuring contrast sensitivity. Vision Res 90, 10–14.
Perrone, J.A., 2018. Visual-vestibular estimation of the body's curvilinear motion through the world: a computational model. J. Vis. 18, 1. https://doi.org/10.1167/18.4.1.
Piccolino, M., Wade, N.J., 2008. Galileo Galilei's vision of the senses. Trends Neurosci. 31, 585–590.
Provis, J.M., Dubis, A.M., Maddess, T., Carroll, J., 2013. Adaptation of the central retina for high acuity vision: cones, the fovea and the avascular zone. Prog. Retin. Eye Res. 35, 63–81.
Reuter, T., 2011. Fifty years of dark adaptation 1961–2011. Vision Res 51, 2243–2262.
Rieiro, H., Martinez-Conde, S., Danielson, A.P., Pardo-Vazquez, J.L., Srivastava, N., Macknik, S.L., 2012. Optimizing the temporal dynamics of light to human perception. Proc. Natl. Acad. Sci. U.S.A. 109, 19828–19833.
Roe, A.W., Chelazzi, L., Connor, C.E., Conway, B.R., Fujita, I., Gallant, J.L., et al., 2012. Toward a unified theory of visual area V4. Neuron 74, 12–29.
Rossion, B., Hanseeuw, B., Dricot, L., 2012. Defining face perception areas in the human brain: a large-scale factorial fMRI face localizer analysis. Brain Cogn. 79, 138–157.
Salomon, R.G., Hong, L., Hollyfield, J.G., 2011. Discovery of carboxyethylpyrroles (CEPs): critical insights into AMD, autism, cancer, and wound healing from basic research on the chemistry of oxidized phospholipids. Chem. Res. Toxicol. 24, 1803–1816.
Sand, A., Schmidt, T.M., Kofuji, P., 2012. Diverse types of ganglion cell photoreceptors in the mammalian retina. Prog. Retin. Eye Res. 31, 287–302.
Sanes, J.R., Masland, R.H., 2015. The types of retinal ganglion cells: current status and implications for neuronal classification. Annu. Rev. Neurosci. 38, 221–246. https://doi.org/10.1146/annurev-neuro-071714-034120.
Sasaki, Y., Nanez, J.E., Watanabe, T., 2010. Advances in visual perceptual learning and plasticity. Nat. Rev. Neurosci. 11, 53–60.
Schmidt, T.M., Do, M.T., Dacey, D., Lucas, R., Hattar, S., Matynia, A., 2011. Melanopsin-positive intrinsically photosensitive retinal ganglion cells: from form to function. J. Neurosci. 31, 16094–16101.
Schwartz, O., Hsu, A., Dayan, P., 2007. Space and time in visual context. Nat. Rev. Neurosci. 8, 522–535.
Schwiegerling, J., 2000. Theoretical limits to visual performance. Surv. Ophthalmol. 45, 139–146. https://doi.org/10.1016/s0039-6257(00)00145-4.
Solomon, S.G., Lennie, P., 2007. The machinery of colour vision. Nat. Rev. Neurosci. 8, 276–286.
Spillmann, L., 2009. Phenomenology and neurophysiological correlations: two approaches to perception research. Vis. Res. 49, 1507–1521.
Sterling, P., 2013. Some principles of retinal design: the Proctor lecture. Invest. Ophthalmol. Vis. Sci. 54, 2267–2275.
Sung, C.H., Chuang, J.Z., 2010. The cell biology of vision. J. Cell Biol. 190, 953–960.
Tanimoto, N., Brombas, A., Muller, F., Seeliger, M.W., 2012. HCN1 channels significantly shape retinal photoresponses. Adv. Exp. Med. Biol. 723, 807–812.
Taylor, W.R., Smith, R.G., 2012. The role of starburst amacrine cells in visual signal processing. Vis. Neurosci. 29, 73–81.
Temple, S.E., 2011. Why different regions of the retina have different spectral sensitivities: a review of mechanisms and functional significance of intraretinal variability in spectral sensitivity in vertebrates. Vis. Neurosci. 28, 281–293.

Todd, J.T., Thaler, L., Dijkstra, T.M., 2005. The effects of field of view on the perception of 3D slant from texture. Vis. Res. 45, 1501–1517. https://doi.org/10.1016/j.visres.2005.01.003.
Toscani, M., Valsecchi, M., Gegenfurtner, K.R., 2013. Optimal sampling of visual information for lightness judgments. Proc. Natl. Acad. Sci. U.S.A. 110, 11163–11168.
Ts'o, D.Y., Zarella, M., Burkitt, G., 2009. Whither the hypercolumn? J. Physiol. 587, 2791–2805.
van Pelt, S., Fries, P., 2013. Visual stimulus eccentricity affects human gamma peak frequency. Neuroimage 78C, 439–447.
Vaney, D.I., Sivyer, B., Taylor, W.R., 2012. Direction selectivity in the retina: symmetry and asymmetry in structure and function. Nat. Rev. Neurosci. 13, 194–208.
Verdon-Roe, G.M., Westcott, M.C., Viswanathan, A.C., Fitzke, F.W., Garway-Heath, D.F., 2006. Exploration of the psychophysics of a motion displacement hyperacuity stimulus. Invest. Ophthalmol. Vis. Sci. 47, 4847–4855.
Vos, J.J., 2006. From lower to higher colour metrics: a historical account. Clin. Exp. Optom. 89, 348–360.
Wandell, B.A., 2011. The neurobiological basis of seeing words. Ann. N. Y. Acad. Sci. 1224, 63–80.
Wandell, B.A., Smirnakis, S.M., 2009. Plasticity and stability of visual field maps in adult primary visual cortex. Nat. Rev. Neurosci. 10, 873–884.
Wandell, B.A., Brewer, A.A., Dougherty, R.F., 2005. Visual field map clusters in human cortex. Philos. Trans. R. Soc. Lond. B Biol. Sci. 360, 693–707. https://doi.org/10.1098/rstb.2005.1628.
Watson, A.B., Ahumada, A.J., 2011. Blur clarified: a review and synthesis of blur discrimination. J. Vis. 11 pii: 10.
Wedeen, V.J., Rosene, D.L., Wang, R., Dai, G., Mortazavi, F., Hagmann, P., et al., 2012. The geometric structure of the brain fiber pathways. Science 335, 1628–1634.
Wei, W., Feller, M.B., 2011. Organization and development of direction-selective circuits in the retina. Trends. Neurosci. 34, 638–645.
Westheimer, G., 2009. The third dimension in the primary visual cortex. J. Physiol. 587, 2807–2816.
Westheimer, G., 2009. Visual acuity: information theory, retinal image structure and resolution thresholds. Prog. Retin. Eye Res. 28, 178–186.
Westheimer, G., 2011. Three-dimensional displays and stereo vision. Proc. Biol. Sci. 278, 2241–2248.
Westheimer, G., 2012. Optical superresolution and visual hyperacuity. Prog. Retin. Eye Res. 31, 467–480.
Wexler, M., 2003. Voluntary head movement and allocentric perception of space. Psychol. Sci. 14, 340–346. https://doi.org/10.1111/1467-9280.14491.
Wexler, M., van Boxtel, J.J., 2005. Depth perception by the active observer. Trends Cogn. Sci 9, 431–438. https://doi.org/10.1016/j.tics.2005.06.018.
Wiechmann, A.F., Summers, J.A., 2008. Circadian rhythms in the eye: the physiological significance of melatonin receptors in ocular tissues. Prog. Ret. Eye. Res 27, 137–160.
Wilson, C.R., Gaffan, D., Browning, P.G., Baxter, M.G., 2010. Functional localization within the prefrontal cortex: missing the forest for the trees? Trends Neurosci. 33, 533–540.
Wise, S.P., 2008. Forward frontal fields: phylogeny and fundamental function. Trends Neurosci. 31, 599–608.
Wong, A.M., 2004. Listing's law: clinical significance and implications for neural control. Surv. Ophthalmol. 49, 563–575.
Wood, S., Loudon, A., 2018. The pars tuberalis: the site of the circannual clock in mammals? Gen. Comp. Endocrinol. 258, 222–235. https://doi.org/10.1016/j.ygcen.2017.06.029.
Wu, J., Yan, T., Zhang, Z., Jin, F., Guo, Q., 2012. Retinotopic mapping of the peripheral visual field to human visual cortex by functional magnetic resonance imaging. Hum. Brain Mapp. 33, 1727–1740.
Zeki, S., 1990. Parallelism and functional specialization in human visual cortex. Cold. Spring. Harb. Symp. Quant. Biol 55, 651–661. https://doi.org/10.1101/sqb.1990.055.01.062.
Zeki, S., 1992. The visual image in mind and brain. Sci. Am. 267, 68–76. https://doi.org/10.1038/scientificamerican0992-68.
Zeki, S., 2011. The mind's eye. Cerebrum 2011, 4.
Zhang, H., Maloney, L.T., 2012. Ubiquitous log odds: a common representation of probability and frequency distortion in perception, action, and cognition. Front. Neurosci. 6, 1.

Chapter 6

General and ocular pharmacology

CONTENTS

Introduction
Pharmacokinetics: drug trafficking in the body
- Basic concepts
- Drug absorption
- Drug absorption is dependent on a drug's lipid solubility
- Drug distribution
- Drugs are metabolized to facilitate clearance
- Metabolized drugs are excreted in urine and bile

Pharmacodynamics: drug handling by the body
Drug–receptor interactions
- Ion channels
- Second messengers
- Calcium is a major mediator of cellular activity
- Regulation of protein synthesis

Ocular pharmacology: drug handling by cells and tissues of the eye
- Mechanisms of ocular drug absorption
- Factors influencing delivery of drugs to the eye

Delivery methods of ocular medication
- Residence in the conjunctival sac
- Drug vehicles affect drug delivery
- Advanced ocular delivery systems
- A gene therapy approach to delivering treatment
- Intracameral and intravitreal administration
- Drugs administered systemically also penetrate the eye
- Topical medications and preservatives

Reconstituting the tear film
- Tear substitutes
- Mucolytics

Ocular drugs and the autonomic nervous system
- The parasympathetic system
- The sympathetic system

Clinical control of intraocular pressure exploits the adrenergic system
A drug-free approach to treating glaucoma
The histaminergic system: histamine is released from conjunctival mast cells during allergic reactions
- Antihistamines
- Sodium cromoglicate
- Local anaesthetics have both local and systemic side-effects

Eicosanoids affect multiple ocular functions
Serotonin: a potent neurotransmitter
- Serotonin receptor subclasses mediate different effects
- 5-HT antagonists

Glucocorticoids
Immunosuppressant agents: combating ocular inflammatory disease
- Biologics

Local anaesthetics: an integral part of ophthalmic examination and surgery
- Local anaesthetics affect excitable membranes
- Local anaesthetics have both local and systemic side-effects

Ocular toxicity from systemic administration of drugs
- Uveal tract
- Photosensitization

Introduction

This chapter covers the basic principles of clinical pharmacology, with particular reference to drugs used in the management of ophthalmic disorders, methods of ocular drug delivery, and the interactions of drugs and the eye. Although it is generally thought that the medical management of ocular disease is mainly administered through topical therapeutic agents, many systemic drugs and agents

are also used. These include diuretics for the control of intraocular pressure, immunosuppressants for control of intraocular inflammatory conditions and antimicrobials for control of infection. Therefore, basic pharmacological principles (pharmacokinetics and pharmacodynamics) are important to ophthalmologists. The basic pharmacology of systemic therapy, including receptor–drug interactions, is reviewed before discussing the more specific topical and systemic therapies in the treatment of ophthalmic disorders.

Pharmacokinetics: drug trafficking in the body

Basic concepts

Pharmacokinetics is the mathematical study of the time-course of drug absorption, distribution, metabolism and excretion. One of the simplest parameters to consider when discussing the pharmacokinetics of any drug is the biological *half-life* of a drug ($t_{1/2}$), which is the time taken for the plasma drug concentration to fall by half after administration. A more accurate method of assessing the efficiency of drug elimination is the estimation of drug clearance from the circulation. For example, after intravenous administration, the $t_{1/2}$ of the drug may be calculated from its plasma concentration–time curve. This simplistic model is based on a single compartment model that states that, following intravenous administration, the distribution of a drug assumes a uniform concentration throughout all compartments (intracellular and extracellular); the elimination by both metabolism and renal excretion is also assumed to be directly proportional to the drug concentration. If this is the case, the volume of distribution may be calculated as:

$$V_d = \text{dose} \;/\; C_0$$

where C_0 is the estimated (plasma) concentration at the time of injection.

The apparent volume of distribution is defined as the volume of fluid required to contain the total amount of drug in the body at the same concentration as that present in plasma. Drugs in general may be confined to the plasma compartment because they are either too large to cross the capillary wall or are highly protein-bound (e.g. heparin and warfarin, respectively). Drugs may also be distributed only to the extracellular compartment because they have low lipid solubility (e.g. gentamicin) or distributed throughout all aqueous compartments if they are lipid-soluble. When both the volume of distribution and renal clearance of a drug are known, the $t_{1/2}$ may be estimated. However, when considering a single compartment model, no account is taken of distribution into tissue compartments and metabolism within tissue compartments. When repeated injections of the drug are given, the plasma concentration becomes a function of the rate of both elimination and administration, and the plasma concentration will equilibrate when these two parameters are equal. A steady-state plasma concentration is therefore reached, which in practice may be established after an interval of about three or four plasma half-life lives of the drug given. To reach a steady-state concentration more quickly, loading doses of the drug are often given (for example, antibiotics and warfarin) (Fig. 6.1). In clinical practice, the $t_{1/2}$ is important because it will determine the frequency of administration of a drug. If continued intravenous boluses are given, there are frequently large peaks and troughs in drug concentration, which can lead to a greater incidence of toxic side-effects.

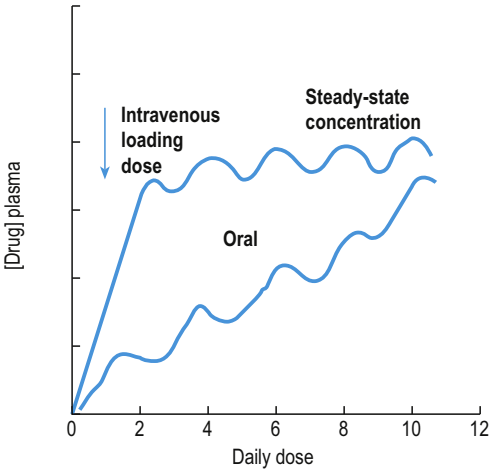


Fig. 6.1 Plasma concentration of drugs after oral and intravenous administration.

Bioavailability describes the amount of oral dose that reaches the systemic circulation and becomes available to the site of drug action. However, this blanket term is not sufficiently precise because rapidly absorbed drugs will reach a much higher plasma concentration than those absorbed slowly and, similarly, rapid elimination would also theoretically lead to low bioavailability. Bioavailability of a drug is measured as the area under the curve of log plasma concentration against time for both intravenous and oral administration, although the reliability of quantitative drug assessment may be variable for several reasons. For instance, the degree of bioavailability may be altered by incomplete absorption of the drug or destruction of the drug by *first-pass metabolism* before the drug reaches the plasma compartment, irrespective of the rate of absorption from the gastrointestinal tract. Drugs instilled topically to the eye are also absorbed from the nasal and nasopharyngeal mucosae directly into the systemic circulation. As such, they escape first-pass metabolism and have a high bioavailability. Thus,

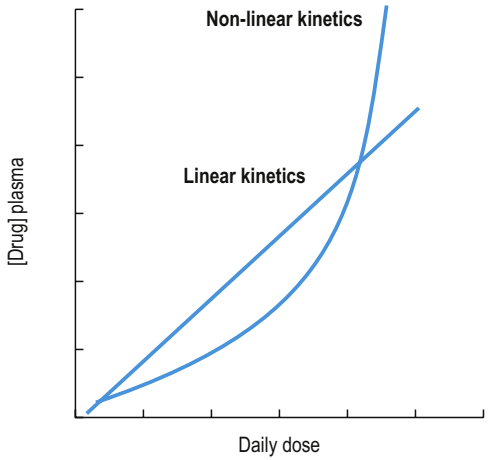


Fig. 6.2 First-order and zero-order pharmacokinetics.

topically administered agents can give rise to quite marked systemic effects (see later). Modification of topical delivery allowing access to the rich conjunctival area and network of transport processes (e.g. amino acids, nucleosides and dipeptides) also allows possibility of access to the posterior segment. However, obstruction of lymphatic clearance, conjunctival blood flow clearance and tear dilution means formulation is paramount for optimal effect.

Drug kinetics can be described as first-order (linear) or zero-order (non-linear, saturation) kinetics (Fig. 6.2). First-order kinetics describes a process where rate is proportional to the amount of drug present and can be defined by linear differential equations. Zero-order kinetics occurs when drug dynamics show *saturation* at high drug concentrations. Saturation can occur, for example, when the capacity of drug-metabolizing liver enzymes is surpassed and there is a wide distribution and tissue depot, leading to unmetabolized drug in the circulation for longer periods. The duration of action of a drug that exhibits saturation kinetics is more dependent on the administering dose than in drugs that exhibit first-order kinetics. Also, there is no direct relationship between drug dose and steady-state plasma concentration in zero-order kinetics, which may explain sudden unexpected drug toxicity in a number of clinical settings. For this reason, close drug monitoring is required (e.g. phenytoin).

Drug absorption

For a drug to reach the site at which it produces its effect, it must first be absorbed from its site of administration. In general, drug penetration of cell membranes increases with lipid solubility. Drugs can cross cell membranes by diffusing directly through lipid, by diffusion through aqueous pores that traverse the lipid, by utilizing carrier molecules, or by pinocytosis by the cell.

The rate of passage (diffusion) through a cell membrane can be predicted by Fick's law

Non-polar substances dissolve freely in lipid and therefore penetrate the cell by diffusion. The lipid solubility, degree of ionization and molecular size of the drug will determine its diffusion coefficient. The rate is determined by Fick's law, which states that the rate at which drugs cross a biological membrane is directly proportional to the concentration gradient across the membrane and the diffusion coefficient, and inversely proportional to the cell membrane thickness (see later):

$$\text{Rate of diffusion} = KA\,(x_1 - x_2) \;/\; D$$

where K is the diffusion constant, A is the diffusion area, $x_1 - x_2$ is the concentration difference between plasma and intracellular compartments and D is the thickness of the membrane.

Active transport

As for active transport of ions (see Chapter 4, p. 166), active transport of large polar drugs requires energy-dependent carrier-mediated mechanisms. These transport systems may be disrupted by inhibiting enzyme-dependent carriers or by blocking the carrier mechanism with structurally similar drugs (analogues). Facilitated transport is carrier-mediated transport that does not require energy because it does not proceed against a concentration gradient, for example glucose transport into erythrocytes.

Drug absorption is dependent on a drug's lipid solubility

Drug properties

The absorption of a drug depends on its lipid solubility and inversely on its polarity or degree of ionization. An important factor in the degree of penetration of a drug through membranes is that many drugs are weak acids or weak bases. The more the drug is in its un-ionized form, the more likely it is to be lipid-soluble and transferred by passive diffusion through the membrane. For a weak acid or base, the pK_a value will determine the degree of ionization, as described by the Henderson–Hasselbalch equation.

For a weak acid the ionizing reaction is:

$$pH = pK_a + \log[A^-]/[HA]$$

and for a weak base it is:

$$pH = pK_a + \log[B]/[BH^+]$$

Box 6.1 Factors altering absorption and bioavailability from gut

- Gut motility
- Intestinal pH, mucus, bile salts
- Enterohepatic circulation
- Exercise
- Reduced absorptive area
- Reduced intestinal blood flow
- Intestinal microflora that may metabolize some drugs

The pK_a is a measure of the relative strength (degree of ionization) of a weak acid or base (the pK_a of a drug is that point at which the compound is 50% ionized). The lipid solubility of the uncharged species also depends on the chemical nature of the drug. For example, although streptomycin and the related aminoglycosides are uncharged, the high percentage of hydrogen-bonding groups within these molecules renders them hydrophilic.

The molecular shape, as well as the charge distribution, of the drug molecule determines which membrane pores it may traverse. Small polar molecules, such as urea, readily traverse small aqueous pores in the membrane, thus accounting for the high permeability of cell membranes to these substances. Most drugs, however, are too large to pass through these pores.

The rate of drug absorption varies with the route of administration

The main routes of administration (besides intravenous) are oral, sublingual, rectal, topical (e.g. skin, conjunctival fornix), subcutaneous and intramuscular.

Absorption of orally administered drugs is affected by gastric pH, rate of emptying of gastric contents, presence of food and surface area of absorptive mucosa (in disorders such as Crohn's disease the absorptive surface area may be reduced). It is important to consider drug interactions during multiple drug therapy and their effect on drug absorption. For example, in migraine, gastric emptying is delayed, reducing the absorption of analgesics such as aspirin and paracetamol. This may be overcome with the adjunctive use of parenteral metoclopramide, which increases the rate of gastric emptying. The presence of food is generally unimportant, except in the case of tetracyclines (used to treat certain forms of external eye disease commonly associated with acne rosacea), which form insoluble salts with magnesium and calcium. Some drugs will be inactivated within the gut lumen (e.g. benzylpenicillin and insulin). Most malabsorption syndromes do not affect drug absorption, but in other disorders, such as congestive cardiac failure, drug absorption may be impaired because of the secondary gastrointestinal mucosal oedema. Factors affecting bioavailability are shown in Box 6.1.

Box 6.2 Factors affecting drug distribution

- Physicochemical properties of drug
- Binding to plasma proteins
- Binding to tissue proteins
- Relative blood flow to different tissues

Drug distribution

Once a drug is absorbed, it has the potential to penetrate most compartments of the body, therefore the distribution of the drug depends largely on the route of administration. Intravenous and intramuscular administration of a drug result in high drug availability. Buccal (sublingual) absorption of a drug is used to reduce the extent of first-pass metabolism by the liver (as with topically applied eye drops), which invariably occurs with orally administered drugs that reach the liver via the portal circulation. The distribution of the drug and its ability to penetrate cells are also dependent on its physicochemical properties and thus the extent of binding to tissue proteins or cell membrane receptors (Box 6.2).

Once the drug has reached the systemic circulation, it may become bound to circulating proteins, commonly *albumin* or α_1-*acid glycoproteins* (for basic drugs). Protein-bound drugs are restricted in their distribution into tissues, which reduces the availability of the free drug for pharmacological effect. This, in turn, depends on the affinity of the protein for the drug. High levels of protein binding may occur with acidic drugs that are bound to albumin. If a drug is less than 90% bound to plasma proteins, changes in the plasma protein concentration make little difference to the overall amount of unbound drug in the circulation. In cases of drugs with a high binding affinity for protein, a decrease in the steady-state total concentration of the drug, and as a consequence a comparative increase in the clearance of the increased amount of free drug, would occur if for any reason protein binding was impaired. For instance, impaired plasma protein binding in the case of phenytoin results in peaks of unbound active drug in the plasma.

Basic drugs are bound in varying degrees to α_1-acid glycoprotein, a protein that increases in concentration in certain pathological conditions such as acute inflammation. Under such conditions, the binding of basic drugs (e.g. propranolol) may be increased, thus reducing their effect.

Systemic factors that alter protein binding include hypoalbuminaemia (plasma concentration of albumin less than 25 g/L), renal failure, competition by other highly protein-bound drugs, and changes occurring during the last trimester of pregnancy, such as the diluting effect of the increased plasma volume. In general, competition for binding by other

drugs is the major factor affecting distribution because adequate compensatory mechanisms can be initiated to counteract the other causes. The distribution of a drug, as mentioned previously, may also be regulated by the binding of a drug to tissue proteins, a process regulated by the abundance of binding sites, affinity constants and the binding of a drug to its receptor. This in turn may give rise to a desired or undesired effect, although receptor numbers are unlikely to be high enough to alter the distribution of the drug appreciably. Additionally, the gut microbiota can have both a direct and indirect effect on the drug availability and metabolism. The gut microbiota are a flora delivering metabolism of some pro-drugs as well as a substrate including acetylation, deacetylation and demethylation.

In the eye, both the blood–retinal barrier and the blood–aqueous barrier (see Chapter 1, p. 51) limit the distribution of drugs. Tight junctions between the retinal pigment epithelium (RPE) and endothelium of the retinal vessel endothelium give rise to a relatively impermeable barrier to water solutes and larger molecules. Under normal conditions only lipid-soluble drugs will move between the blood and retina. Similarly, the apical membranes of the non-pigmented ciliary body epithelium and the capillary endothelium of the iris are bound by tight junctions, raising a barrier to all but lipid-soluble drugs (see Chapter 1, p. 30).

Drugs are metabolized to facilitate clearance

The metabolism of most drugs occurs almost entirely in the liver, enzymatically altering the drug to increase its water solubility in preparation for excretion and simultaneously making the compound metabolically and pharmacologically active or inactive. Metabolism can affect the drug in various ways. The first is activation of the parent drug, which may itself be inactive (known as a *prodrug*). The drug may be metabolized to form active metabolites, as for example in the case of diamorphine and diazepam. However, metabolism of a drug can also produce toxic metabolites that may persist in the circulation for longer than the parent molecule and thus restrict the continued use of such drugs (e.g. lidocaine). In general, most drug metabolism modifies the drug so that it can then be excreted, usually in the urine but sometimes in bile.

Drug metabolism takes place in two stages: phase I (oxidation) and phase II (conjugation). Phase I reactions are carried out by a heterogeneous group of microsomal enzymes called cytochrome P_{450}, of which various forms exist (Box 6.3).

This enzyme system exists in abundance in the liver but is also found in some peripheral organs, including the eye. It is important to note that the activity of these enzymes can be induced by other drugs (e.g. phenytoin and carbamazepine), which accounts for many well-recognized drug interactions. Conjugation reactions appear not to be affected by enzyme-inducing drugs to the same extent as oxidative metabolism. There are also other oxidative enzymes that are not part of the cytochrome P_{450} system but which are involved in drug metabolism; these include xanthine oxidase (e.g. purine metabolism), alcohol dehydrogenase and monoamine oxidase (e.g. catecholamine metabolism). Many drugs also affect the function of microsomal enzyme systems (Box 6.4).

The metabolism of a drug is dependent on several factors. Oxidative metabolism is affected by age. Premature babies metabolize poorly and, similarly in the elderly, oxidative drug metabolism is reduced because of reduced liver size. Smoking may induce certain liver enzymes, necessitating an increased dose of drug for a required effect. Alcohol, on the other hand, inhibits drug metabolism, particularly

Box 6.3 **Drug-metabolizing systems**

Phase I metabolism: oxidation

Cytochrome P_{450} (microsomal)
- aromatic hydroxylation
- aliphatic hydroxylation
- *N*-deamination
- *N*-dealkylation
- *S*-oxidation
- desulphuration

Phase II metabolism: conjugation

Conjugation occurs with:
- glucuronic acid
- glycine
- glutamine
- sulphate
- acetate

Box 6.4 **Drugs acting as microsomal enzyme manipulators**

Enzyme inducers

- barbiturates
- phenytoin
- phenothiazines
- rifampicin
- griseofulvin
- nicotine

Enzyme inhibitors

- isoniazid
- chloramphenicol
- metronidazole
- warfarin
- carbon monoxide

during and after binge drinking. Both severe liver disease and poor nutrition may markedly impair drug metabolism.

Metabolized drugs are excreted in urine and bile

Renal excretion

Drugs differ greatly in their excretion via the kidney. Some drugs are cleared in a single transit through the kidney, while others are poorly cleared. This difference is dependent on the kidney's ability to handle the drug and on physicochemical properties of the drug. Certain drugs may be filtered through the glomerulus (depending on molecular weight), others are actively secreted by the tubules, and still others passively diffuse across the tubular epithelia (reabsorption).

Glomerular filtration. Glomerular excretion of a drug is only possible when drugs are not bound to plasma proteins and the drug is of a molecular size that can be filtered freely, irrespective of charge (less than 20,000 Da). The clearance of the drug is therefore related to the unbound fraction of the drug and is dependent on the glomerular filtration rate. Glomerular filtration at most removes only about 20% of drug reaching the kidney; the remaining drug passes to the capillaries lining the tubules. Most drugs are presented to the kidney in a carrier form and thus their clearance is slow. Carrier-mediated transport can increase the clearance of the drug even if it is highly bound. Benzylpenicillin is 80% bound and cleared very slowly by glomerular filtration but is almost completely removed by tubular secretion.

Secretion. Drugs may be actively secreted by the tubules into the glomerular filtrate. Different secretory systems, which are relatively non-selective carrier systems, exist for acidic, basic and neutral drug transport into the tubular lumen. Competition for active sites occurs, so that the secretion of drugs can be blocked by other drugs, for example secretion of penicillins is blocked by probenecid, thus decreasing the excretion of the drug.

Reabsorption. If the tubules were freely permeable to drugs, the drug concentration in the filtrate would be similar to that in plasma. Drug reabsorption occurs mainly as a result of the enormous reabsorption of water achieved by the nephron. This builds up a concentration gradient that drives the drug back into the plasma. Thus, drugs that are highly lipid-soluble are excreted slowly and the reabsorption of drugs, especially those that are weak acids or bases, depends on the drug being in its un-ionized form. Therefore, altering the pH of the urine will increase the elimination of the drug, so that acidic drugs are eliminated more quickly in alkaline urine. This manipulation of urine pH is used clinically for the treatment of aspirin overdoses, where forced alkaline diuresis is an effective way of increasing the excretion of the acidic drug. Measurement of glomerular filtration can be achieved with the use of agents that are completely filtered and neither secreted nor reabsorbed, for example inulin. Conversely, the clearance of a drug that is completely secreted in one transit through the kidney will correspond to the renal plasma flow, for example *p*-aminohippuric acid.

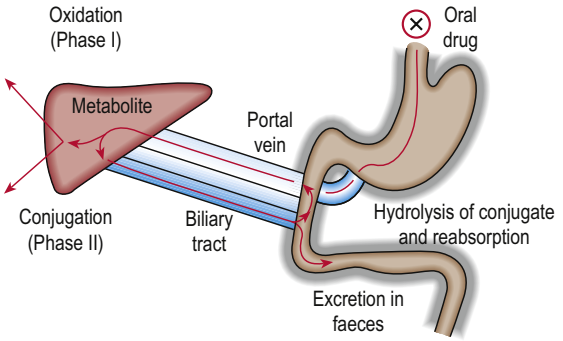


Fig. 6.3 Drug metabolism: enterohepatic circulation.

Biliary excretion

Liver cells also possess transport systems similar to those in the renal tubules, which can transfer drug metabolites from blood to bile. Conjugated drugs (particularly with glucuronate) are concentrated in the bile and delivered to the intestine, where the conjugate may be hydrolysed, releasing the active drug (Fig. 6.3). The drug may then be reabsorbed and its duration of action prolonged (enterohepatic circulation). This is particularly important for digoxin, which is excreted in the bile in an unconjugated form, and for morphine, which is transported as a glucuronide. Some drugs, for example rifampicin, are excreted into the bile and will be excreted unchanged in the faeces.

Pharmacodynamics: drug handling by the body

Pharmacodynamics considers the effects of a drug and the relationship between drug concentration and response. The drug effect is usually initiated by its binding to a cellular membrane receptor, which is specific for the drug, as is the case with autonomic nervous system neurotransmitters, or by non-specific mechanisms, where specific cell membrane receptors do not exist but drug action is dependent solely on its physical properties (i.e. lipid solubility), or by its ability to inhibit specific biochemical enzymes. The effects of enzyme inhibition may be direct, for example by blocking sodium/potassium ATPase pumps or closing ion channels, or indirect, for example by acting on calcium channels.

Box 6.5 **Receptors**

Receptors are unique cell surface proteins capable of binding specific associated substances (ligands). Receptors may be divided into subtypes, depending on the type of agonist or antagonist associated with them. Often receptors can be downregulated or upregulated in different disease states or with chronic drug usage.

A ligand is a specific compound (drug or natural substance) that binds to a receptor.

Agonists are substances that, when bound to a receptor, produce a response that may result in stimulation or inhibition of cell function.

Antagonists are substances that prevent receptor activation.

A partial agonist is a ligand that possesses both antagonist and agonist properties. As such, its maximal effect is less than that of a pure agonist. However, when occupying a receptor, it prevents the actions of other agonists, which are therefore less efficacious.

Some drugs act on intracytoplasmic receptors or cell nucleus receptors. The variety of drug actions is presented in Box 6.5.

Further information on drugs and receptors is available online at https://expertconsult.inkling.com

Drug–receptor interactions

Drugs bind to target molecules on the cell, which then initiate their pharmacological effect. Receptors are proteins that, in general, are situated on cell membranes and are able to bind ligands (e.g. neurotransmitters and hormones) (see Box 6.5). The recognition of a ligand by a receptor is analogous to antibody–antigen binding, that is the lock and key principle (see Chapter 7, p. 407). The molecular mechanisms involved in this receptor–effector pathway are known as *transduction* mechanisms, where the receptor is linked either directly or indirectly to the effector system altering cell function. Several forms of receptor–effector linkage are recognized and include:

- direct regulation of membrane permeability to ions
- regulation of cell function via an intracellular second messenger
- regulation of cell function by regulating DNA transcription and protein synthesis.

Ion channels

Receptors can be linked directly with ion channels, which function only when the receptor is occupied by an agonist. This is the fastest type of receptor response, as, for example, when a neurotransmitter acts on a nerve ending. The excitatory neurotransmitters acetylcholine and glutamate cause a direct increase in both sodium and potassium permeability, which results in depolarization of the cell. An equilibrium between open and closed ion channels exists and random fluctuations in conductance occur when the ion channels are opened (or closed) after receptor stimulation. The duration of this response can be measured and is referred to as the *mean channel lifetime*.

Second messengers

Receptor binding on extracellular receptors alters intracellular functions by activating a secondary messenger, for example the activation of the enzyme *adenyl cyclase* and calcium ion influx. Activation of these secondary messengers regulates various intracellular activities. In many cases they generate protein kinase activation and phosphorylation of membrane proteins (see Chapter 4, p. 185). The consequence of any change in protein conformation of the cell membrane may lead to opening of ion channels (e.g. sodium/potassium or calcium channels). Adenyl cyclase may also be activated by G proteins, which are discussed later.

Cyclic adenosine monophosphate (cAMP)

Cyclic AMP is a nucleotide synthesized within the cell from adenosine triphosphate (ATP) by the enzyme *adenyl cyclase* (see also Chapter 4, p. 185), and is inactivated enzymatically to 5′AMP by hydroxylation with a group of enzymes called *phosphodiesterases*. The following components are required to generate a secondary messenger: a receptor, which faces outwards from the cell; a regulatory protein (G protein), which faces inward towards the cytoplasm; adenyl cyclase; and a cAMP-dependent protein kinase. Inhibition of the secondary messenger activation can be induced with cholera toxin, which binds specifically to the G protein in the intestinal mucosa and thus prevents hydrolysis of another membrane nucleotide, guanine triphosphate (GTP), and adenyl cyclase (Box 6.6).

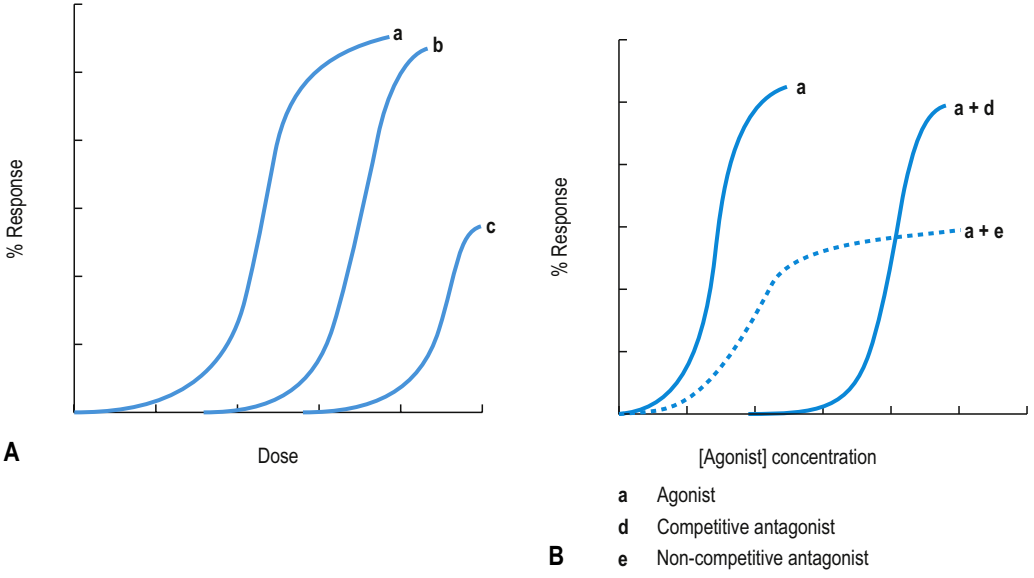
Calcium is a major mediator of cellular activity

Virtually all the calcium in the body is in the skeleton as hydroxyapatite. Intracellular calcium only constitutes 1% of total body calcium. The plasma concentration of calcium is 2.5 mmol/L; 50% is in its ionized form and 50% is bound to proteins or complexed with anions. A rise in the intracellular free calcium concentration can occur as a response to hormones and transmitters, by a net influx of calcium ions into the cell. It may also occur as a result of a release of sequestered

DRUGS AND RECEPTORS

A drug that acts on a receptor may act as an *agonist* or *antagonist*, depending on the response that is elicited. In eFig. 6.1A, drugs a and b are both agonists, but drug b is less potent. Drug c, on the other hand, acts as a partial agonist. Antagonists can be either competitive or non-competitive. If the antagonist is displaced by increasing concentrations of the drug (agonist), then competitive inhibition is present. Non-competitive antagonism describes the situation where the antagonist blocks the action of the agonist without competing with the receptor, which may not be overcome by increasing the concentration of the agonist, hence shifting the curve to the right and depressing the maximal response (eFig. 6.1B).

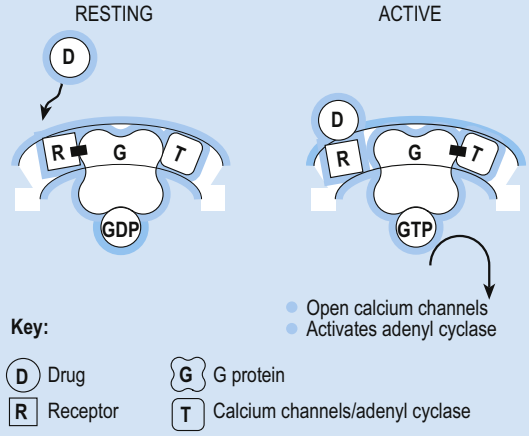
Drugs often give a graded dose–response curve, where increasing the drug concentration will increase the drug effect. This graded response is seen with drugs that are not permanently bound to the receptor. Thus, the *efficacy* of a drug is defined as the maximal response it can give, whereas *potency* describes the amount of drug required to give the desired response. Thus, some drugs may be efficacious but not potent, requiring large doses to give an effect. However, if the drug is irreversibly bound to the receptor, its effect will continue well after the elimination of the drug from the bloodstream. Alternatively, the efficacy of a drug may diminish with time, an effect known as *tolerance*, which is thought to be the result of downregulation of specific drug receptors. Some drugs also produce active metabolites that continue to give a pharmacological effect well beyond the half-life of the parent drug. All these effects may complicate the study of pharmacokinetics.



eFig. 6.1 (**A**) Potency of receptor stimulation by agonists and partial agonists. (**B**) The effects of agonist action by competitive and non-competitive antagonists.

Box 6.6 G proteins regulate second messenger activity

G proteins are so-called because of their affiliation with the guanine nucleotides GTP and GDP. They consist of three subunits, which catalyse GTP conversion to GDP. In the resting (empty receptor) state, the G protein–GDP complex is associated with the receptor. The presumed conformational change in the cell membrane when the receptor is occupied acts to increase the affinity of the neighbouring G protein for GTP and to promote the binding of the G protein–GTP complex to the effector site. This complex can activate ion channels, adenyl cyclase and other secondary messengers, as shown in the figure.



calcium without a net influx into the cell. The intracellular calcium concentration is 10^{-7} mol/L, most of which is protein bound, and small changes in free calcium concentration may affect many intracellular processes. Intracellular calcium is normally regulated by an ATP-active transport channel. Accompanying the increase in intracellular calcium level, there is also an increase in the rate of degradation of minor membrane phospholipids (phosphatidylinositols) by phospholipase C, which in turn alters the membrane permeability and control of membrane phosphorylation, either directly, via its effect on protein kinases, or indirectly, through activation of adenyl cyclase. At the same time a phospholipid flip-flop mechanism occurs within the cell membrane, where methylation of phospholipids converting phosphatidylethanolamine to phosphatidylcholine takes place. This further increases calcium permeability and secondarily regulates adenyl cyclase activation. Intracellular effects of calcium are controlled by an intracellular acidic protein, calmodulin. The calcium–calmodulin complex can thus activate many enzyme systems, including protein kinases, adenyl cyclase, phosphodiesterases and calcium-dependent ATPase (Fig. 6.4). The calcium–calmodulin

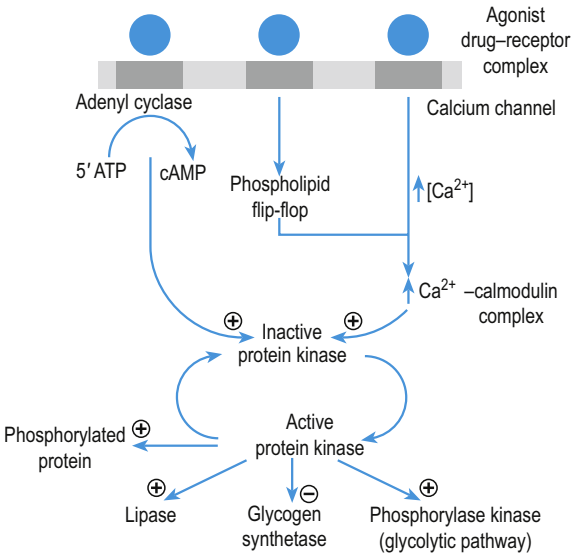


Fig. 6.4 Activation of membrane-associated enzymes by calcium–calmodulin secondary messengers.

complex is inactivated via the binding of trifluoperazine (an antipsychotic agent), which can be used to study the mechanism of action of certain therapeutic agents.

Regulation of protein synthesis

Steroid hormones, because they are lipophilic, can diffuse across the cell membrane and bind to cytoplasmic and nuclear receptors and nuclear chromatin. Because nuclear receptors bind directly to DNA they act as transcription factors, regulating gene expression. As a result of their receptor–ligand role as transcription factors their response regulates homeostasis. Certain regions of the DNA sequence show a high affinity for steroid–receptor complexes, and binding of these complexes generates an increase in RNA polymerase activity and generation of messenger RNA (mRNA) and thus production of proteins. One steroid may result in the production of several different species of mRNA within each cell, which may explain the great diversity of steroid actions.

Ligand–receptor affinity determines selectivity

The affinity of a ligand for its receptor is measured by the amount of ligand required to achieve half-maximal binding (EC_{50}). The binding force of the receptor–ligand complex can also be measured by the length of time taken for dissociation of the complex to occur. Low-affinity agonists include those required to potentiate a rapid response (e.g. neurotransmitters) and therefore concentrations of the ligand close to the receptor during stimulation must be high. The *selectivity* of a

ligand for receptors is a ratio of the EC_{50} for the receptors being compared. If the ligand (or drug) is more selective for receptor A than for receptor B, it will achieve its clinical effect by stimulation of receptor A at lower doses but will still stimulate receptor B if sufficient drug is given. Therefore, clinically, it is important to establish the EC_{50} ratio of a given drug; ratios lower than 50 are considered unsafe for clinical use if stimulation of other receptors is to be avoided.

Genetic control of drug handling influences response

Defining the genetic make-up within populations or, more pertinently, the individual has assisted in understanding responses to drugs. The hope is that with such understanding we may not only increase our knowledge of disease pathogenesis but also develop tailored therapies for the individual, whilst reducing adverse effects.

There are historical examples of clear distinctions in the individual's capability to metabolize drugs. This is seen with acetylation of drugs. N-acetylation may be either fast or slow, depending on the amount of enzyme present, which is controlled by a single recessive gene associated with low hepatic acetyltransferase activity (e.g. in isoniazid metabolism). Patients with low acetyltransferase activity are known as *slow acetylators*. Another genetic variation in the rate of drug metabolism is seen with suxamethonium (a depolarizing neuromuscular blocker used in general anaesthesia). About 1 in 3000 individuals fails to inactivate suxamethonium by hydrolysis (pseudocholinesterase), which is the result of a recessive gene that gives rise, in homozygotes, to an abnormal cholinesterase with a much lower substrate affinity. Patients who have a hereditary erythrocyte glucose-6-phosphate dehydrogenase enzyme deficiency might develop a haemolytic anaemia when treated with a number of drugs, including chloroquine, vitamin K, acetylsalicylic acid (aspirin) and probenecid.

P450 gene. The drug-metabolizing cytochrome P450 enzyme genes have been identified and cloned to include a large family of mono-oxygenases. Genes encoding P450 enzymes are polymorphic and within three families. As a result, there are broadly four functional outcomes: *fast* metabolizers who inherit multiple gene copies; *normal* metabolizers who have two normal alleles; *intermediate* metabolizers who have one deficient allele; and *poor* metabolizers who lack the functional enzyme.

The genes encoding families of enzymes are called *cyp* and are given numbers associated with a specific group of enzymes within the gene family, subfamily and then specific gene itself: e.g. *CYP27A1* relates to group 27, subfamily A and gene 1.

In the eye *CYP1B1* is a gene that encodes for an enzyme localized to the endoplasmic reticulum, which is involved in metabolizing aromatic hydrocarbons and steroids. In addition to metabolism, within the eye mutations are linked with congenital glaucoma as well as rare but significant retinal dystrophies such as Bietti crystalline dystrophy. The aryl hydrocarbons have implications in ageing and cell disfunction secondary to smoking and cataract formation or age-related degeneration (see Chapter 9, p. 518) (Box 6.7). In addition, the aryl hydrocarbon receptor (AhR) has a fundamental role in regulating immune mechanisms particularly those affecting inflammation and autoimmunity.

Box 6.7 Genetic control of immunosuppressives dictates drugs response

Azathioprine is a commonly used steroid-sparing immunosuppressant for inflammatory disorders, including for the eye in thyroid eye disease and uveitis. It is a purine analogue, whose active metabolite component, 6-mercaptopurine's conversion to cytotoxic 6-thioguanine analogues inhibits proliferation of cells. Most importantly, it carries a significant potential for adverse effects. These have been reduced with improved understanding of pharmacogenetics. 6-mercaptopurine is metabolized via thiopurine methyltransferase (TPMT). Individuals (up to 1:1000 patients) who inherit two non-functional alleles for the polymorphic *TPMT* gene will have dramatic life-threatening myelosuppression. Importantly, up to 15% of the population are heterozygotes and will have increased incidence of myelosuppression. Genetic testing is now available as well as phenotyping by measuring TPMT activity in red blood cells.

Ocular pharmacology: drug handling by cells and tissues of the eye

This section describes the pharmacokinetics, pharmacodynamics and modes of drug delivery (excluding systemic administration) that are used to treat ocular surface and intraocular conditions.

Mechanisms of ocular drug absorption

There are several methods of administering ocular medications, including extraocular routes (topical) via either conjunctival/episcleral absorption (non-corneal) or transcorneal absorption, and direct intraocular administration of drugs. The non-corneal route of absorption may be significant for drugs that do not penetrate the cornea well. Corneal absorption still represents the major route of absorption for most

ocular medication, requiring in the main, high concentrations of drug in a small volume for effect.

Factors influencing delivery of drugs to the eye

Drugs can be administered in many different topical forms, including solutions, gels and ointments. The efficacy of treatment is usually dependent on intraocular penetration, which in turn is dependent on the permeability of the drug across the cornea, and the anatomical and physiological influences of the local environment, including lacrimation, tear drainage and the composition of the precorneal tear film.

Routes of administration

Conjunctiva. Topical administration into the inferior fornix of the conjunctiva is by far the most common route of ocular drug delivery. Both lacrimation and blinking profoundly influence the residence time of fluid in the fornix. Therefore, the efficacy of such delivery systems depends on the anatomy and physiology of the lids, the precorneal tear film, and the health of the conjunctiva, cornea and lacrimal system.

The conjunctival sac has a capacity of approximately 15–30 μL (dependent on blinking) and the natural tear film volume is 7–8 μL. The tears turn over at approximately 16% per minute during a normal blink rate of 15–20 blinks per minute. Most solution applicators deliver between 50 and 100 μL per drop, so a substantial amount of drug will be lost through overspill on administration. The turnover of tears is also highly dependent on environmental conditions, particularly temperature and humidity. The epithelium of the conjunctiva is continuous with that of the cornea and epidermis of the lids (see Chapter 1, p. 88) and contains goblet cells, which produce mucus and are integral to the stability of the tear film (see Chapter 4, p. 204). Drug absorption through the conjunctiva therefore first requires transport through the epithelium. In the subconjunctival stroma, which is a highly vascular conjunctiva owing to the rich superficial venous plexus and lid margin vessels, drugs may be absorbed in significant concentrations into the circulation. Also, after administration into the inferior fornix, drugs drain directly through the nasolacrimal duct into the nose, where measurable systemic absorption of drugs via the nasal and nasopharyngeal mucosa occurs. Restricting the entry of a topically applied ophthalmic dose into the nasal cavity by nasolacrimal occlusion for 5 minutes or by making appropriate alterations to the vehicle (i.e. from solution to ointment) increases the residence time of the drug in the fornix and increases ocular absorption.

Precorneal tear film and cornea

Tears are considered to act as a buffering system for many substances. The pH of normal tears varies between 6.5 and 7.6, while many drug delivery systems are often formulated at pH of less than 7; the return to physiological pH after drug instillation is, however, more likely to be a function of increased tear turnover than the result of a buffering effect. The precorneal tear film is composed of an outer lipid layer (mixed lipids), a middle aqueous layer (including proteins) and a deeper mucin layer (glycoprotein) (see Chapter 4, p. 204). The mucin layer contributes to the stability of the tear film, as well as promoting adherence of the tears to the lipophilic corneal and conjunctival epithelium. Any alteration in the components of the tear film will result in instability of the tear film and a reduced conjunctival residence time of the drug. At the same time alteration in the pH of the tear film may affect the ionization of the drug and thus its diffusion capacity. In spite of the extensive losses to the exterior and to systemic absorption, topical (conjunctival) administration of drugs achieves acceptable intraocular levels, mainly because of the very high concentrations that are administered.

The pH of the tear fluid is important in drug formulation because of the physiological homeostatic mechanisms. Drug penetration may be enhanced by changing the degree of drug ionization and enhancing the product stabilization over a range of pH changes. The epithelium of the cornea represents the most important barrier to intraocular transport of drugs via this route. First, the stratified cellular epithelium is bound by desmosomes between the lateral borders of the superficial cells. Second, the corneal epithelium is hydrophobic (as are all cell membranes), and so will allow only lipid-soluble drugs to pass through. In addition, Bowman's membrane, an acellular collagenous sheet (10 μm thick) between the basement membrane of the epithelium and the stroma of the cornea, acts as a further barrier to the penetration of drugs. In contrast, the stroma, which accounts for 90% of the corneal substance and its ground substance (glycosaminoglycans and water), permits ionized water-soluble drugs to pass more efficiently than lipid-soluble drugs. Finally, transport across the single-layer endothelium of the cornea is relatively free because it contains gap junctions that permit good penetration of most drugs into the aqueous humour. Many topical eye medications are weak bases, for example tropicamide, cyclopentolate and atropine, and exist in both ionized and un-ionized forms within the pH range of the tear film (pH 7.4). Altering the solution pH of timolol (pK_a 9.2) from 6.2 to 7.5 increases its corneal penetration and systemic absorption. The partition coefficient (ratio of concentrations in the two compartments) may therefore be increased by raising the pH of the water phase, rendering the drug non-ionized and

Box 6.8 **Factors influencing topical drug absorption**

- Environmental conditions: temperature and humidity
- Volume of drug application
- Drug formulation: pH, preservative, vehicle type
- Blink rate
- Stability of tear film
- Absorption through conjunctival vessels and nasal mucosa
- Corneal epithelium and stroma
- Nasolacrimal drainage of tears

Box 6.9 **Enzymes involved in drug metabolism in the eye**

- *Ketone reductase:* this is a cytosolic enzyme dependent on NADPH. It is thought to play a key role in the metabolism of timolol and analogues of propranolol. Found in corneal epithelium, lens, iris and ciliary body.
- *Esterases:* important in the activation of ester pro-drugs. Both acetyl- and butyrylcholinesterases are found in rabbits. They are widely distributed throughout the anterior segment.
- *Classic phase I and II oxidizing and conjugating enzymes:* cytochrome P_{450} reductase, demethylase, sulphatase and glucuronidase.

more lipid-soluble. The factors affecting topical drug absorption are summarized in Box 6.8.

Delivery methods of ocular medication

Residence in the conjunctival sac

The bioavailability of ocular medication depends on the precorneal fluid dynamics, drug binding to tear proteins, conjunctival drug absorption, systemic drug absorption, resistance to corneal penetration, drug binding to melanin and intraocular drug metabolism. Both the absorption and the efficacy of the drug can be increased by altering the formulation of the drug and/or by changing the local conditions.

To increase the residence time of the drug in the inferior fornix, and thus its delivery to the corneal epithelium, attempts at reducing the instilled volume of drug and increasing the viscosity of the solution have been made. For instance, polymers that increase solution viscosity include polyvinyl alcohol, hydroxypropylcellulose and other cellulose derivatives. However, increasing the viscosity of the delivery solution only produces a modest gain in ocular drug absorption, particularly for lipid-soluble drugs.

The residence time for many drugs is especially reduced by their pH, tonicity and the direct effect of certain drugs on lacrimation, which all affect the residence time of drugs in the lower fornix. For example, formulation at acidic or alkaline pH causes irritation to the eye and increases lacrimation and the blink rate and clearance of the drug. Attempts to reduce the tonicity of the drug solution by using dilute buffers (e.g. phosphate buffers) to prevent stinging and lacrimation, and therefore increase transit time (reduce clearance), have also been used. Drugs that have a direct pharmacological action on the lacrimal gland, increasing lacrimation and subsequently altering the precorneal fluid dynamics, include muscarinic agonists. Some drugs are also bound to the tear proteins (albumin, globulins and lysozyme), reducing the concentration of the free drug available for absorption.

The corneal epithelium presents a considerably greater barrier to hydrophilic than to lipophilic drugs (10:1). Corneal epithelial permeability, however, may be increased during ocular inflammation so that some drugs, for example dexamethasone, are more rapidly absorbed across the corneal epithelium into the eye. Preservatives such as benzalkonium chloride have also been shown to enhance the ocular absorption of drugs. Benzalkonium chloride and other cationic surfactants increase the ocular absorption of drugs by increasing corneal permeability (by compromising corneal integrity), depending on the molecular size and lipophilicity of the drug. Increased absorption can be obtained with pilocarpine, prednisolone and homatropine. For a drug to penetrate optimally, it must be able to exist in both ionized and un-ionized forms. Drugs will be buffered by the precorneal tear film and any alteration in the pH will change the ratio of ionized to un-ionized forms of the drug, dependent on the pK_a of the drug. In general, a drug that exists in a purely ionized form will not penetrate the cornea unless the cornea has been damaged. Once absorbed into the eye, drugs may be bound to melanin within the pigment epithelium of the iris and the ciliary body, which may in turn reduce the bioavailability of the drug and also retard its clearance, leading to increased and prolonged drug levels. Similarly, after penetrating into the eye, drugs may be rendered inactive by intraocular metabolism. Enzymes that participate in ocular drug metabolism include those involved in inactivating neurotransmitters, for example monoamine oxidase, catechol *O*-methyltransferase, esterases, cytochrome P_{450}, and other enzymes including ketone transferase, glucuronidase and aldose reductase within the lens (Box 6.9). The majority of enzyme activity is microsomal, although some enzymes, for example esterases, are

Box 6.10 **Drug delivery vehicles**

- Solutions, colloids, emulsions, suspensions
- Ointments
- Slow-release preparations
- New ophthalmic delivery systems
- Particulates
- Liposomes
- Intravitreal preparations

both cytosolic and extracellular. Within the anterior segment, the corneal endothelium, the non-pigmented cells of the iris and the ciliary body are metabolically most active, therefore drugs that are good substrates for these enzymes may suffer substantial degradation during absorption.

Drug vehicles affect drug delivery

Several topical drug delivery systems are used in ophthalmology (Box 6.10) and there are a plethora of advances in the pipeline and in early-phase development such as nanoparticles and viral vectors for gene delivery. Nanoparticles are colloidal drug carrier systems in nano- and micro-size range. The advantages are overcoming lipid solubility, improved clinical targeting and depot longer-acting delivery. Similarly, there is increasing development of viral delivered genes to replace mutated or loss of gene function in degenerative disease or as a drug depot to release the active component at the tissue site (see Chapter 3, p. 154)

Solutions

Solutions are a common mode of delivery because they cause less blurring of vision than ointments. They are easily administered and achieve high intraocular concentrations if applied regularly. They do, however, possess a short contact time and are quickly washed away at a rate proportional to the volume instilled. Polyvinyl alcohol or methylcellulose added to the solution increases the viscosity and/or lowers the surface tension, and will thus prolong contact time. Biologically active drug compounds that are sparingly soluble in water are often formulated as suspensions. Ophthalmic suspensions, particularly steroids, are thought to be useful delivery systems because it is assumed that the drug particles persist in the conjunctival sac and give rise to a sustained-release effect. Suspensions tend to form precipitates and thus need to be resuspended in the mixing bottle before application.

Semisolids (ointments)

Ointments consist of any one or a combination of hydrocarbons, mineral oils, lanolin and polymers such as polyvinyl alcohol, carbopol and methylcellulose. Drugs applied by this method provide an increase in the duration of action because of reduced dilution, reduced drainage and prolonged corneal contact time. Although these preparations melt at the temperature of the ocular tissue and disperse within the tear film, they are still retained longer than other ophthalmic preparations. They do, however, give rise to blurring of the vision and an increased incidence of contact dermatitis, related commonly to the preservative within the preparation.

Slow-release preparations

The problem of short residence times has been addressed by the development of ingenious vehicle supports. For instance, controlled release of ocular medications can be achieved with conjunctival inserts or with hydrophilic soft contact lenses. Slow-release preparations allow the constant release of drug while minimizing the drainage rate of the drug.

Ocular inserts. Controlled-release delivery systems deliver a bioactive agent to the target site at a controlled concentration over a desired time course. Ocular inserts are flexible, elliptical devices, consisting of three layers. The two outer coats of ethylene vinyl acetate enclose an inner coat of drug/alginate mix. The ocusert with pilocarpine relies on the solubility properties of pilocarpine-free base, which exhibits both hydrophilic and lipophilic properties. Because the drug is miscible in both aqueous and organic solvent media, it will permeate the hydrophobic controlling membranes by diffusion through the pores. Biodegradable systems such as Lacrisert® contain drugs within carboxymethylcellulose wafers or polyvinyl alcohol discs or rods placed in the lower conjunctival fornix. They are manufactured principally for the treatment of dry eyes.

Collagen shields. Collagen is thought to be a suitable carrier of drugs. The three polypeptide α-helical chains are held together by cross-linking between proline and hydroxyproline, which account for 30% of the amino acid content of the molecule. Because of the ability to control the amount of cross-linking in the collagen subunits by exposure to ultraviolet light during manufacture, the time taken to dissolve when placed on the cornea can be altered. Also, collagen acts as an ion exchanger and is semipermeable, facilitating controlled release of drugs. Thus, the collagen bandage shields prolong contact between drug and cornea. Drugs can be incorporated into the collagen matrix, absorbed on to the shield during rehydration, or applied topically over a shield when in the eye. As the shield is erodible, release of the drug occurs gradually into the tear film, maintaining higher concentrations.

Box 6.11 **Criteria for successful ocular inserts**

- Comfortable and easy handling and insertion
- No expulsion of insert during wear
- No interference with vision
- Negligible interference with oxygen access to cornea
- Reproducibility of kinetics of drug release
- Sterility

Soft contact lenses. In this case, the polymer of the contact lens is hydrophilic and thus water-soluble drugs are absorbed into the lens. The lens is hydrated once placed on to the cornea and so releases the drug until equilibrium is reached between drug concentration in the contact lens and in the conjunctival sac.

A primary concern with all ocular inserts is comfort, while many other criteria determine their usability (Box 6.11).

Intravitreal inserts. Development of inserts directly applied to the vitreous cavity has gained increasing impetus following the advent of successful trial evidence supporting intravitreal drug administration for macular degeneration and vascular occlusions and cytomegalovirus (CMV) viral retinitis (see Chapter 9, p. 502). Such inserts include, for example, the following:

- Ozurdex® is a dexamethasone intravitreal implant inserted directly by injection into the vitreous cavity, using a Novadur® polymer system of poly D,L-lactide-co-glycoside (PLGA), which is biodegradable over months.
- Retisert™ is a controlled delivery system of pellets containing the steroid fluocinolone coated in polyvinyl alcohol and silicon laminate, which is directly inserted to the vitreous and sutured into the sclera for stability.
- Ilivien® is a directly injectable-free floating non-biodegradable insert containing the steroid fluocinolone acetonide in a polyamide cylindrical tube.

Advanced ocular delivery systems

New ophthalmic delivery system (NODS)

NODS is a method of administering a drug as a single unit volume within a water-soluble preservative-free form. The system is easily administered and may offer a significant improvement in bioavailability over drops. Essentially the device consists of a water-insoluble drug-loaded flange attached to the end of a water-soluble handle by a soluble membrane film. The flange end is placed into the inferior fornix and the membrane rapidly dissolves, releasing the flange, which then hydrates within the inferior fornix and releases the drug.

Particulates

Microspheres and nanoparticles represent promising particulate polymeric drug delivery systems for ophthalmic medications. These systems may avoid the potential disadvantages of other delivery systems, which include discomfort and difficulty of use with inserts, blurring of vision with viscous solutions, and instability of liposomes (see later). They may not represent real particles (capsular wall containing an aqueous or solid core) but are matrix-type structures of lipid base and drug. The particles are formed by polymerization (with ultraviolet light), during or after which the drug can be added. This leads to covalent drug binding to the polymer, so that when the drugs are absorbed into the polymer matrix they form a solid–solution matrix. The binding of the drug depends on its physicochemical properties as well as the nature of the polymer (polybutylcyanoacrylate). Smaller particles are better tolerated by patients than larger ones and also have increased drug absorption and much slower elimination rates.

Liposomes

Liposomes are vesicles composed of lipid membranes enclosing an aqueous volume. They form spontaneously when a mixture of phospholipids is agitated in an aqueous medium to disperse the two phases and are composed of phospholipids, including lecithin, phosphatidylserine and phosphatidylglycerol. They therefore share the properties of the bilayer of an outer cell membrane. The drugs can be trapped in either the lipid or aqueous phase. They provide the possibility of controlled and selective drug delivery and thus increased ocular bioavailability, although liposomes carry lipophilic drugs more readily than hydrophilic ones. The advantage of liposomes is that they are easily prepared, non-irritant, and do not cause any blurring of vision. Altering their surface charge or binding specific ligands to them increases their adherence to cells and subsequent endocytosis of the liposome–drug complex. Shen et al. (2018) have reviewed delivery systems for the eye and the associated pharmacokinetics (see further reading).

Ocular iontophoresis

Iontophoresis is a method of drug delivery that uses an electric current to drive a polar drug across a semipermeable membrane; this may be achieved by either a cathode or an anode, depending on whether the molecule is negatively or positively charged. Applications in clinical ophthalmology have yet to be determined but a possibility may be the iontophoretic application of antibiotics in bacterial keratitis, enhancing tissue penetration of the agent. Corneal iontophoresis of gentamicin and aprofloxacin has been successful in treating experimental pseudomonas

keratitis. Trans-scleral iontophoresis allows direct drug penetration into the vitreous but has a major disadvantage of discomfort and may produce small areas of retinal necrosis at sites of application.

A gene therapy approach to delivering treatment

As discussed in Chapter 3, the advances to modulate disease outcome via altering protein expression through gene therapy has begun the revolution for treatment of monogenetic inherited disorders. There are now many gene therapy trials for inherited retinal degenerations, including a first phase III trial leading to approval for treatment of Leber Congenital Amaurosis and retinitis pigmentosa caused by RPE65 gene mutations. Most genes being targeted are of a size to be packaged in adeno-associated virus (AAV). Trials for Stargardt's disease (*ABCA4*) and Usher's syndrome type 1 (*MYO7A*) involve larger genes and therefore use the equine infectious anaemia virus (*EIAV*) vector. For most therapies, the loaded vectors are administered via a vitrectomy and subretinal delivery, although intravitreal administration is being used for the trial for X-linked juvenile retinoschisis. Other ongoing studies currently include: choroideraemia, X-linked retinitis pigmentosa, achromatopia and *MERTK*-retinitis pigmentosa.

Such therapeutic approaches can be adapted to package a gene to secrete therapeutic proteins. Examples include AAV-sFlt-1 (or other anti-VEGF proteins) or AAV-complement regulators (such as CD59 or CFI) delivery for age-related macular degeneration or a non-viral electotransfection in ciliary muscle of DNA plasmid to secrete p55 TNF receptor for uveitis.

Intracameral and intravitreal administration

Despite the considerable ingenuity applied to the development of topical drug preparations, the treatment of many ocular disorders is hampered by poor penetration into the eye. Systemic drug administration also does not guarantee high intraocular drug levels, in part as a result of the integrity of the blood–retinal and blood–aqueous barriers. For instance, the treatment of bacterial endophthalmitis is often inadequate unless vitrectomy and intravitreal antibiotics are used. The recommended dose of intravitreal antibiotics is based on doses that are not toxic to the rabbit or primate retina, but little is known about dose–response in the human vitreous. Experimentally, intravenous administration of the newer cephalosporins produces an incremental increase in the vitreous levels in inflamed eyes (but not in healthy eyes). In inflammatory conditions, water-soluble antibiotics (penicillins) have a prolonged half-life as the retinal pump mechanism is damaged in such eyes. However, the commonly used aminoglycosides are more rapidly cleared from inflamed eyes, despite low lipid solubility, because they are eliminated through the aqueous circulation. To achieve continual maximal therapeutic concentrations in the vitreous, repeated injections may be required. This may lead to complications of lens injury, haemorrhage, infection and retinal injury. Methods to reduce the toxicity of this form of treatment have included development of liposome-encapsulated drugs, which may be administered directly into the vitreous, or subcutaneously/subconjunctivally, acting as controlled-release formulations. Similar approaches are being sought both in humans and experimental models with other agents, for example ganciclovir for cytomegalovirus retinitis, ciclosporin A for endogenous uveitis and encapsulated anti-fibroblastic drugs (5-fluorouracil) for proliferative vitreoretinopathy.

Further information on drugs used for management of CMV retinitis is available online at https://expertconsult.inkling.com



Intravitreal delivery of drugs has revolutionized care for vascular retinal disorders

Understanding that in neovascular (wet) age-related macular degeneration creating choroidal neovascular membranes or diabetic macular oedema is in part driven by pro-angiogenic growth factor, vascular endothelial growth factor (VEGF) has led to development and implementation of anti-VEGF therapies in clinical practice. There are numerous intravitreal injections of drug available to neutralize VEGF activity (Fig. 6.5). In common, the VEGF inhibitors act to bind to soluble VEGF and block activity by preventing binding to VEGF receptors. VEGF actions are mediated principally via receptors VEGF-R1 and VEGF-R2. *Ranibizumab* is a monoclonal recombinant human Fab VEGF; recombinant human FabV2 is a humanized Fab fragment that binds and neutralizes VEGF and *aptamers* (i.e. *pegaptanib sodium*) are nucleic acid ligands that are specific for given proteins and act as decoys, neutralizing their function. Another decoy mechanism is that of high-affinity VEGF blocker that is a fusion of VEGF receptor 1 to a human immunoglobulin G constant region (aflibercept) (Box 6.12). The advancement in biologics is to develop them to act longer at the site and reduce absorption to the systemic circulation. All three agents can rapidly move into the circulation. Bevacizumab (Avastin, a full light chain and truncated heavy chain monoclonal antibody) and aflibercept have evidence of higher systemic exposure.

DRUGS FOR TREATMENT OF CMV RETINITIS

Ganciclovir was the first agent found to demonstrate therapeutic activity against CMV retinitis (see Chapter 8, p. 480). At least 80% of patients respond to initial treatment with parenteral ganciclovir. Complications following parenteral administration include thrombocytopenia and leucopenia.

Zidovudine (AZT) has no direct antiviral action against CMV but may act by diminishing human immunodeficiency virus (HIV) enhancement of CMV infection and by improving immune function. However, AZT also causes bone marrow suppression, making combined therapy incompatible. Human recombinant granulocyte–macrophage colony-stimulating factor stimulates proliferation, differentiation and chemotaxis of neutrophils and, in combination with anti-CMV and HIV treatment, may prevent neutropenic episodes.

Cidofovir is a DNA polymerase chain inhibitor that can be used at maintenance dose once a week.

Foscarnet may also be used to treat CMV infection, but up to 30% of patients develop renal toxicity. Despite this, it is safer for administration in combination with AZT.

Control of CMV retinitis may occur with intravitreal injections of either ganciclovir or foscarnet. Injections are not directly retinotoxic even though intravitreal levels of these agents are within the toxic range for the drug. Intravitreal injections of liposome-encapsulated drugs reduce the requirement for repeated injection, without any evidence of retinal toxicity.

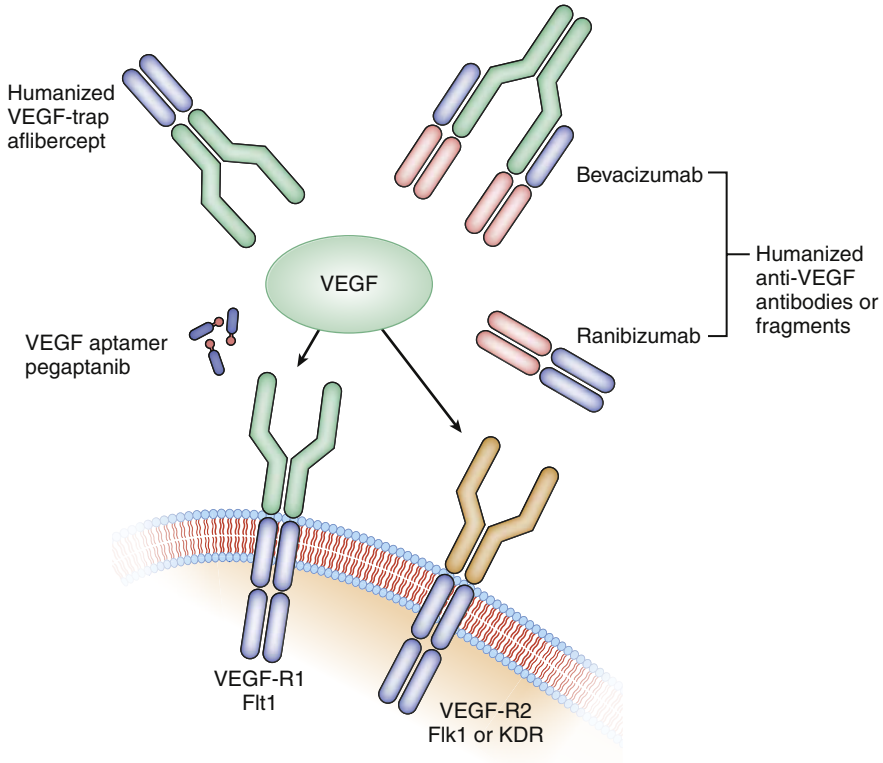


Fig. 6.5 Intravitreal anti-VEGF therapy. Advent of current biologics to inhibit VEGF activity are shown. They target one of the VEGF family of proteins: VEGF A (there are VEGF A–D). *VEGF* binds to numerous receptors, including VEGF-R1-R3 and neuropilin receptor. However, the greater receptor affinity for major pro-angiogenic VEGF A actions is via *VEGF-R1* and *VEGF-R2*, as depicted in this schematic.

Drugs administered systemically also penetrate the eye

The preceding sections focus on the factors that affect the efficacy of topical or direct ophthalmic therapy with minimal systemic side-effects. However, drugs for ocular conditions may be administered systemically to achieve sufficient drug concentration in ocular tissue, although in humans the intraocular drug levels that are reached are largely unknown for most substances.

Commonly used drugs in ophthalmology include the carbonic anhydrase inhibitors (acetazolamide and dichlorphenamide, see later), which are administered orally or intravenously to reduce intraocular pressure. Various animal studies have also demonstrated the ability of systemic antibiotics to reach intraocular infections in concentrations that are bactericidal to certain pathogens. For example, ciprofloxacin penetrates the aqueous humour following oral administration. Similarly, both non-steroidal anti-inflammatory drugs and steroids penetrate the eye when given orally.

Conversely, drugs applied topically may also reach the systemic circulation and affect both eyes. This has been recorded with timolol therapy of chronic open-angled glaucoma with unilateral intraocular pressure rise, which resulted in a significant drop in intraocular pressure in both treated and untreated eyes. In experimental models, a drug effect on the contralateral eye has also been shown with apraclonidine. This phenomenon may occur with a drug possessing a long systemic half-life or unique tissue-binding characteristics within the eye. If supplied in too high a dose or concentration, systemic side-effects may occur, as with 10% phenylephrine drops.

Topical medications and preservatives

Ophthalmic solutions and ointments must be sterile and a wide variety of preservatives are used for this purpose. Most are toxic to the precorneal tear film and epithelium, impeding epithelial healing and disrupting the tear film. The

Box 6.12 **VEGF**

VEGF belongs to a platelet-derived growth factor family that includes VEGF A–D.

VEGF A is a product of the *vegfA* gene encoding multiple isoforms identified by their amino acid number (e.g. $VEGF_{165}$).

Alternative splicing may result in two protein families from the same gene, which for the *vegfA* gene, for example, are denoted $VEGF_{165}$ or $VEGF_{165}b$. The latter is anti-angiogenic and is important for tissue homeostasis.

There are a number of VEGF receptors, which include the major pro-angiogenic receptors VEGF-R1 (Flt1), VEGF-R2 (KDR) mediating via tyrosine kinase's VEGF A effect.

VEGF-R3 (Flt4) mediates lymphangiogenesis through VEGF C and D, while the fourth receptor, neuropilin receptor (NRP1 receptor), which binds VEGF A, is important for cell survival and axonal guidance.

commonly used preservatives are benzalkonium chloride, thiomersal, chlorbutol and organomercuric compounds. Benzalkonium chloride is a surfactant preservative (cationic preservative) that attains its bactericidal activity by attaching to the bacterial cell wall, increasing permeability and eventually rupturing the cell wall. While these preservatives and other cationic surfactants compromise the corneal integrity, they have also been shown to enhance the ocular absorption of drugs. Benzalkonium chloride is most effective at an alkaline pH (approximately pH 8.0) but is inactivated by the presence of soaps and salts, for example magnesium and calcium. As such, some contact lens solutions also combine ethylenediaminetetra-acetic acid (EDTA; a chelating agent) to overcome this problem. Severe toxicity may result from direct cellular damage or from a hypersensitivity reaction to components of the drug and give rise to papillary conjunctivitis, punctate keratitis and corneal oedema. Chlorbutol also reduces oxygen utilization of the cornea and may result in epithelial desquamation. Mercurial compounds include phenylmercuric acetate and thiomersal. Hypersensitivity to these compounds is the most dramatic and common complication of preservatives in 10% of patients, and mercurial deposits can develop in corneal tissues. All these preservatives are absorbed to various degrees by soft contact lenses.

Reconstituting the tear film

With increased knowledge of the anatomy and physiology of the precorneal tear film (see Chapter 1, p. 15 and Chapter 4, p. 206), tear substitutes have been generated to provide symptomatic relief by artificially reconstituting individual tear film components. Ocular surface disease results from abnormalities in one, but generally more, of the tear film components. *Aqueous* deficiency is observed classically in keratoconjunctivitis sicca, as occurs in, for example, Sjögren's syndrome. Instillation of artificial tears is the mainstay of treatment in conjunction with reducing drainage by occluding the lacrimal puncta. *Mucin* deficiency occurs in conditions that affect goblet cell function, for example cicatricial conjunctival disease and hypovitaminosis A, the effect of which is an unstable tear film because the surface tension is reduced and so is insufficient to maintain the aqueous layer. Therefore, areas of non-wetting occur and if left untreated corneal ulceration and scarring can result. Finally, *lipid* abnormalities are associated with chronic inflammation of the meibomian glands and, again, result in an unstable tear film and inadequate wetting of the cornea.

Tear substitutes

Artificial tears or ocular lubricants are generally formulated as solutions consisting of inorganic ions (0.9% NaCl) and polymers to increase wettability and retention time, i.e. they additionally act to replace one of the actions of the mucin component of the tears. Commonly used polymers include polyvinyl alcohol and semisynthetic celluloses. Methylcellulose, hydroxypropyl methylcellulose and hydroxycellulose are still widely used, although polyvinyl alcohols have additional surfactant properties to stabilize the tear film for longer periods. Recently, hyaluronic acid has been used, which as a hygroscopic molecule has much greater water retaining capacity, and thus retention time, than celluloses or polyvinyl alcohols and improves tear film stability. Polyacrylic acid (carbomer) is a gel that is hydrophilic, helping it to form a stable tear film, increasing retention time and thus enabling it to be used less frequently.

Artificial tears are now more commonly available without preservatives. Preservatives affect corneal epithelial stability and thus ability to maintain the precorneal tear film.

Mucolytics

Acetylcysteine, which is a derivative of the amino acid L-cysteine, has been used as a mucolytic agent, dissolving the mucus threads that occur in keratoconjunctivitis sicca, but it

does not appear to effect corneal wetting so it has to be used in conjunction with other artificial tear preparations.

Ocular drugs and the autonomic nervous system

Parasympathetic and sympathetic divisions of the autonomic nervous system supply both ocular and extraocular tissues vital for normal ocular function. Agents that influence neurotransmission in the autonomic system are extremely important in the diagnosis and management of many ophthalmic disorders.

The parasympathetic system

Acetylcholine is the major neurotransmitter of the parasympathetic system. It is formed enzymatically from choline and acetyl coenzyme A in the nerve endings (Fig. 6.6). Acetylcholine acts on two types of receptors: muscarinic, which are situated at the effector organ (postganglionic), and nicotinic, which are situated at ganglion synapses and also at neuromuscular junctions. In the eye such receptors are found in the motor endplates of extraocular muscles and levator palpebrae superioris, sympathetic and parasympathetic ganglia, iris sphincter, ciliary body and lacrimal glands. The nicotinic–acetylcholine receptor complex consists of two pairs of polypeptide chains and an additional single polypeptide chain (250,000 Da), providing a hydrophilic channel through which ions can traverse the lipid bilayer. This is known as a *ligand-gated ion channel*. As the acetylcholine binds to the receptor in the synaptic cleft the configuration of the polypeptide chains alters, allowing an influx of sodium and potassium ions down the concentration gradient, thus depolarizing the motor endplate. Cobratoxin, bungarotoxin and tubocurarine (active ingredient of curare) block acetylcholine receptors by preventing the opening of these ion channels. Following the release of acetylcholine, it is largely metabolized by hydrolysis, either within the synapses by acetylcholinesterase or it is absorbed into plasma and then hydrolysed by butyrylcholinesterase. Drugs can act as agonists or antagonists at either muscarinic or nicotinic receptors.

Parasympathomimetics mimic the action of acetylcholine

Parasympathomimetics are a group of drugs that act either by directly stimulating the muscarinic receptor, for example pilocarpine, or by inhibiting the enzyme acetylcholinesterase, which hydrolyses the acetylcholine in the synapse.

Pilocarpine is used in the treatment of chronic open-angle glaucoma, facilitating aqueous drainage via its miotic action on the iris and contraction of the longitudinal muscle of the ciliary body. This draws on the scleral spur and opens the uveotrabecular meshwork. Parasympathetic stimulation has been shown experimentally to reduce aqueous outflow resistance with concomitant breakdown of the blood–aqueous barrier. Pilocarpine thus not only reduces outflow resistance but also reduces the rate of aqueous secretion. In addition, pilocarpine blocks the uveoscleral drainage route of aqueous (see Chapter 4, p. 224) so that drainage is confined to the canal of Schlemm. The ocular hypotensive effect of pilocarpine is further achieved by a reduction in blood flow within the ciliary body, which in turn reduces aqueous secretion (Box 6.13).

The side-effects of direct muscarinic agonists include conjunctival toxicity, iris cysts, cataracts and systemic absorption, which can give rise to muscarinic stimulation resulting in sweating, salivation, vomiting and bradycardia.

Parasympathetic antagonists affect the pupil and ciliary muscle separately and are exploited therapeutically

These drugs block the effect of acetylcholine at muscarinic receptor sites. Their principal effects are those of mydriasis (dilatation of pupil) and cycloplegia (inhibiting ciliary body muscle contraction), but additional effects also include reduction of lacrimal secretions (Table 6.1).

They are used clinically in the management of iritis (preventing adherence of the iris to the anterior capsule of the lens called posterior synechiae), for cycloplegic refraction, to facilitate routine fundal examination and for the provocation test in narrow-angle glaucoma. The side-effects of this group of drugs are seen more often with the longer-acting drugs, for example atropine sulphate. Systemic absorption occurs, particularly with atropine, via conjunctival vessels and via nasal and nasopharyngeal mucosal absorption. Systemic side-effects include dry mouth, facial flushing, sweating and tachycardia. Ocular side-effects include hypersensitivity reactions and conjunctival hyperaemia, blurring of vision and photophobia, a transient rise in intraocular pressure and, occasionally, precipitation of acute-angled glaucoma.

The sympathetic system

Acetylcholine, adrenaline (epinephrine) and noradrenaline (norepinephrine) all act as neurotransmitters within the sympathetic autonomic nervous system (Fig. 6.7).

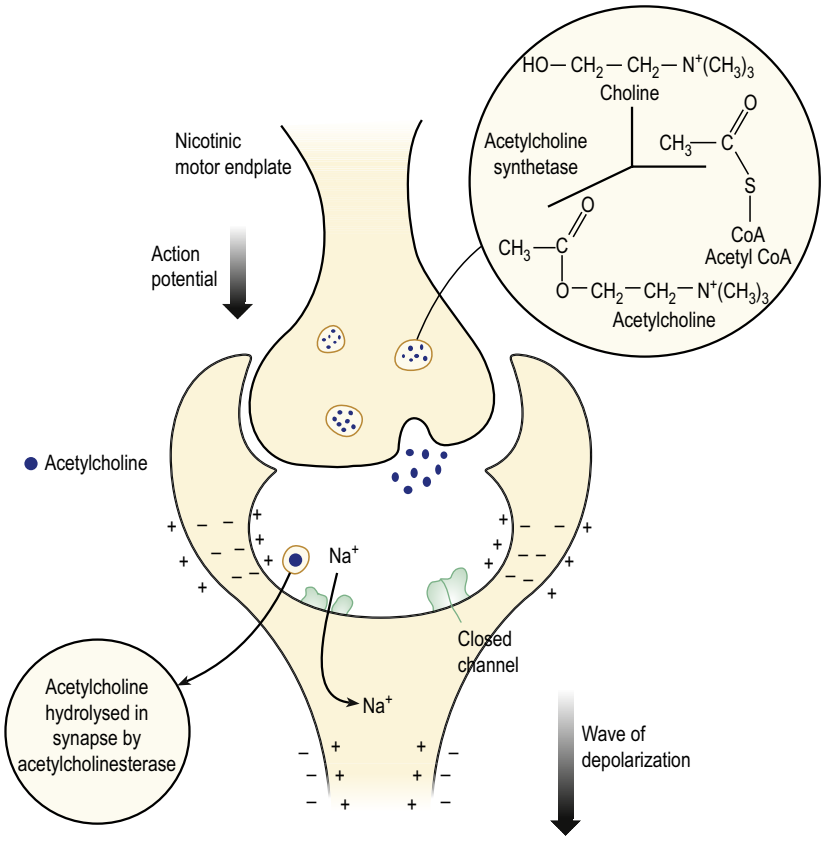


Fig. 6.6 The parasympathetic system: motor endplate.

Box 6.13 **Using parasympathomimetics in diagnosis**

Edrophonium (a competitive inhibitor of acetylcholinesterase) is used in the diagnosis of ocular or systemic myasthenia gravis (tensilon test). After intravenous administration, any improvement in ptosis or diplopia confirms the diagnosis. Longer-acting anticholinesterases (neostigmine) can be used to maintain some neuromuscular function. Immunosuppressive therapies are also used to treat myasthenia because they impair the autoimmune response generated by the acetylcholine receptor antibodies.

Low concentrations of pilocarpine are used in the diagnosis of Adie's tonic pupil, by manifesting the miotic response of the supersensitive iris sphincter caused by loss of postganglionic nerve fibres in this condition.

Table 6.1 **Antimuscarinic agents**

Drug	Maximal mydriasis/ cycloplegia (minutes)	Duration
Atropine sulphate	40	7–10 days
Homatropine	40	1–2 days
Cyclopentolate	30	12–24 hours
Tropicamide	20	3–4 hours

Acetylcholine is the neurotransmitter at autonomic ganglia, which includes the sympathetic nervous system. Adrenaline and noradrenaline are generated from hydroxylation of tyrosine to form DOPA and further modified enzymatically to form dopamine and finally adrenaline and noradrenaline (Fig. 6.8). Once released from the

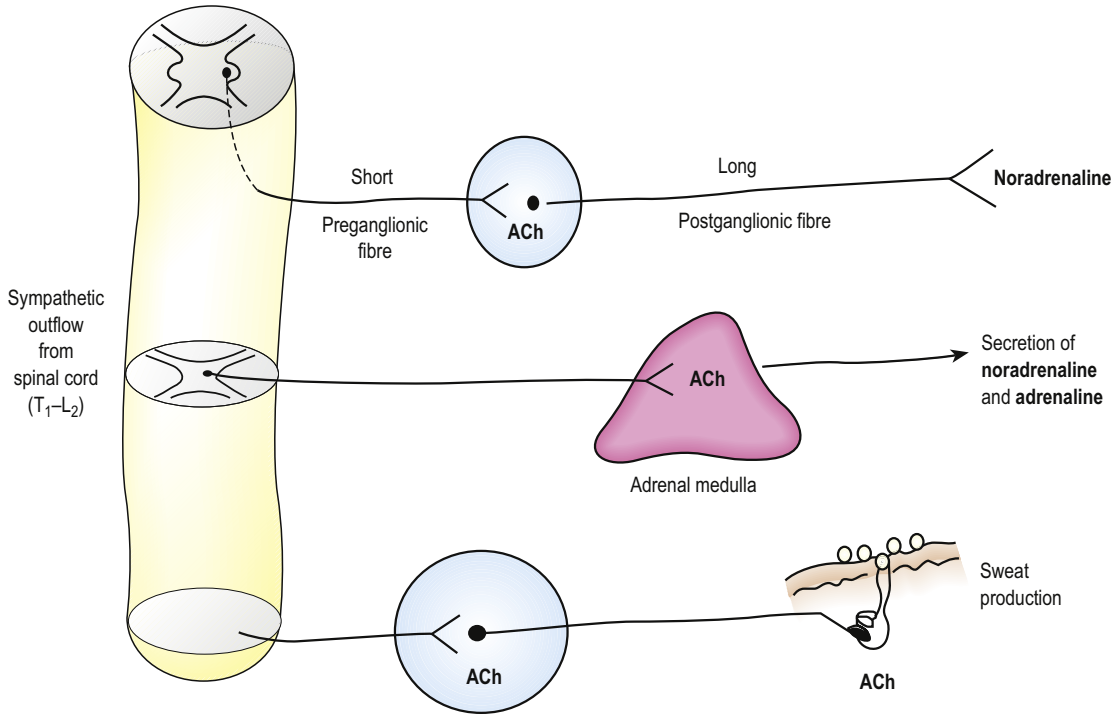


Fig. 6.7 The sympathetic nervous system. *ACh*, Acetylcholine.

nerve ending, they act on several well-defined subtypes of receptors. The action of the neurotransmitter is terminated by enzyme degradation with monoamine oxidases and catechol *O*-methyltransferase and by active reuptake into the nerve endings (Fig. 6.9). Understanding of sympathetic outflow facilitates the diagnosis of Horner's syndrome.

Further information on Horner's syndrome is available online at https://expertconsult.inkling.com

The control of intraocular pressure is facilitated by similar and opposite effects of adrenergic receptors

The control of intraocular pressure (IOP) and aqueous production and outflow has already been discussed (see Chapter 4, p. 228) and is regulated through α_1-, α_2-, β_1-, and β_2-agonists mediating their function through specific receptors (see Box 6.14). It is well documented that certain α-agonists cause lowering of the IOP, for example clonidine, which is principally mediated by an α_2-agonist action. On the other hand, α_1-agonists (e.g. phenylephrine) cause a rise in IOP with concomitant pupillary dilatation. As α_2-agonists produce an immediate fall in IOP, their effect is likely to be central, mediated by stimulation of the medullopontine sympathetic centre. Conversely, α_1-hypertensive effects are thought to be muscular in origin because the effect is antagonized by muscle relaxants.

β-receptors are the predominant receptor on ciliary epithelium, stimulation of which will reduce IOP experimentally. However, β-blockers (e.g. timolol) also reduce IOP. In humans, stimulation of β-receptors does not cause a reduction in aqueous secretion. Indeed adrenaline (a combined α- and β-agonist) causes an initial rise in IOP that is also associated with an initial rise in aqueous secretion. This effect is mediated by concomitant α_1-receptor stimulation, the hypotensive effect being a purely α_2-agonist action. Experimental evidence suggests that direct application of cAMP also increases aqueous drainage, and adenylate cyclase, which is also activated by β_2-agonists, produces a similar response. Certainly, a rise in aqueous cAMP levels correlates well with reduction of IOP, which may be mediated by both α- and β-agonists, but as yet the exact mechanism is unknown (Box 6.15).

Further detail on adrenergic agonists and antagonists is available online at https://expertconsult.inkling.com

HORNER'S SYNDROME

Horner's syndrome is important to recognize given that it can indicate life-threatening disease, for example Pancoast's syndrome (apical lung tumour) or dissecting aortic aneurysm. The signs are a result of interference with sympathetic nervous system and anatomical diagnosis of location of pathology can be assisted with response to pharmacological block of sympathetic outflow.

Classical triad of symptoms

- Partial ptosis (loss of sympathetic innervation of Müller's muscle)
- Anhidrosis (decreased sweating affecting side of face)
- Miosis (loss of sympathetic tone to iris).

Making a diagnosis

Even with the advent of medical imaging to locate the lesion precisely within the sympathetic outflow pathway, diagnosis may be confirmed and then localized via pharmacological manipulation. The site depends on where the sympathetic outflow is perturbed.

- First-order neurones: central lesions involving the medulla oblongata where autonomic centres reside (e.g. as seen in lateral medullary syndrome)
- Second-order neurones: where preganglionic fibres are involved with compression of sympathetic chain (Pancoast's syndrome)
- Third-order neurones: where postganglionic lesions at level of internal carotid artery (and within cavernous sinus) are affected.

The use of drops

- Cocaine prevents the reuptake of noradrenaline, resulting in pupil dilatation. Therefore, in Horner's syndrome with the loss of sympathetic drive, there is a failure of cocaine-induced pupil dilatation.
- Hydroxyamphetamine 1% will cause neurotransmitter release from the postganglionic fibres (third-order neurones). Therefore, in Horner's syndrome affecting the preganglionic fibres, amphetamine will cause normal pupil dilatation, whereas if there is disruption of postganglionic fibres there will be no effect as there will be no noradrenaline to release.
- Both apraclonidine and phenylephrine drops will dilate postganglionic Horner's syndrome as a result of denervation hypersensitivity.

AGENTS ACTING ON THE SYMPATHETIC (ADRENERGIC) SYSTEM

Adrenergic agonists

- *Adrenaline* stimulates both α- and β-receptors. Adrenergic receptors are found in cell membranes of iris dilator muscle, Müller's muscle, ciliary process epithelium and smooth muscle of ocular blood vessels. Phenylephrine is a synthetic sympathomimetic that acts directly on α-receptors. Mydriasis occurs within 1 hour and lasts 4–6 hours. The mydriasis of phenylephrine can be overcome by the powerful parasympathetic light reflex during fundoscopy. Thus, for maintained pupillary dilatation, a combination of phenylephrine and an antimuscarinic, for example cyclopentolate, is used. Systemic absorption is common with phenylephrine and care must be taken when using it in the elderly, in patients with hypertension or in neonates and children.
- *Hydroxyamphetamine* acts by releasing noradrenaline from the nerve terminals. It is therefore used in the diagnosis of postganglionic Horner's syndrome.
- *Cocaine*, whose membrane-stabilizing effect gives rise to its local anaesthetic properties, also prevents the reuptake of noradrenaline into the nerve endings and is used in the diagnosis of pre- or postganglionic Horner's syndrome in cases of anisocoria.
- *Apraclonidine hydrochloride* is a selective α-adrenergic agonist acting mainly on α_2-receptors. It can be applied topically for the prevention and management of raised IOP after anterior segment laser treatment. It lowers IOP by decreasing aqueous humour formation. It does not penetrate the blood–aqueous barrier easily and therefore has minimal systemic side-effects, unlike clonidine.
- *Brimonidine* is a selective α_2-receptor antagonist, which over longer periods can reduce IOP without significant cardiovascular or pulmonary function effects.
- *β_2-adrenergic agonists* also lower IOP, but their effect is mainly to increase the uveoscleral outflow and they are often used in combination with β-blocking agents.

Commercially available L-adrenaline (Eppy 1%) can produce local irritation, allergy and adrenochrome deposits in the subconjunctiva. These adrenochrome deposits are most likely to be oxidative products of adrenaline and are most common in the palpebral conjunctiva. Dipivalyl adrenaline (Propine® 0.1%) is also used in the treatment of glaucoma. It is a lipophilic prodrug of adrenaline and can penetrate the anterior chamber, where it is hydrolysed into the active agent. If adrenaline or its prodrug analogues are used to treat aphakic glaucoma, there is a high incidence of cystoid macular oedema. This macular toxicity has been reported to occur in 20%–30% of aphakic eyes undergoing adrenaline treatment. The mechanism of cystoid macular oedema is unknown, but the angiographic features are identical to those of other conditions causing macular oedema, such as Irvine–Gass syndrome (vitreous–wound adherence following cataract surgery). Experimental evidence has shown that the high incidence of cystoid macular oedema in aphakic eyes is the result of higher concentrations of adrenaline in the retina of aphakic eyes (eTable 6.1).

Adrenergic antagonists

Since the introduction of timolol maleate, topical β-blockers have become the most important therapeutic agents in the medical treatment of glaucoma. The lowering of IOP is a response to a blockade of β-adrenergic receptors in the ciliary epithelium, which produces up to 50% reduction in aqueous production. The site of action probably resides within the ciliary body, but it is not known whether adrenergic antagonists specifically affect ciliary body perfusion or the pumping mechanism of the ciliary epithelium. β-blockers certainly interact with β-adrenergic receptor-coupled adenyl cyclase of the ciliary epithelium. Clinically, there are now many β-blockers available, which vary according to their β-receptor selectivity. However, because of the opposing pharmacological actions of adrenaline and β-blockers, these drugs do not have an additive effect in lowering IOP. Timolol maleate is the most commonly used topical β-blocker and is non-selective, acting on both β_1- and β_2-receptors. Timolol binds reversibly with β-receptors and can reduce IOP in healthy human eyes by reducing aqueous production by 15%–48%. Betaxolol hydrochloride is a lipid-soluble β_1-antagonist that is relatively cardioselective. It reduces IOP by decreasing the production of aqueous at the ciliary body, with no effect on aqueous outflow. Carteolol hydrochloride is a β-blocker that possesses intrinsic sympathomimetic activity and is thought to be beneficial in glaucoma because it not only lowers IOP but also increases optic nerve head perfusion. Its intrinsic sympathomimetic activity does not appear to confer any protection from systemic side-effects of non-selective β-blockade. New β-blockers are frequently under investigation for their effectiveness in reducing IOP and preventing the progression of glaucomatous visual field loss. The ideal β-blocker would block β-receptors in the eye without having any effect on systemic β-receptors.

Systemic side-effects of agonist agents include hypotensive episodes, cardiac arrhythmia, headache and anxiety. The systemic effects of β-blockade include bronchospasm, bradycardia, syncopal attacks and central nervous system depression. It is still not confirmed whether systemic absorption of β-blockers raises levels of serum triglycerides and cholesterol, as with systemic β-blockade, which is of particular importance when treating patients with ischaemic heart disease or more widespread vascular disease.

eTable 6.1 **Sympathomimetic drug receptor profile/action**

Sympathomimetic	Receptor/action
Noradrenaline	α and β
Adrenaline	α and β
Phenylephrine	α (non-selective)
Brimonidine	Selective α_2-agonist
Apraclonidine	α_2 partial agonist
Hydroxyamphetamine	Releases noradrenaline from nerve endings; inhibits monoamine oxidases
Cocaine	Inhibits uptake of noradrenaline at nerve endings

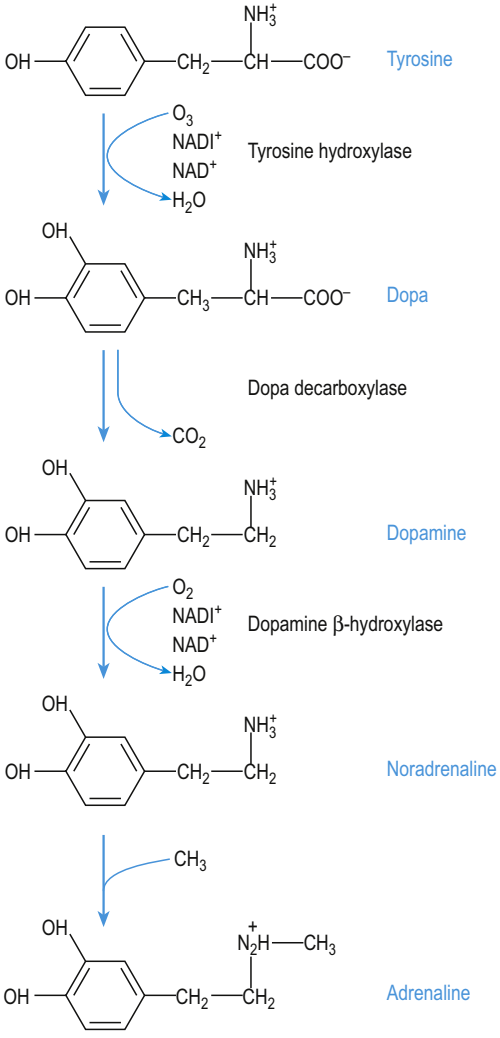
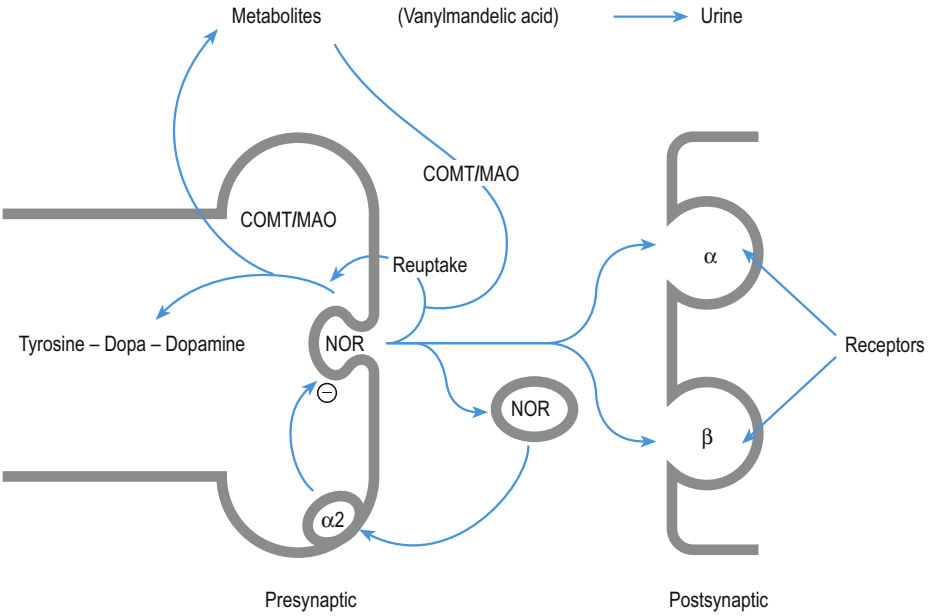


Fig. 6.8 Synthesis of noradrenaline (norepinephrine) and adrenaline (epinephrine).



COMT Catechol *O*-methyltransferase
MAO Monoamine oxidase
NOR Noradrenaline

Fig. 6.9 Adrenergic receptors and metabolism of adrenergic neurotransmitters.

Box 6.14 Adrenergic receptors

Adrenergic receptors are classes of G-coupled protein receptors of two types, α and β; further subtypes exist within each group, classified according to the elicited response and distinguished by selective agonists and antagonists.

- α_1-receptors mediate excitatory responses: principally smooth muscle contraction.
- α_2-receptors are located mostly on the presynaptic nerve endings and are inhibitory because, when stimulated, they prevent further release of the neurotransmitter from the presynaptic terminal (Fig. 6.9).
- β_1-receptors situated in the heart are excitatory and give rise to a positive inotropic and chronotropic response.
- β_2-receptors can mediate inhibitory responses, relaxing the smooth muscle of blood vessels and bronchi.
- β_3-receptors are located in adipose tissue and related to thermogenesis and lipolysis.

Clinical control of intraocular pressure exploits the adrenergic system

Control of IOP is the main aim of therapy in primary open-angle glaucoma, a disorder that contributes greatly to the demands on eye health care systems. IOP control may be achieved surgically or medically and numerous topical therapies based on the pharmacology of the autonomic nervous system are in use, as described previously.

However, medical treatment of glaucoma with respect to control of IOP may also be achieved by systemic or local inhibition of carbonic anhydrase, an enzyme central to the formation of aqueous humour (see Chapter 4, p. 231), enhancing outflow pathways of aqueous flow or protecting the neurones (Box 6.16). The common drugs used topically are summarized in Box 6.17.

The ocular effects of adrenergic agonists include mydriasis, slight ciliary muscle relaxation (probably not relevant in humans), increased formation and increased outflow of aqueous humour (although a stimulation will decrease aqueous humour production), contraction of Müller's muscle and constriction of conjunctival and episcleral vessels (Table 6.2).

Box 6.15 Adrenergic antagonists

Non-selective β-blockers (β_1 and β_2)

- Carteolol
- Levobunolol
- Timolol

Selective β_1-blockers

- Betaxolol

Box 6.16 Protecting nerve fibres without lowering IOP: neuroprotection

- *Glutamate-receptor blocking agents:* reduce excessive calcium influx and ATP synthesis inhibiting excitatory cell death. Agents may act at different sites (receptors) inducing amino acid-induced excitotoxicity.
- *NMDA (N-methyl D-aspartate) receptor inhibition*: NMDA is a postsynaptic ligand-gated ion channel requiring NMDA and glycine for activation. Antagonists can either be competitive or non-competitive, both of which ultimately reduce glutamate secretion and reduce excitotoxicity.
- *Nitric oxide synthase inhibition*
- *Antioxidants*

More detail on ocular carbonic anhydrase is available online at https://expertconsult.inkling.com

A drug-free approach to treating glaucoma

There has been a renaissance in developing minimally invasive glaucoma surgery (MIGS) through the engineering of microstents acting as aqueous shunts. There are broadly three classes. The trabecular shunts are placed into the trabecular meshwork between the iris and cornea during concomitant cataract surgery. The suprachoroidal shunts are larger and drain to the suprachoroidal space and placed between the iris and cornea (the Cypass implant). Finally, subconjunctival devices are inserted through the trabecular meshwork and drain directly into the subconjunctival space (the Xen implant). While this approach to glaucoma treatment attempts to avoid both the complications of conventional surgery (trabeculectomy +/– antimetabolites +/– drainage tubes) and the side-effects of pharmacologic

eLink 6.5 CARBONIC ANHYDRASE IN THE EYE IS A TARGET ENZYME FOR DRUG ACTION

Aqueous humour is secreted actively into the posterior chamber by the non-pigmented epithelium of ciliary processes, which in turn is dependent on the active transport of sodium using the sodium/potassium ATPase pump. Aqueous secretion can be decreased by inhibiting bicarbonate formation, an essential component in aqueous production. The production of bicarbonate is catalysed by the ubiquitous enzyme carbonic anhydrase, which exists in at least six isoforms. Carbonic anhydrase catalyses the hydration of carbon dioxide to bicarbonate, which then dissociates to form hydrogen ions and bicarbonate. The subtypes of carbonic anhydrase enzymes located in ocular tissues include carbonic anhydrase isoenzyme II, which is found in the ciliary body, and isoenzyme IV, which is thought to be a membrane-bound fraction of the enzyme found in the apical region of the RPE cells. Inhibition of carbonic anhydrase reduces aqueous production. This may be a direct effect of enzyme inhibition or secondary to the altered intracellular pH and blockade of the sodium/potassium ATPase ion channel. The effect of acetazolamide therapy is perhaps somewhat surprising. Carbonic anhydrase has a high turnover and the drug must not only reach the active site but also remain in high enough concentrations to have a sufficient duration of action. The drug binds avidly to red blood cell carbonic anhydrase and, when saturation is reached, the drug may effectively distribute to other tissue-binding sites, e.g. ciliary processes and proximal tubule of the kidney. Carbonic anhydrase inhibitors are useful in selected cases of glaucoma. However, these drugs are usually sulphonamide derivatives and can have severe systemic side-effects, including potassium depletion, dermatitis, renal stones (the incidence is said to be 11 times higher than normal), acidosis and, most commonly, fatigue and paraesthesia of the extremities, which can make up to 50% of patients intolerant to its long-term use. Acetazolamide is the most frequently used agent and recently a new sustained-release preparation has become available that is said to reduce the incidence of systemic side-effects. Dorzolamide is a topical carbonic anhydrase inhibitor used to lower IOP in patients resistant or intolerant to β-blockers. It irritates the corneal and conjunctival surface and systemic absorption can occasionally give rise to sulphonamide-like side-effects.

Box 6.17 **Actions of topically applied drugs that lower IOP**

- *β-blockers* (timolol, betaxolol, levobunolol, carteolol) decrease aqueous production by up to 50% and thus lower IOP, even in healthy eyes. Their precise action remains unknown but they are likely to affect either vascular perfusion of the ciliary body or adenyl cyclase of the ciliary epithelium.
- *Parasympathomimetics* (pilocarpine) lowers IOP by facilitating aqueous outflow via its direct action on the scleral spur and ciliary body muscle.
- *$\alpha_2$adrenostimulatants* (brimonidine and aproclonidine) stimulate receptors in the trabecular meshwork (increasing both intracellular and aqueous cAMP levels), which increases the facility of outflow. Stimulation of the α-receptors on blood vessels supplying the ciliary body causes vasoconstriction and reduced blood flow and consequently reduced aqueous production.
- *Prostaglandin analogue* (latanoprost, travoprost) is a prostaglandin $F_{2\alpha}$ analogue and a prostanoid FP receptor (Gprotein-coupled receptor) agonist, which reduces IOP by increasing the outflow of aqueous humour, mainly ureoscleral outflow.
- Carbonic anhydrase inhibitors (systemic: acetazolamide or topically, brinzolamide, dorzolamide).

Table 6.2 **Profile of adrenergic receptors in the eye**

	α	β
Iris dilator	+++	+/–
Iris sphincter	+/–	+
Ciliary process epithelium	+/–	+++
Conjunctival blood vessels	+	+
Müller's muscle	+	–

drugs, their effectiveness for long-term control of IOP in glaucoma patients is still being evaluated.

The histaminergic system: histamine is released from conjunctival mast cells during allergic reactions

Histamine is synthesized and stored in most tissues. It is derived from the amino acid histidine and excreted in the urine after being enzymatically degraded in the liver by histaminase. Histamine is a modulator of the inflammatory response, particularly in allergic type I hypersensitivity reactions (see Chapter 7, p. 448). Histamine also plays an integral role in neurotransmission, for example regulating gastric acid secretion. Four common histamine receptors (7 transmembrane G protein-coupled receptors) are described, H_1–H_4 on the basis of the structure of the specific antagonist that binds to them (Box 6.18).

H_1-receptors are found in abundance in human bronchial muscle and at many other sites, including the CNS. The histamine receptors in these tissues have the same affinity for histamine as for the histamine-competitive antagonist cetirizine. H_2-receptors are found in the stomach, heart and uterus. The receptors involved have a common affinity for the competitive histamine antagonist cimetidine. Activation of H_1-receptors results in an increase in intracellular calcium concentration and activation of H_2-receptors results in stimulation of adenyl cyclase and second messenger production (Box 6.19).

Mast cells are an abundant source of histamine. Histamine release is mediated by allergen-induced immunoglobulin E (IgE) hypersensitivity responses. Control of this allergic response can be obtained by preventing mast cell degranulation at the mucous membranes with mast cell stabilizing agents. This is of particular relevance when considering the management of allergic eye disease (Box 6.20).

Antihistamines

H_1 antihistamines inhibit histamine-induced contraction of smooth bronchial muscles and increased vascular permeability caused by histamine. Some H_1 antagonists have pronounced central nervous system side-effects, including drowsiness, but also have the benefit of being antiemetics. Recently introduced H_1 antihistamines have markedly reduced sedative or anticholinergic action and do not readily cross the blood–brain barrier when administered systemically.

H_2 antagonists are effective in the pharmacological control of gastric acid secretion by decreasing basal and food-stimulated acid secretion (up to 90%). The two drugs commonly used are cimetidine and ranitidine, which are structural analogues of histamine. Cimetidine also inhibits cytochrome P_{450} and decreases the metabolism of drugs (e.g. anticoagulants, phenytoin and aminophylline), thus potentiating their effects. Cimetidine binds to androgen receptors, sometimes leading to gynaecomastia and reduced sexual function.

Sodium cromoglicate

Disodium cromoglicate is used in the treatment of allergic hypersensitivity reactions and is administered topically, either into the conjunctival fornix or by inhalation to the bronchial mucosa. It acts by inhibiting the release of histamine and slow-releasing substance of anaphylaxis (SRS-A)

Box 6.18 Histamine receptor antagonists

H_1: diphenhydramine, loratadine, cetirizine
H_2: ranitidine, cimetidine
H_3: ciproxifan
H_4: thioperamide

Box 6.19 Actions of histamine

H_1 actions

- Increases vascular permeability and vasodilatation of arterioles
- Arteriolar dilatation of superficial skin vessels (axon reflex)
- Capillary dilatation and oedema of dermis
- Smooth muscle contraction/bronchospasm
- Increases mucus secretion
- Central nervous system depressant

H_2 actions

- Increases pepsin and acid production
- Increases myocardial stroke volume

H_3 actions

- Receptors associated with neural tissue at presynaptic sites, which, when stimulated, inhibit histamine release, the significance of which is unclear

H_4 actions

- Mediates mast cell chemotaxis and calcium mobilization in mast cells

Box 6.20 Treating allergic eye disease

Topical antihistamines

- Azelastine
- Levocarbastine

Mast cell stabilizers

- Sodium cromoglicate
- Lodoxamide

from the mast cells within the mucosa, by stabilizing mast cell membranes. It must therefore be administered before mast cell priming with IgE and allergen because it will have no effect once the mast cells have degranulated. Lodoxamide has also been developed as a topical application to prevent mast cell degranulation, alongside increasing numbers of topical choices in this class of agent.

Eicosanoids affect multiple ocular functions

Eicosanoids are not found preformed in tissues but are generated *de novo* from cellular phospholipids after a wide range of stimuli. They are important mediators of the inflammatory response, and the non-steroidal anti-inflammatory drugs (NSAIDs) owe some of their activity, especially prostaglandin $F_{2\alpha}$, to the inhibition of the synthesis of eicosanoids. The name prostaglandin derives from reports that semen contained a substance that contracted the uterus and was thought to be derived from the prostate gland. However, the first prostaglandins were described in the eye, although not named as such.

The principal eicosanoids are the prostaglandins, thromboxanes and leukotrienes (see Chapter 7, p. 404). The main source of these substances is from the 20-carbon unsaturated fatty acid arachidonic acid, which is found esterified in the phospholipids of cell membranes. Fig. 6.10 describes the generation of these molecules, where the initial rate-limiting step is the generation of arachidonate by phospholipase A_2 or C. Stimuli that liberate these enzymes include thrombin in platelets, C5a from complement and bradykinin, as well as general cell damage. The free arachidonic acid is then metabolized via two pathways mediated by the enzymes cyclo-oxygenase, generating prostaglandins, and lipoxygenase, generating leukotrienes. The anti-inflammatory action of the NSAIDs is mainly the result of the inhibition of cyclo-oxygenase, and thus prostaglandin synthesis (Box 6.21).

Each prostaglandin appears to act on specific receptors that have yet to be fully characterized. However, receptor antagonists are gradually being developed for each eicosanoid. A recent example is misoprostol, which reduces gastric acid secretion and gastric erosion by NSAIDs, by the antagonism of prostaglandin E_2 class 1 receptors.

Aspirin, among other compounds, inactivates cyclo-oxygenase and reduces inflammation and pain. Other systemically administered NSAIDs, such as indomethacin and flurbiprofen, are used in an attempt to reduce the inflammatory response in uveitis and scleritis. Indomethacin is also used to reduce cystoid macular oedema, but its effects have yet to be substantiated. Both ibuprofen and diclofenac are now available as topical applications; their principal use is to prevent perioperative miosis during cataract surgery and to reduce postoperative inflammation.

More recently, prostaglandin analogues have been studied for use in primary open-angle glaucoma, e.g. prostaglandin $F_{2\alpha}$.

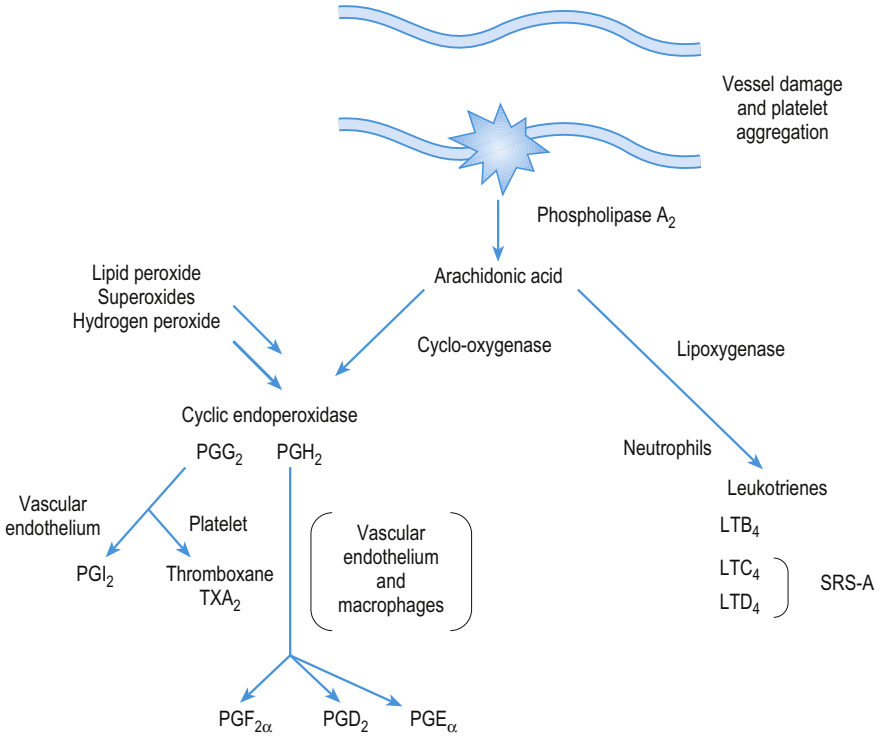


Fig. 6.10 Synthesis of eicosanoids.

Box 6.21 **Biological actions of eicosanoids**

Prostaglandins

- Prostaglandin I_2: vasodilatation, decreases platelet adhesion
- Prostaglandin $F_{2\alpha}$: bronchial smooth muscle contraction
- Prostaglandin E_2: vasodilatation, bronchodilatation, uterine contraction, pyretic, stimulates release of pituitary hormones, adrenal cortex steroids and insulin from pancreas, macrophage activation

Thromboxane

- Thromboxane A_2: vasoconstriction, platelet aggregation, bronchoconstriction

Leukotrienes

- Leukotriene B_4: aggregation of neutrophils, chemotactic, stimulation of phospholipase A_2
- Leukotriene C_4
- Leukotriene D_4: contraction of smooth muscle, bronchoconstriction, vasoconstriction, leukotrienes C_4 and D_4 together form SRS-A.

In an attempt to increase the efficacy of NSAIDs and reduce, in particular, gastric mucosal erosions, newer NSAIDs are being generated that are specific for one of the isoenzymes of cyclo-oxygenase, cyclo-oxygenase 2 (COX-2). COX-2 specificity is said to confer a more specific inhibition of inflammatory prostaglandin synthesis, without impairing synthesis of gastroprotecting prostaglandins, e.g. prostaglandin I_2, which is synthesized by COX-1. Conventional NSAIDs are strong inhibitors of COX-1 and less active against COX-2. However, Cox-2 inhibitors have a poor track record of side-effects, the first two on the market having to be withdrawn because of increased risk of myocardial infarction, leaving only celecoxib as an available anti-inflammatory.

Serotonin: a potent neurotransmitter

Neurotransmitters are integral to retinal and cortical function and drugs that modulate their function may alter visual perception. In addition, some neuromodulatory drugs are used to treat conditions with visual symptoms, such as migraine.

5-Hydroxytryptamine (5-HT; serotonin) is biosynthesized in a manner similar to noradrenaline from the precursor amino acid tryptophan, which is taken up into the nerve endings and converted by tryptophan hydroxylase to 5-hydroxytryptophan, and then decarboxylated to serotonin. Serotonin is degraded by oxidative deamination by the action of a group of enzymes called *monoamine oxidases* to form the aldehydes 5-hydroxyindoleacetaldehyde and 5-hydroxyindoleacetic acid (5-HIAA), analogous to noradrenaline metabolism. The aldehydes are then excreted in the urine and may be detected in hypersecretory conditions such as carcinoid syndrome.

Serotonin receptor subclasses mediate different effects

Serotonin is a neurotransmitter that is widely distributed throughout the body, particularly in platelets, mucosa of the gastrointestinal tract and neurones of both the central and peripheral nervous system. There are four main types of serotonin receptors, designated 5-HT_1, 5-HT_2, 5-HT_3 and 5-HT_4. Although 5-HT is a simple molecule, it has a wide variety of effects in the cardiovascular, gastrointestinal, respiratory and central nervous systems.

5-HT_1-receptors can be further divided into subtypes 5-$HT_{1a–d}$. They are differentiated by the kinetics of agonist and antagonist binding to the receptor, and by their regional distribution within the central nervous system. 5-HT_{1a} has been found in the raphe nucleus and the hippocampus, activation of which causes hypotensive attacks and behavioural changes. This has been studied with particular reference to the physiology of anxiety. 5-HT_{1b} is found in rodents and is thought to have an inhibitory effect on the release of 5-HT, similar to the 5-HT_{1d}-receptor in humans. 5-HT_{1c} has been localized to the choroidal plexus, but no specific agonist or antagonist has yet been identified. The 5-HT_2-receptor is located in the hippocampus, frontal cortex and spinal cord. These receptors are also located on smooth muscle in the bronchus and blood vessels. They have a direct excitatory effect on smooth muscle. 5-HT_3-receptors are found in the nerves along the gastrointestinal tract and peripheral nervous system. 5-HT exerts an excitatory effect through these receptors, particularly excitation of nociceptive nerve endings. They are also found in the limbic and cortical areas of the brain and are thought to play an important role in the development of anxiety and psychotic states. Recently, 5-HT_4-receptors have been discovered in the CNS but their clinical significance remains unclear.

5-HT antagonists

Selective 5-HT_2-receptor antagonists, which are used for the prevention of migraine attacks (e.g. methysergide and pizotifen), probably act by inhibiting the release of serotonin at the aura stage of the attack. During this stage the activation of 5-HT neurones is at its greatest, and therefore antagonism of the receptors prevents the sequelae of vascular smooth muscle contraction, local inflammation and nociceptor stimulation. Selective 5-HT_3-receptor antagonism (e.g. ondansetron) reduces the 5-HT nociceptive stimulation,and thus the headache of migraine.

Glucocorticoids

Steroids are applied topically to suppress the inflammatory reaction of many conditions, commonly anterior uveitis, postoperative inflammation and corneal graft rejection. They may also be administered subconjunctivally or systemically in more severe intraocular inflammatory conditions. The mechanism of their anti-inflammatory actions is mediated by the drug's effects on both the number and function of lymphocytes, polymorphonuclear leucocytes and macrophages and on vascular permeability. They also affect inflammatory mediators by inhibiting phospholipase A_2, prostaglandin, thromboxane and leukotrienes; they also inhibit histamine release. Despite the pleiotropic and what at first seems significantly powerful anti-inflammatory effects, steroids have many ocular side-effects, including cataract formation, reactivation of viral keratitis or increased incidence of bacterial infection and steroid-induced rises in IOP. The latter side-effect is dependent on the duration and strength (potency) of individual steroid preparations as well as genetic predisposition. The rise in IOP is thought to occur via the accumulation of glycosaminoglycans and water in the trabecular meshwork, reducing aqueous outflow. Thirty per cent of normal subjects have a hypertensive response when challenged with corticosteroids. This can be characterized further into poor responders, moderate responders (heterozygous responders) and strong responders (homozygous responders). If the individual also has open-angled glaucoma, a greater overall response is seen. It is therefore important that IOP is monitored at regular intervals during the course of prolonged steroid therapy.

The anti-inflammatory potency of any particular steroid is dependent on its ability to penetrate the cornea. Increasing the steroid concentration results in higher intraocular concentrations, but this may also be achieved with different formulations of the same parent steroid that increase the contact time of the topically applied drug. Steroids such as prednisolone phosphate are hydrophilic and therefore penetrate the corneal epithelium poorly; in contrast, the acetate forms of both dexamethasone and prednisolone give

rise to a much greater intraocular concentration. To date, it is still not established which concentrations of steroids are desirable for low-grade intraocular inflammatory conditions. If this could be established, ocular side-effects might be kept to a minimum.

Box 6.22 **Immunosuppressants and their mode of action**

- *Corticosteroids* act on cytosolic receptors and block transcription of cytokine genes (e.g. interleukins 1, 2, 3 and 5, tumour necrosis factor-α and interferon-γ)
- *Ciclosporin* acts via inhibition of calcineurin and NFAT, resulting in inhibition of interleukin-2 production
- *Tacrolimus* inhibits interleukin-2; mechanism similar to ciclosporin
- *Azathioprine* inhibits purine synthesis, blocking RNA and DNA synthesis
- *Methotrexate* folic acid antagonist, inhibiting dihydrofolate reductase and suppressing DNA synthesis
- *CellCept* blocks *de novo* pathway of purine synthesis, which is selective for lymphocytes

Immunosuppressant agents: combating ocular inflammatory disease

Controlling immune responses, particularly T-cell responses (Chapter 7, p. 405), that mediate allograft rejection, endogenous uveitis, associated or not with other autoimmune conditions, and chronic allergic eye disease may necessitate systemic therapy with immunosuppressants. In addition to steroids, there is now a significant armamentarium of agents used in ophthalmology to treat such conditions. Such agents have been generated as a result of our increased understanding of T-cell biology, in particular the interactions of T cells with specific antigens and antigen-presenting cells, activation of T cells and the effects of mediators of inflammation, such as cytokines.

Commonly used agents in addition to the traditional use of steroids for suppressing inflammatory responses (Box 6.22) include ciclosporin A, purine antagonists and cytotoxic agents such as methotrexate, and occasionally alkylating agents such as cyclophosphamide. Newer generations of similarly acting agents are now used, especially in

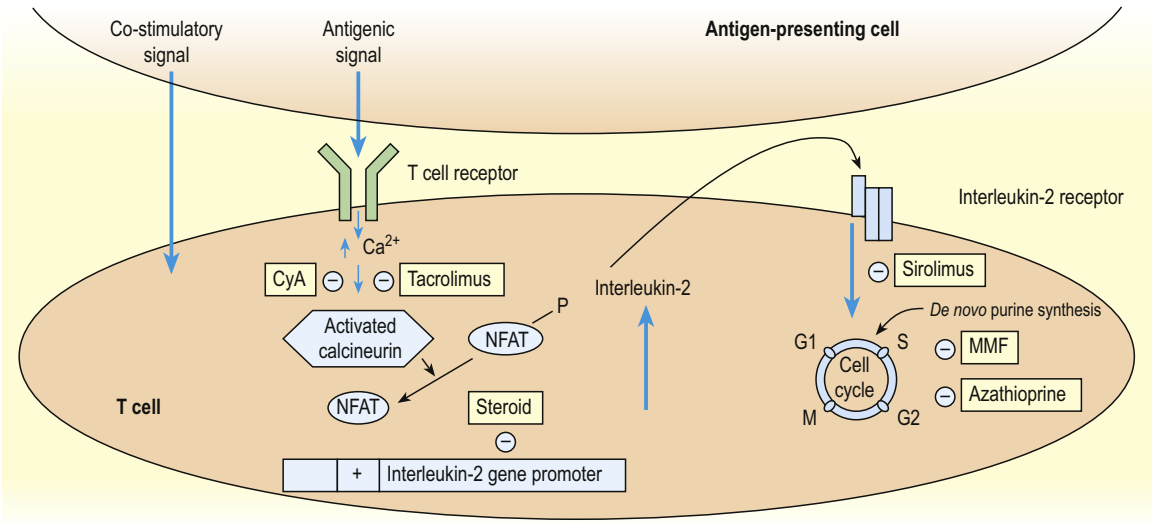


Fig. 6.11 Stages of T-cell activation: multiple targets for immunosuppressive agents. Signal 1: stimulation of T-cell receptor results in calcineurin activation, a process inhibited by ciclosporin *(CyA)* and tacrolimus. Calcineurin dephosphorylates nuclear factor of activated T cells *(NFAT)*, enabling it to enter the nucleus and bind to *interleukin-2 promoter*. Corticosteroids inhibit cytokine gene transcription in lymphocytes and antigen-presenting cells by several mechanisms. Signal 2: co-stimulatory signals are necessary to optimize T-cell interleukin-2 gene transcription, prevent T-cell anergy and inhibit T-cell apoptosis. Experimental agents but not current immunosuppressive agents interrupt these intracellular signals. Signal 3: interleukin-2 receptor stimulation induces the cell to enter the cell cycle and proliferate. Signal 3 may be blocked by interleukin-2 receptor antibodies or *sirolimus*, which inhibits the second messenger signals induced by interleukin-2 receptor ligation. Following progression into the cell cycle, azathioprine and mycophenolate mofetil *(MMF)* interrupt DNA replication by inhibiting purine synthesis. (From Denton et al., 1999, with permission from Elsevier.)

prevention of solid organ allograft rejection, including tacrolimus (FK 506) and mycophenolate mofetil (CellCept®). The mechanism of the action of such agents and its effect on T-cell activation are represented in Fig. 6.11.

Although successful, immunosuppressive therapy is limited by its relative non-specificity, the refractoriness of some patients to therapy and the high incidence of side-effects (Box 6.23). To reduce unwanted systemic effects, some agents, for example ciclosporin A and tacrolimus, are currently being formulated for topical delivery to treat chronic allergic disease and corneal allograft rejection.

Box 6.23 Common side-effects of immunosuppressants

- *Corticosteroids:* osteoporosis, hypertension, glucose intolerance, altered habitus
- *Ciclosporin and tacrolimus:* nephrotoxicity, hypertension, hyperlipidaemia, glucose intolerance, hirsutism and gingival hyperplasia
- *Azathioprine/CellCept:* bone marrow suppression, diarrhoea and gastrointestinal upset

Biologics

The current expertise in molecular biological and engineering technology has permitted over the past decade the generation of specific molecules that can be engineered to target specifically receptors, membrane proteins or soluble proteins. With reference to eye diseases, this has resulted in newer treatments for ocular inflammatory disorders and retinal choroidal angiogenesis such as 'wet' age-related macular degeneration, as discussed earlier. For example, one of the principal mediators of T-cell responses in autoimmune disorders, such as rheumatoid arthritis, inflammatory bowel disease and posterior uveitis, is tumour necrosis factor-α (TNF-α). Engineering either specific antibodies or

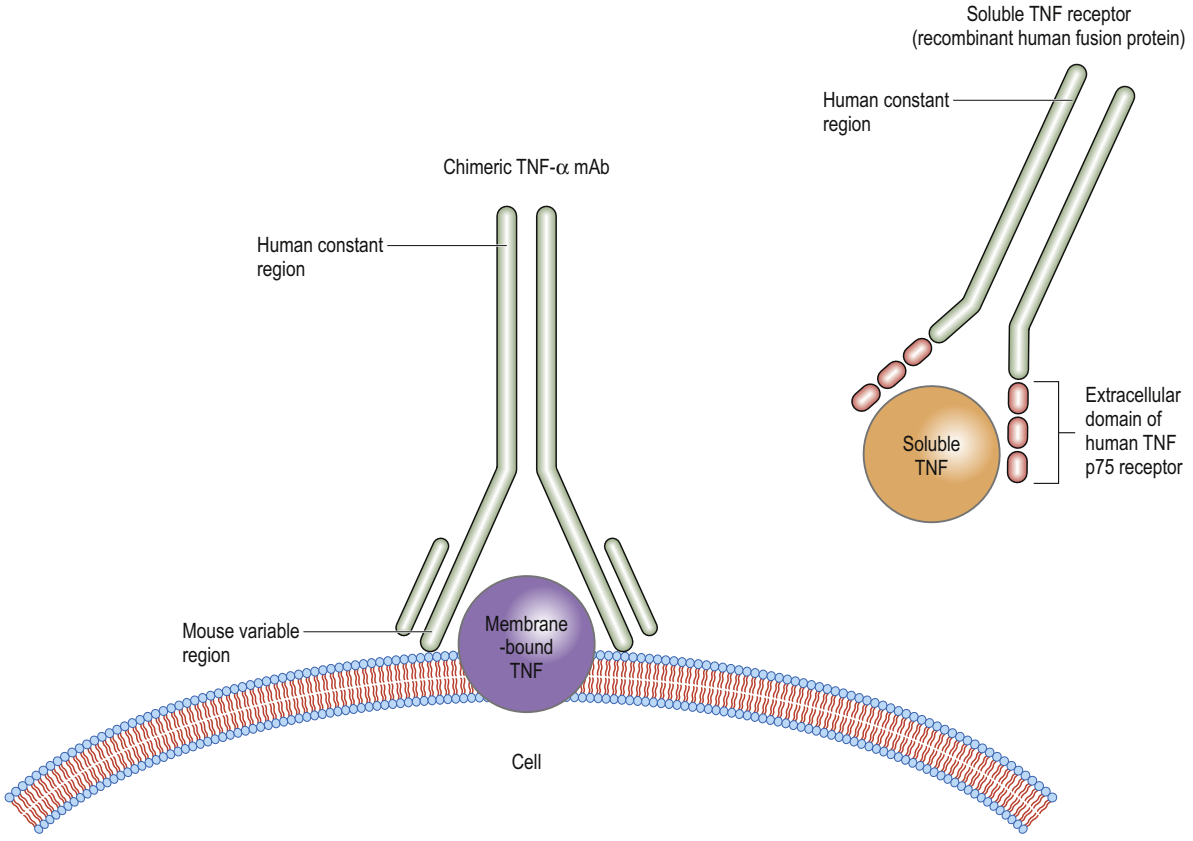


Fig. 6.12 Schematic representation of anti-TNF agents infliximab and the fusion protein etanercept. (From Cochrane and Dick, 2007, with permission from Springer Science and Business Media.)

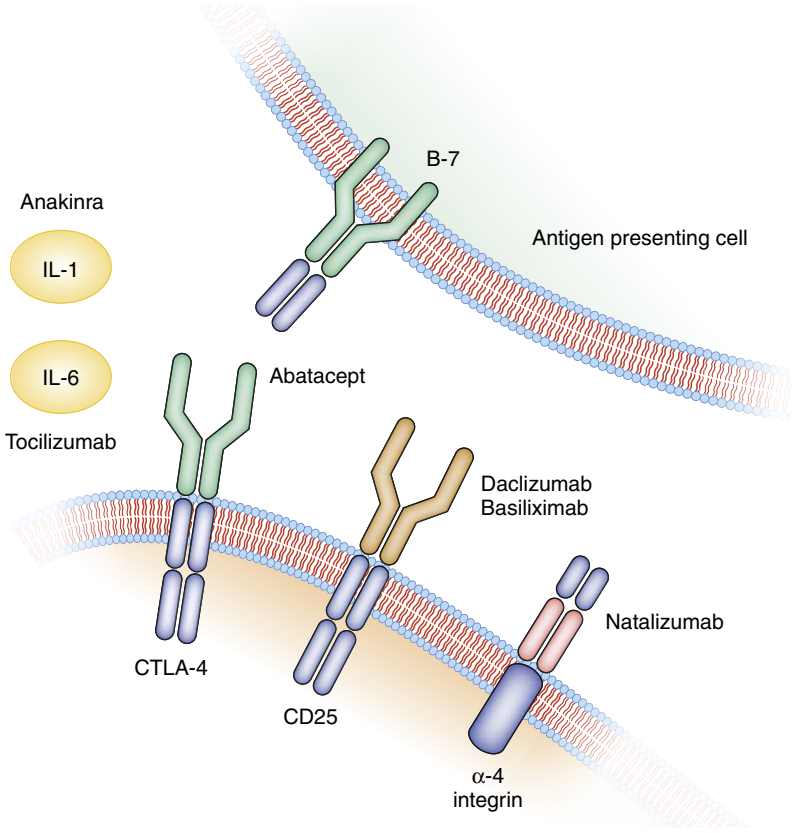


Fig. 6.13 Biologics target key mediators of inflammation. Chimeric or humanized monoclonal antibodies or fusion proteins can deliver potent inhibition of cell response and inflammation. The schematic shows some currently available targets and biological therapies that have been generated and are being brought into practice for the treatment of inflammatory eye diseases.

immunoadhesins (fusion proteins of their receptors bound to a human immunoglobulin tail) can recognize membrane-bound and soluble TNF-α and neutralize its activity. TNF-α binds two receptors (p55-TNFR1 or p75-TNFR2). The commercially available biologics include *infliximab*, a chimeric antibody (human with mouse-derived variable region recognizing TNF-α) or Humira® (adalimumab, a humanized monoclonal antibody) that inhibits TNF-α by neutralizing both membrane-bound and soluble TNF, and *etanercept*, which is a fusion protein of the p75 receptor that successfully binds TNF, thereby preventing further binding (Fig. 6.12). There are now many targets that can be specifically inhibited with monoclonal antibody or other biological therapy (Fig. 6.13). For uveitis, *adalimumab*, a humanized monoclonal anti-TNF mAb is licensed for use in children and adults for non-infectious uveitis. Table 6.3 shows the spectrum of biologics available and their targets.

There is development in small molecule inhibitors of inflammatory pathways including Janus Kinase inhibitors (Tofacitinib), and Sphingosine 1 Phosphatase receptor (Fingolimod).

Local anaesthetics: an integral part of ophthalmic examination and surgery

Local anaesthetics consist biochemically of an aromatic residue linked to an amide or basic side-chain. As such, local anaesthetics are both hydrophobic (aromatic

Table 6.3 **Current biologic monoclonal antibody therapies for treatment of ocular inflammatory disease**

Biologic agent		Target
Anti-cytokines		
Infliximab	Chimeric mAb	Membrane and soluble TNF
Adalimumab	Humanized mAb	Membrane and soluble TNF
Etanercept	Fusion protein of P75 TNF receptor	Soluble TNF
Certolizumab pegol	PEGylated mAb fragment	Membrane and soluble TNF
Golimumab	Humanized mAb	Soluble and membrane bound TNF
Tocilizumab	Humanized mAb	Interleukin (IL)-6
Anakinra	Recombinant IL-1 receptor antagonist	Interleukin (IL)-1
Ustekinumab	Humanized mAb	Interleukin (IL)-12/ IL-23
Secukinumab	Humanized mAb	Interleukin (IL)-17A
Anti-lymphocyte		
Rituximab	Humanized mAb	CD20 B cells
Abatacept	Fusion protein of CTLA4	T cells
Anti-Adhesion/trafficking		
Natalizumab	Humanized mAb	Alpha-4 integrin

TNF, Tumour necrosis factor.

residue) and hydrophilic (amide group) and tend to accumulate at aqueous–non-aqueous interfaces. Because the aromatic residue and side-chain in some local anaesthetics are linked by esters, they are susceptible to metabolic hydrolysis. These compounds are usually inactivated in the liver and plasma by non-specific esterases. As amide links are more stable, this group of anaesthetics has a longer $t_{1/2}$.

Local anaesthetics affect excitable membranes

Local anaesthetics block the initiation and propagation of action potentials by preventing the voltage-dependent increase in sodium conductance via a direct action on sodium channel function and, to a lesser degree, by stabilizing membranes. The action of local anaesthetics is strongly pH dependent. Under alkaline conditions the proportion of the local anaesthetic that is ionized is low but it is also lipid-soluble. This allows the anaesthetic to penetrate the myelin of the medullated nerve fibres. However, once inside the myelin sheath, the ionized cationic form of the drug is the active species, giving rise to the local anaesthetic effect. If the local anaesthetic is unable to penetrate the myelin sheath, its action can be mediated only via the node of Ranvier at several sites along the nerve. However, in unmyelinated fibres only a short length of fibre needs to be functionally interrupted to induce anaesthesia. In general, local anaesthetics block small-diameter nerves (myelinated more than unmyelinated) more readily, but practically it is not possible to produce a block of pain sensation without affecting other modalities, and there is no inherent difference between the susceptibility of motor and sensory nerves to local anaesthetics.

Local anaesthetics used in ophthalmology are usually aromatic residues linked to tertiary amide chains. These agents are more stable in acidic solutions where they are in their cationic form. However, when applied to the conjunctival sac and orbital tissues (pH 7.4), only 15% will be in the non-cationic form and lipid soluble, allowing the drug to penetrate myelin. Topical application of the anaesthetic blocks parasympathetic and sympathetic fibres first, followed by sensory (pain and temperature) and finally motor (large, myelinated) nerves.

Box 6.24 **Maximal safe dose of regional local anaesthetics**

Lidocaine

- 10–15 mL 2% solution (200 mg)
- 20–25 mL 2% solution with adrenaline, 1:200,000 (500 mg)

Bupivacaine

- 15–20 mL 0.75% solution (150 mg)

During infiltrative regional anaesthesia for surgery (e.g. peribulbar or sub-Tenon's anaesthesia for intraocular surgery), a 1:200,000 concentration of adrenaline may be added to constrict blood vessels and retard vascular absorption and hydrolysis of the agent. Adrenaline is inactivated by heat and may therefore have reduced potency in precombined formulations. In addition, the pH of the local anaesthetic may be altered when mixing with an adrenaline solution, which itself is acidic. This results in a reduced amount of the non-cationic form that is able to penetrate medullated nerve fibres.

Local anaesthetics have both local and systemic side-effects

Topically applied anaesthetic agents are well recognized as inhibitors of wound healing. They disrupt tight junctions between cells and interfere with corneal epithelial metabolism and, ultimately, the repair of corneal epithelial wounds. Systemic effects include numbness and tingling, dizziness, slurred speech and aggressive behaviour (Box 6.24). Central nervous system toxicity may ultimately lead to convulsions with respiratory and myocardial depression. In the UK it has been recommended by the Royal College of Ophthalmologists that intravenous access (an infusion line) is available and monitoring of both heart rate and oxygen saturation is performed during infiltrative regional local anaesthesia.

Ocular toxicity from systemic administration of drugs

Ocular side-effects from the systemic administration of drugs are well recognized (see previously in relation to glaucoma treatment and the use of topical steroids). Access to the globe by systemically administered drugs is restricted by the blood–retinal and blood–aqueous barriers, as stated earlier. In the absence of ocular inflammation, penetration of the drug into the eye is a function of the drug's physicochemical properties. Drugs or their active metabolites may accumulate in the eye, particularly within the melanocytes of the uveal tract (bound to melanin), the cornea (because of its differential solubility characteristics) and the lens.

Uveal tract

Some drugs possess a high affinity for binding to melanin, from which they are only slowly released. A good example of this is chloroquine, which is known to concentrate in the melanin of the RPE and persist in this tissue for prolonged periods. If chloroquine is taken in large enough doses and for a long duration, the drug–melanin complex will result in retinal toxicity, although the exact mechanism remains unknown. The mechanism of drug binding to melanin is complex and involves electrostatic (van der Waals') forces as well as possible cation exchange, which may displace free radicals from the melanin and, in turn, give rise to retinal toxicity. However, binding of the drug alone is not sufficient to give rise to retinal toxicity. For instance, β-blockers and benzodiazepines are irreversibly bound to melanin, although the amount of bound drug does not correlate with the damage to the RPE or uveal tract. Indeed, phospholipid metabolism within the RPE is still normal after 6 months of chloroquine treatment. In the case of chloroquine, the retinopathy is seen usually in patients receiving more than 100 g total dose of drug or in patients taking the drug for more than 1 year. In general, there is a low incidence of retinal toxicity from hydroxychloroquine and some centres regard this drug as safe and not requiring monitoring of patients.

Photosensitization

Photosensitizing agents absorb visible and ultraviolet radiation and, as a result, generate free radicals. These photosensitizing agents may become bound to macromolecules in the cornea, lens and retina. Amiodarone, phenothiazines and psoralens are well-known examples of photosensitizing agents. The cornea, lens and retina may also act as drug depots. For example, once the therapeutic agent has circulated through the uveal tract into the aqueous it can rapidly penetrate the corneal endothelium and deposit in the stroma or, if lipophilic, accumulate in the corneal epithelium.

Oculotoxic drugs

Examples of oculotoxic drugs include steroids and ethambutol. Long-term steroid treatment is well recognized as a cause of both cataracts and glaucoma. Cataract formation correlates well with daily dosage and prednisolone-induced cataracts can be seen in most patients receiving 15 mg prednisolone daily. Topical application of steroids causes glaucoma by decreasing the aqueous humour outflow. The rise in IOP is related to the anti-inflammatory strength of the steroid used and the genetic disposition of the individual.

Further reading and references are available online at https://expertconsult.inkling.com

Further reading

Cochrane, S., Dick, A.D., 2007. Tumor necrosis factor alpha-targeted therapies in uveitis. In: Pleyer, U., Foster, C.S. (Eds.), Essentials in Ophthalmology. Uveitis and Immunological Disorders. Springer, Berlin-Heidelberg, pp. 177–192.

Denton, D.M., Magee, C.C., Sayegh, M.H., 1999. Immunosuppressive strategies in transplantation. Lancet 353, 1084.

Gaudana, R., Ananthula, H.K., Parenky, A., Mitra, A.K., 2010. Ocular drug delivery. A. A. P. S. J. 12, 348–360.

Ingelman-Sundberg, M., 2004. Pharmacogenetics of cytochrome P450 and its applications in drug therapy: the past present and future. Trends Pharmacol. Sci. 25, 193–200.

McGhee, C.N.J., 1992. Pharmacokinetics of ophthalmic corticosteroids. Br. J. Ophthalmol. 76, 681–684.

Shen, J., Lu, G.W., Hughes, P., 2018. Targeted ocular drug delivery with pharmacokinetic/pharmacodynamic considerations. Pharm. Res. 35, 217–237. https://doi.org/10.1007/s11095-018-2498-y.

Chapter 7

Immunology

CONTENTS

Introduction

Overview of the immune system

Cells and molecules of the immune system

- The myeloid system and innate immunity
- The lymphoid system and acquired immunity
- Blurring the margins between innate and adaptive immunity

Initial response of the host to injury (the innate immune response)

- The acute inflammatory response
- Resolution of inflammation
- Chronic inflammation
- The systemic response to acute inflammation: the acute-phase reaction

Development of adaptive immunity and immunological memory

- Antigen recognition is made possible by antigen-presenting cells
- T cells respond to antigen by clonal expansion
- T and B cells participate in the effector response
- How does the organism deal with intracellular antigen?

Effector mechanisms

- Antibodies
- Cytokines are the effector elements released by cells during innate and adaptive immune responses
- Complement
- Cellular mechanisms of tissue damage

Organization of the immune system

- Functional anatomy of lymphoid organs
- The mucosal immune system
- The immune system as a police force

Antigen recognition

- The APC makes the antigen recognizable

The major histocompatibility system

- What are MHC antigens and where are they found?
- Organization of the MHC genes in the genome

T-cell activation

- The TCR and antigen binding
- Co-stimulation: presentation of antigen requires 'help' from other molecules

B-cell activation

- Antigens, B cells, and T cells: what's the story?
- B-cell differentiation

Tolerance and autoimmunity

- What is tolerance?
- Sites of tolerance induction
- Mechanisms of tolerance induction
- Failed tolerance
- Auto-inflammatory disease

Allergy and immediate hypersensitivity

- Atopy, asthma and allergic eye disease
- Immunoglobulin and helper T cells
- Mechanism of disease production in allergic disease
- Mast cells also have a protective role

Organ and tissue transplantation

Tumours induce immune responses but not very effectively

The eye and the immune system

- The innate immune system and the eye
- The adaptive immune system and the first line of defence in the eye
- Progressive ocular surface disease

Conclusion

Introduction

Immunology is the study of host defence mechanisms. Even the simplest organisms have the ability to mount a variety of specific and non-specific responses to invasion or attack by foreign organisms. Where the host and pathogen meet and interact is where the immune response is initiated and the nature of both players determines the outcome. Thus, humans can respond to pathogens in many different ways, depending on their genetic make-up, the type of foreign organism (e.g. virus, bacterium, fungus, etc.) and prevailing conditions or the setting in which the host is under threat. Anything that generates an immune response in the host is known as an antigen.

The immune response is also determined by the nature of the tissue invaded. For instance, the eye (and the brain) respond when under foreign attack, but under certain circumstances the predicted response does not occur but is rather modulated by the tissue; this is called 'immune privilege' and is related to the special microenvironment and immunoregulatory mechanisms operating in these tissues.

Most of this chapter outlines the basic principles of immune responses, with reference to the eye as appropriate. At a very simple level, however, it must be remembered that the organism has very effective means to prevent host and pathogen meeting each other in the first place: namely, barriers such as the skin and mucous membranes with their surface coverings (including that most important layer of mucus present on mucosal surface tissues).

Overview of the immune system

Immunity is defined as the ability of the host to protect itself against a foreign organism or pathogen. To do this it requires an *immune system* comprising the cells and molecules used in the host's defence. For unicellular hosts this may simply mean certain molecules on the cell surface that enable it to recognize foreign organisms. However, for hosts in the animal kingdom the immune system is a highly organized network of tissues, cells and molecules.

Hosts, including vertebrates such as humans, defend themselves by mounting an *immune response* involving the activation and recruitment of the cells and molecules of the immune system. The immune system is considered in terms of its specificity: the *innate (natural or native) immune system* and the *acquired (adaptive) immune system*. The innate immune system provides the first line of defence against attack since this is rapidly mobilized and is not dependent on previous exposure to the foreign invader, i.e. does not involve 'memory' (but see later). This form of response is considered relatively *non-specific* in that the same sort of response occurs to most foreign organisms and even to injury itself. It involves activation of cells (such as macrophages, see p. 386) and molecules (such as complement, see p. 411) which are present in the tissues and fluids of all higher organisms having progressively evolved to their current level of efficiency from primitive ancient systems.

Innate immunity, however, has some degree of specificity in that innate immune cells such as macrophages 'sense' pathogens through receptors that recognize generic molecular 'patterns', which characterize different types of pathogens (pathogen-associated molecular patterns or PAMPs). Thus, there are receptors for broad groups of molecules present in different organisms such as viral DNA, fungal carbohydrates and bacterial endotoxins (see Chapter 8, pp. 480–481). However, as stated previously, immune responses can be initiated not only by pathogens but simply by injury, or by molecules such as plant pollens (causing allergies), and especially by alloantigens (those molecules that differentiate individuals within a species and cause rejection of donor-unrelated transplants). Accordingly, Matzinger and Kamala (2011) developed Janeway's PAMP hypothesis to suggest that innate immunity has evolved to allow the host to sense danger or damage to tissues (DAMPs) and includes anything that threatens host or tissue viability, namely pathogens of all varieties, sterile trauma such as chemical or thermal injuries, non-self proteins such as those present in allografts, and even 'self' proteins if these have become 'abnormal', as in the case of cancer cell proteins, prion proteins, and altered or unaltered self proteins seen by the immune system in the wrong context (auto-antigens).

Further information on innate immune receptors and pathogen recognition available online at https://expertconsult.inkling.com

Accordingly, the molecular patterns associated with danger signals and tissue damage are known as danger- or damage-associated molecular patterns (DAMPs) and include PAMPs. Thus, bacterial and viral products such single- or double-stranded RNA, endotoxin, complement fragments, reactive oxygen species, heat shock proteins, high mobility group box-1 (HMGB-1) proteins, and many other molecules (sometimes called 'alarmins') can act as DAMPs. Another group of molecules, 'alarmins', are released not only from damaged cells but also from cells under stress and include cytokines such as interleukin -IL)1, IL-16 and IL-33 (see p. 411)

The host-defence immune system thus includes:

- physicochemical barriers such as the skin, eyelids, tears (see Chapter 4, p. 206);
- molecules normally present in body fluids such as blood, tears and aqueous humour (e.g. complement, lysozyme, antiproteases). Antibacterial defensins are also on this list of ocular surface proteins;

INNATE IMMUNITY, PAMPS AND DAMPS

The field of innate immunity continues to expand. After the initial discovery of the phagocytic leucocyte (macrophage) by the Russian zoologist Metchnikoff at the beginning of the 20th century, much of immunology was devoted to immunochemistry of antigens and antibodies and their unique specificity. In the latter half of the 20th century, the emphasis shifted to developments in cellular immunology, involving cells of the adaptive immune system such as T cells and B cells. Specificity for pathogen recognition by innate immune cells was considered non-existent until Janeway and others identified receptors in the fruit fly *Drosophila* that were important not only for wing patterning in development but also for recognition of pathogens, such as fungi.
His earlier proposal that there were classes of receptors recognizing broad groups of pathogens such as viruses, fungi, bacteria, including mycobacteria, and parasites was shown to be true by the discovery of Toll-like receptors (TLRs). Other classes of pathogen-associated molecular patterns were later discovered and include the nucleotide-biding oligomerization domain-like (NOD-like) receptors (NLRs) C-type lectins (which recognize a wide array of carbohydrate ligands), the retinoic acid inducible-receptors (RLRs) and the AIM2-like receptor (ALR). Interestingly, pathogens may activate more than one receptor and induce cell signalling responses via supramolecular organizing centres (SMOC) (Kieser and Kagan, 2017) such as the myddosome and the inflammasome. However, as Matzinger pointed out, pathogens were not the only agents that activated the innate immune system. Allergens also initiate immune responses, as do allografts, where infectious pathogens were considered not to play a part. Her concept of molecules being either 'dangerous' or 'harmless' has changed the thinking of immunologists in many ways. In addition, she has developed her concepts to consider that different tissues respond to the same insult immunologically in different ways: for instance, the skin and the liver might produce a modulated immune response to the same antigen. In one sense this is a rewriting of the notion of immune privilege, which has been long accepted as a modified immune response taking place in the eye and the brain (Forrester et al., 2018) (see p. 454). Furthermore, the recognition of different pathogens by individual germ-line encoded receptors (PAMPs) implies a level of inflexibility which would not allow for pathogen strain evolution, particularly of viruses. Accordingly, it has recently been proposed that PRRs may not specifically bind ligands but may respond to perturbations in intracellular homeostasis (HAMPs) (Liston and Masters, 2017).

- phagocytic and cytotoxic cells such as polymorphonuclear leucocytes, macrophages, eosinophilic granulocytes, natural killer (NK) cells;
- molecules released by cells (alarmins) responding to attack and acting on other cells (cytokines), such as interleukins and tumour necrosis factor-α (TNF-α) (see later) and complement.

As indicated earlier, the innate immune system is classically considered not to have 'memory' and therefore if, after successfully repelling pathogen invasion and clearing the pathogen from the system, a second attack by the same pathogen occurred, the organism would have to go through the same process again. However, innate myeloid immune cells and specifically one subset of specialized myeloid cells, termed dendritic cells (DCs), after devouring and digesting a pathogen, have the potential to present titbits of the devoured material to specialized immune cells, termed T cells, which, once activated in the throes of battle with the pathogens (inflammation), become specialized not only to assist in clearing the pathogen (effector T cells) but also to become long-lived memory cells that can respond rapidly and with minimal fuss to a repeat challenge by that same pathogen. This is acquired immunity, which is exquisitely specific and not only involves T cells (cytotoxic T cells) that can kill pathogens and infected cells but also can activate other immune cells, namely B cells, to produce very high levels of pathogen-specific killer molecules (antibodies), which are the basis of vaccination protocols now used against many common pathogens.

The acquired immune system includes:

- specific immune systems associated with barrier surfaces (the skin immune system and the mucosa-associated immune system, MALT) where memory T cells reside and have 'first call' on invading pathogens;
- a constant source of naive T and B lymphocytes with potential to modulate their receptors to specifically recognize foreign organisms and molecules (antigens);
- molecules that specifically counteract foreign antigens (antibodies), i.e. immunoglobulins of which there are five types (see p. 406), which are based on ancient evolutionary building blocks;
- non-specific molecules (e.g. cytokines, proteases) released by antigen-specific cells (e.g. lymphocytes), which will directly or indirectly recruit further myeloid cells to sustain the response as needed.

Specific immunity is described as *humoral* when antibodies (derived from B lymphocytes; see later) and complement are involved in removing the antigen or as *cell-mediated* (cellular) when T lymphocytes and macrophages are involved.

Immunity after infection is normally termed *active immunity* in that the host has responded actively to the stimulus. However, immunity may be transferred *passively* by antibodies or cells. Vaccination procedures that involve the administration of antibodies are termed *passive immunization*, while those that involve inducing a response to the antigen either to pathogen-derived protein or even to the attenuated live organism are termed *active immunization*.

The development of acquired immunity involves a number of discrete phases, including:

- an afferent phase in which the foreign antigen is transported from the site of entry and presented by specialized antigen-presenting cells (APCs, usually dendritic cells) to the lymphocytes in the lymphoid tissue (see later)
- a phase of T-cell activation and clonal expansion in which T cells are transformed from a resting to an active state
- an effector stage in which T cells induce other cells, such as B cells and macrophages, to remove the antigen. If the antigen remains intracellular, as with virus-infected cells, the T cells themselves attack and kill the infected host cell (cytotoxic T cells).

This is known as the *primary immune response* and is accompanied by the appearance of antigen-specific T cells and antibody-secreting B cells. As indicated previously, the acquired immune system evolved to provide a memory-based rapid reaction force to heighten immune defence. Thus, on second exposure to the same pathogen, antigen-specific memory T and B cells are recruited much sooner and more efficiently, such that antibody levels are considerably higher than on the first exposure. This is known as the *secondary immune response*. In addition, the type (isotype; see later) of antibody in secondary immune responses is different.

Throughout the development of the acquired immune response there are several checkpoints that prevent a runaway overwhelming inflammatory reaction. Many of these regulatory mechanisms involve induction of specific cells such as regulatory T cells (Tregs) cells and B cells (Bregs), and they have a primary role in preventing adaptive immune responses to self-antigens. However, they also participate in downregulatory responses to foreign antigens and help to restore homeostasis.

Innate immunity with its early warning, rapid-response system provides a reliable means of protecting the host against most extracellular organisms (pathogens) and is a property of every living being including insects such as drosophila, which have contributed a great deal to our understanding of innate immunity. It might be asked, therefore, why the acquired immune system has evolved in vertebrates such as mammals, including humans. In part this is a result of the remarkable ability of pathogens to evade the immune system. Many pathogens reside and hide within the host cell, as in the case of protozoa such as *Toxoplasma*, *Mycobacterium tuberculosis*, obligate parasites such as *Chlamydia*

or, more frequently, viruses. Viruses use the host cell machinery for survival and replication by incorporating their genome, or at least part of it, into the DNA of the host cell. This may lead to viral latency or persistence, but may also allow the immune system to recognize the infected cells by the expression of the foreign antigen on the surface of the cell in addition to the self molecules (also termed antigens because they can induce an immune response). It is the recognition of the foreign antigen in the context of self-antigen, i.e. together with self-antigens, the peptide–major histocompatibility complex (MHC) (see p. 428), that led to the evolution of the acquired immune system (Flajnik, 2018). However, many pathogens continue to evade the immune system by subverting the function of the innate immune cells in which they reside or by becoming 'latent' as in the case of herpes simplex-infected neurones.

The innate immune system is the initial primary responder to pathogens or other challenges. The acquired immune system has developed a considerable degree of sophistication, based on memory of previous pathogen encounters, with an increasingly recognized range of different cell types: T and B lymphocytes; T and B cell subsets (e.g. T helper (Th), T cytotoxic (Tc), T and B regulatory (Treg and Breg) cells; and even subsets of these (e.g. Th0, Th1, Th2, Th17) (see later), each designed for specific cellular functions. In addition, T and B cells have evolved in fundamentally different directions: T cells to deal with surface-bound antigen (usually cell-associated) and B cells to deal with soluble (extracellular) antigen. The sophistication and specificity of the acquired immune system involving T cells has thus been harnessed to assist the innate immune system in dealing more efficiently with extracellular organisms, for instance via B cells.

Certain basic concepts about immune mechanisms can therefore be derived from these considerations:

- extracellular foreign antigen is normally cleared by the innate immune system, with some assistance from B-cell activity
- intracellular foreign antigen may evade the immune system and remain latent during the lifetime of the host unless (1) it kills the cell, is released and generates a second innate immune response to remove the pathogen; (2) it is expressed on the surface of the cell in the context of self-antigen and is recognized by the acquired immune system, which kills the infected cells and clears the pathogen (applies to many viruses, e.g. influenza); (3) the host immune system is compromised, in which case the intracellular pathogen proliferates freely, killing many cells and ultimately the host
- all cellular defence mechanisms involve interactions of cell surface molecules (receptors) with complementary molecules (ligands).

So far we have focused on *immunity*, defined as the response of the host to foreign organism. This suggests that all foreign organisms are pathogens but this is not so. Indeed, the great majority of microbes and other living organisms are not pathogenic, as evidenced by the very large numbers of *commensal* organisms that colonize several regions of the body, particularly the GI tract (the microbiome) (Box 7.1), and indeed are required for normal health (homeostasis).

This lack of immune responsiveness to foreign antigens can be described as 'tolerance' of foreign organisms and indeed the immunological obverse of 'immunity' is 'tolerance'. Tolerance as a defining characteristic can be applied to any antigen but is especially relevant to self-antigens, since clearly we do not normally react to our own tissues. It has been realized that tolerance to self-antigens, and indeed to any antigen, is an active process in which Tregs are critical. Tregs to self-antigens are generated in the thymus (thymic (t) or natural (n) Tregs). Many effector immune responses to infectious foreign antigens are also controlled by the induction of specific Tregs to that organism (induced (i) Tregs). Tregs are part of our normal circulating T-cell population and when Treg function goes awry, autoimmune diseases can ensue; diseases such as various forms of dry eye disease (Sjögren's syndrome, see Chapter 9, p. 506), rheumatoid arthritis and some forms of uveitis are considered autoimmune diseases. Indeed, there is a growing recognition that both the innate and the acquired immune response co-evolved with microbes, each exerting counter-regulation on the other with the suggestion that the hygiene hypothesis might underpin autoimmune disease as well as allergy .

What is the hygiene hypothesis: a brief summary is available online at https://expertconsult.inkling.com

To summarize, acquired immune responses have certain features that are inherent to them:

- specificity: based on certain features/determinants (known as epitopes) of the antigenic structure
- non-responsiveness to many antigens: (prototypically, self-antigens), also known as tolerance, the loss of which underpins autoimmune disease
- diversity: around 10^9 individual epitopes can be distinguished: this is known as the lymphocyte 'repertoire' and represents the range of potential foreign antigens to which immune responses can be generated
- memory: the secondary immune response, specific to that antigen and present in both B and T cells
- specialization: the immune response is customized to different microbes
- ability to downregulate: the immune response is strictly regulated/limited in magnitude and time through specific and non-specific mechanisms.

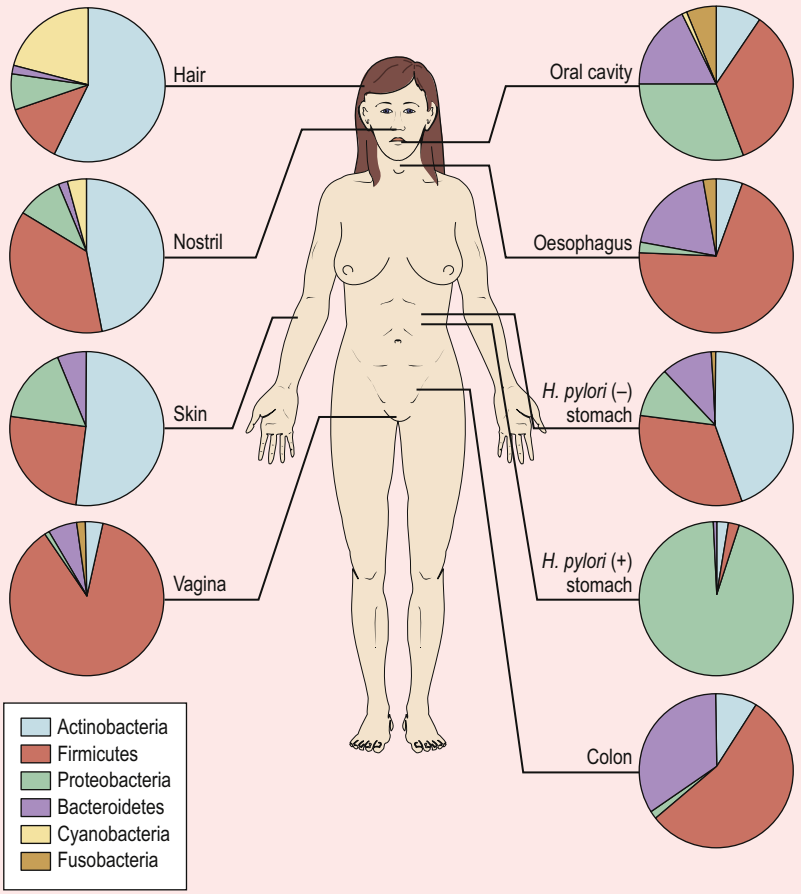
THE HYGIENE HYPOTHESIS

The epidemiological evidence showing a marked increase in allergic diseases such as asthma, eczema and allergic/atopic conjunctivitis has been attributed since around the turn of the 21st century to a reduction in the environmental burden of infectious disease, particularly in the West. Exposure to non-infectious agents such as environmental pollutants has replaced infection as the main suspect. More recently Bach (2018) has suggested that the parallel rise in autoimmune disease may be associated with a loss of a protective role for pathogens in preventing abnormal T-cell responses of autoantigens. For instance, he has shown that a specific strain of mice is more likely to develop spontaneous diabetes if they are reared in a pathogen-free environment. He has also provided impressive epidemiological evidence of the incidence of autoimmune disease versus infectious disease in developed countries compared with developing ones. The proposal is that the induction of regulatory cells (Tregs) in response to infection generates cells that are not only specific for the infection but have broad regulatory function against other T cells, responses not only to other infections but also to autoantigens. Included in his concept is the role of commensal bacteria especially in the gut. The situation cannot be so simple and is likely to work in both directions with host-pathogen interplay modifying disease for better and for worse.

Box 7.1 **The microbiome**

The total number of separate genomes from microorganisms colonizing a specific site (known as an environmental niche) constitutes a microbiome for that site. There are 10 times more microbial cells than human cells in the body and although the microbiome does not weigh much more that 500 grams, its importance to normal physiological processes is such that some regard it as a separate and essential organ. Most of the organisms that constitute the microbiome cannot be cultured and have been identified by genetic mapping techniques. The most important role for the microbiome may be in immunoregulation and prevention of disease including autoimmune disease, diabetes, atherosclerosis, cancer and other diseases associated with chronic inflammation and obesity. It may also have a role in psychiatric disease through its effects on neurotransmitter production.

High-throughput sequencing has revealed substantial intra-individual microbiome variation at different anatomical sites and inter-individual variation at the same anatomical sites (see figure). However, the microbiome at specific anatomical sites can remain remarkably stable over time in each individual. The figure indicates the relative proportion of sequences determined at the taxonomic phylum level at eight anatomical sites. Certain features, such as the presence (+) or absence (–) of *Helicobacter pylori*, can lead to permanent and marked perturbations in community composition. (Adapted from Costello et al., 2012).



Cells and molecules of the immune system

The cells of the immune system are mostly specialized to participate in processes belonging either to the innate or to the acquired immune systems (Box 7.2). Some cells are central to both and the distinction between innate and adaptive immune cells is becoming blurred with the discovery of cells such as innate lymphoid cells (see p. 391 and Video 7.1).

The myeloid system and innate immunity

The cells directly involved in bacterial killing and removal of damaged host tissue at the site of entry of the foreign organism are the cells of the myeloid system. Some of these cells are recruited to the site of injury during acute inflammation. Their defining feature is the presence of cytoplasmic organelles (lysosomes) containing a battery of hydrolytic enzymes for both intracellular and extracellular killing.

Neutrophilic granulocytes

Neutrophilic granulocytes (polymorphonuclear leucocytes) are the most common white cells in the circulation and are attracted to sites of inflammation by chemotaxis (see p. 398). These are fully differentiated cells with no capacity for proliferation. They are primary scavengers, causing much of their effect via release of free radicals and proteases from their numerous cytoplasmic granules and lysosomes. These include defensins, lysozyme, lactoferrin and oxidative enzymes (e.g. NADPH-dependent oxidases, myeloperoxidases and catalase), which are also present in ocular fluids such as tears. The half-life of neutrophils is 1–2 days.

This limited 'clearing-the-bugs' role has been revised recently with many new functions being ascribed to neutrophils. PMNs have been shown to use a remarkable mechanism for 'trapping' bacteria and other 'dangerous' material by extruding NETS (neutrophil extracellular traps), composed predominantly of nuclear DNA material, during the process of cell death. NETS are very 'sticky' and powerfully prevent bacterial dissemination. However, they can also have deleterious effects by trapping metastatic cells and protecting them from immune surveillance.

Further information on nets and netosis is available online at https://expertconsult.inkling.com

Neutrophils may also participate in the adaptive immune response through release of cytokines and extracellular vesicles (evs), which mediate intercellular communication. They may also have the ability to present antigen to T cells, probably through evs.

Myeloid mononuclear cells

Monocytes, macrophages and dendritic cells are cells of the mononuclear phagocyte system. Like neutrophils, these cells are derived from bone marrow haematopoietic stem cells and differentiate into a variety of tissue macrophages (histiocytes) with specific functions (Box 7.3).

Macrophages have many functions as part of the innate immune system: they phagocytose dead and damaged cells and organisms in inflammatory exudates; they release cytokines of various sorts, which may activate other cells such as lymphocytes and eosinophils; and they are involved in the acquired immune system as APCs, as effector cells involved in the process of cell lysis, and by removing antibody-coated (opsonized) cells and particulate material.

Macrophages are, however, not simple cells. There are several varieties of macrophages, some with 'housekeeping' roles as resident cells whose job is to remove dead and dying cells during normal tissue cell turnover. Other macrophages are recruited to sites of inflammation, differentiating from circulating blood monocytes, and are highly active in tissue destruction. These cells are sometimes referred to as classically activated M1 macrophages with a strongly pro-inflammatory function, while resident and other macrophages are considered to be alternatively activated M2 macrophages and function more in healing processes including fibrosis and new vessel formation.

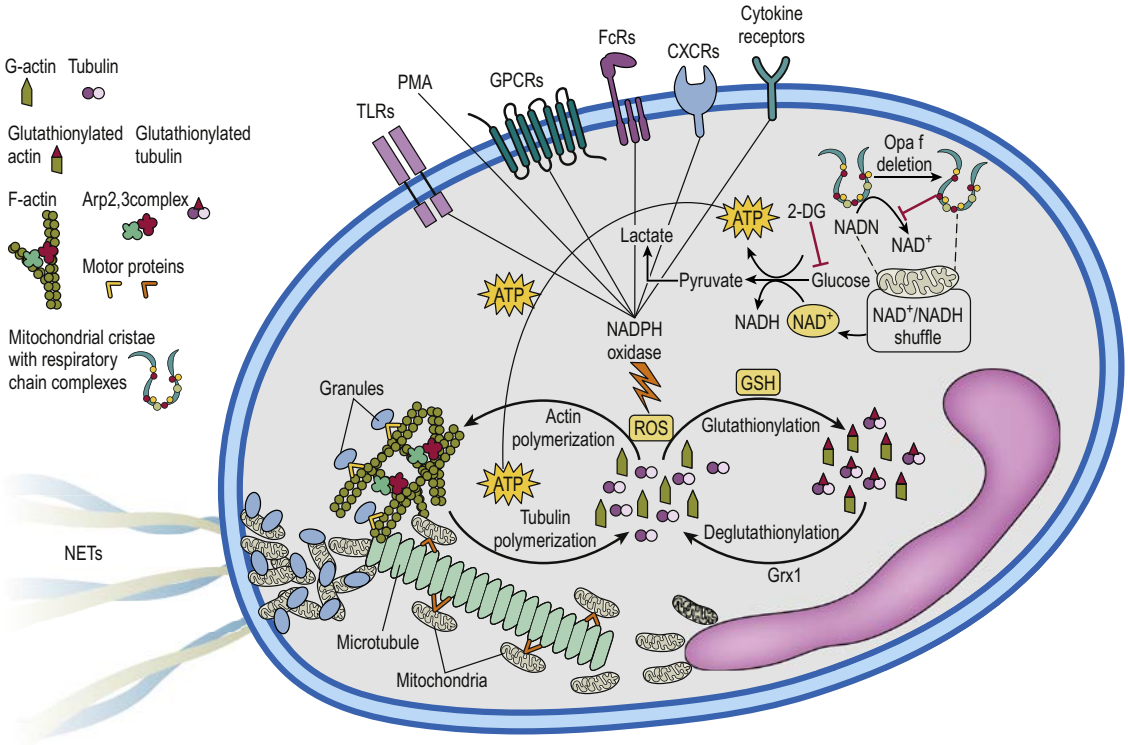
Macrophages are not all the same: more details are available online at https://expertconsult.inkling.com

Macrophages, like B cells, usually capture antigen (e.g. opsonized bacteria) for processing and presentation to T cells via surface receptors such as the complement receptor and the Fc receptor. The Fc receptor on macrophages binds the Fc portion of antibody–antigen complexes while B cells express antibody (surface immunoglobulin, sIg) on their surface and use the Fab portion (see later) of the antibody to capture antigen. In both cases, preformed antibody to that antigen is needed and so the immune system must have already been exposed to antigen. It is clear therefore that macrophages and B cells are only effective as APCs in the later stages of an ongoing immune response.

In contrast, the most potent APC in the immune system is the DC. The DC is the key cell in linking the adaptive and innate immune systems. While the function of most myeloid cells is to engage in innate immune defence by marshalling forces against infectious and non-infectious danger signals, the DC takes this one step further: after phagocytosing, killing and digesting an invasive microorganism the DC, unlike the macrophage or the B cell, or even 'non-professional' APCs, can present the processed

NETS AND NETOSIS

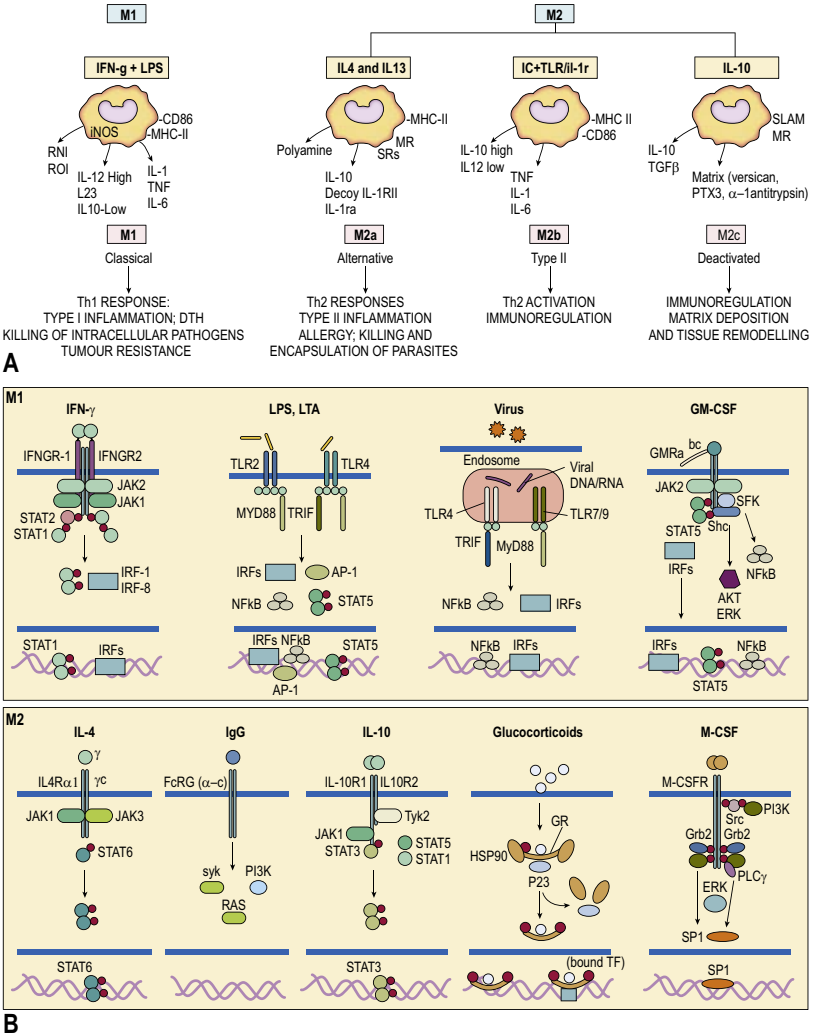
PMNs release mitochondrial DNA binding granule proteins in the form of neutrophil extracellular traps (NETS), which increases their ability to trap extracellular microorganisms and facilitate their killing. Release of NETS requires production of ATP by glycolysis, which drives the changes in cytoskeletal microtubule and F-actin necessary for the movement of the mitrochondria to the plasma membrane where they release the NETS (eFig. 7.1). Release of NETS is not to be confused with NETOSIS, a form of PMN necrotic programmed cell death in which nuclear DNA is released in the form of an extracellular 'cloud' and presages the death of the cell (Yousefi et al., 2019). NETS can also serve as substrates for complement activation and coagulation, thus promoting the inflammatory response. However, PMN NETS not only trap bacteria and fungi but other potentially larger organisms including parasites and helminths and it has been suggested that trapping of some pathogens may actually sequester them from the immune system.



eFig. 7.1 Glycolytic ATP production and active cytoskeletal rearrangements are required for neutrophil extracellular traps *(NETs)* formation. Physiological activation of neutrophils through a range of receptors results in reactive oxygen species *(ROS)*: induced actin and tubulin glutathionylation thereby control cell cytoskeletal dynamics and NET formation. This process is tightly regulated by glutaredoxin 1. Disruption of the Grx1 gene or deficient ROS production (e.g. in chronic granulomatous disease patients) prevents the cytoskeletal rearrangements required for NET formation. Successful NET formation requires ATP produced via glycolysis, which is indirectly controlled by OPA1 (an inner mitochondrial membrane protein) through supply of NAD^+. *GPCR*, G-protein coupled receptors; *GSH* glutathione *NADH*, nicotinamide adenine dinucleotide *NADH*, Reduced nicotinamide adenine dinucleotide phosphate. (From Yousefi et al., 2019)

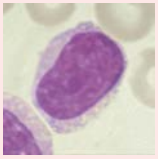
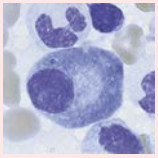
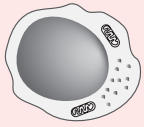
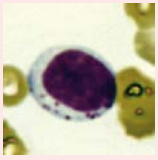
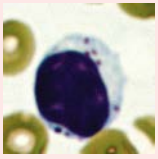
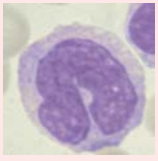
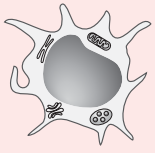
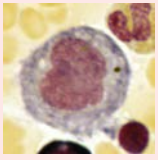
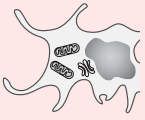
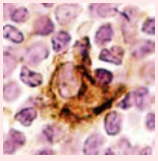
MACROPHAGE HETEROGENEITY

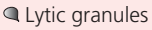
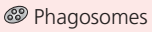
The functional subtyping of macrophages is of considerable importance to the understanding of disease processes including many ocular diseases such as uveitis and macular degeneration. The initial classification into M1 classically activated macrophages and M2 alternatively activated macrophages has developed further into a realization that there is greater plasticity in macrophage function than was once thought (eFig. 7.2). Classically activated macrophages remain major players in innate immune defence with the job of removing dangerous foreign and endogenous antigens, clearing necrotic and damaged tissue and setting the scene for repair of tissue. Repair and restoration of tissue function is conducted by alternatively activated macrophages involved in such processes as generating new blood vessels to restore blood supply to the tissues and produce factors that rein in the excesses of the inflammatory response. However, the dichotomy in macrophage subtypes is not as clear as with T cells (which are also of diverse types beyond Th1/Th2 cells) and there is increasing research in this field. Most recently, single cell RNA sequencing studies in a model of myocardial infarction have shown that there are four subsets of resident macrophages in healthy tissue (at least one of which is derived from yolk sac precursors rather than haematopoeitic stem cells). During inflammation, there are seven additional subtypes that have different roles in clearing inflammatory debris, restoring function and generating new blood vessels (Koelwyn and Moore, 2019). Similar heterogeneity in macrophage subtypes has been shown for instance in response *M. tuberculosis*. (From Marakala et al., 2018)



eFig. 7.2 The M1/M2 paradigm, origin and molecular basis. (**A**) Mantovani et al. proposed an M1–M2 macrophage model, in which M1 included interferon-gamma (IFN-γ) + lipopolysaccharide *(LPS)* or tumour necrosis factor *(TNF)* and M2 included three subtype particular signatures: *(IL-4)* (M2a), immune complex + toll-like receptor *(TLR)* ligands (M2b) and IL-10 and glucocorticoids (M2c). (**B**) Similarities and differences between M1 and M2 receptors and key signalling mediators are shown. (From Marakala et al., 2018.)

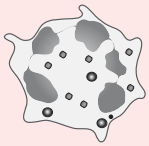
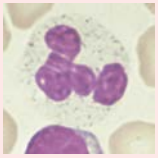
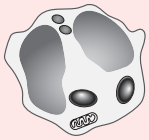
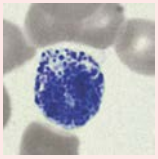
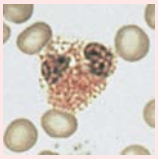
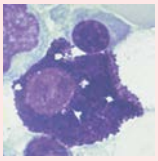
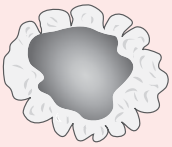
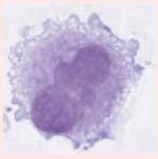
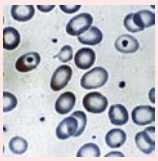
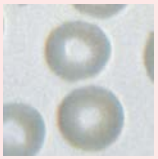
Box 7.2 **Cells and tissues of the innate and acquired immune systems**

	Mature T or B lymphocyte		6–9 μm; round or slightly indented nucleus; sparse cytoplasm; few granules; few mitochondria
	Plasma cell		5–30 μm; round or oval nucleus; abundant cytoplasm; no granules; abundant endoplasmic reticulum
	NK cell		10–12 μm; round nucleus; abundant cytoplasm; many granules; scattered mitochondria
	NKT cell		Phenotypically similar to NK cell
	Monocyte		12–20 μm; round, oval, notched, or horseshoe-shaped nucleus; abundant cytoplasm; abundant granules; well-developed Golgi apparatus; abundant mitochondria
	Macrophage		15–80 μm; elongated, indented, or oval nucleus; abundant cytoplasm; many granules and vacuoles; few mitochondria; abundant lysosomes
	Dendritic cell		Irregularly shaped cell and nucleus; many cellular processes; few intracellular organelles; prominent mitochondria

 Lytic granules
 Mitochondria
 Smooth endoplasmic reticulum
 Granules
 Phagosomes
 Golgi apparatus
 Rough endoplasmic reticulum

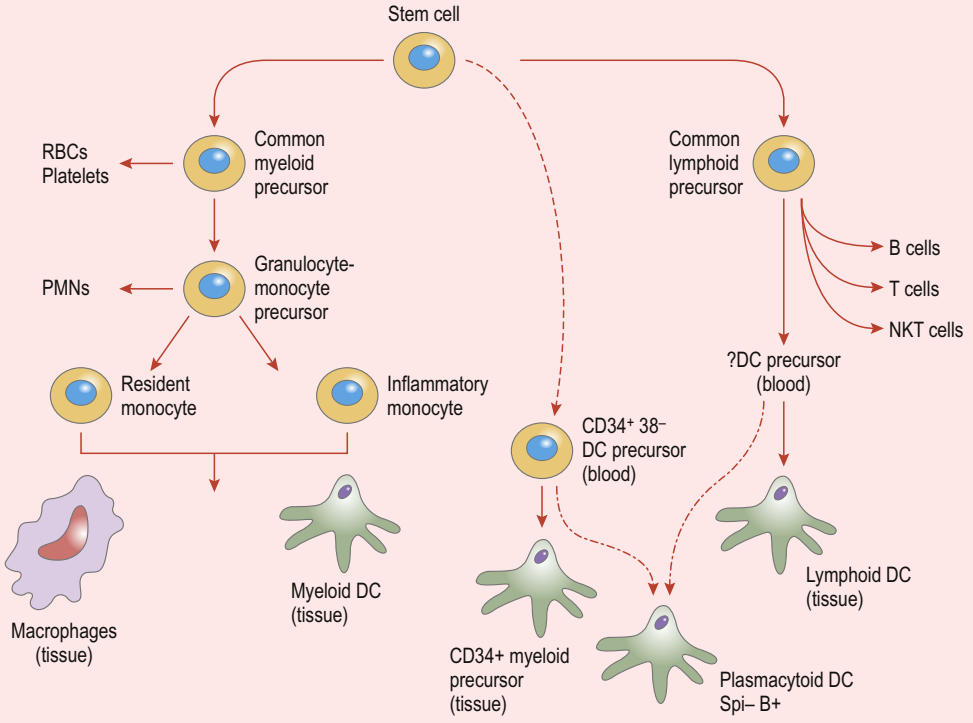
Continued

Box 7.2 **Cells and tissues of the innate and acquired immune systems—Cont'd**

	Neutrophil		10–15 μm; 2–5 distinct nuclear lobes; abundant cytoplasm; numerous granules; few mitochondria; abundance of glycogen
	Basophil		14–16 μm; 2–3 nuclear lobes; abundant cytoplasm; large, coarse granules; many ribosomes and mitochondria; abundance of glycogen
	Eosinophil		12–17 μm; 2–3 nuclear lobes; abundant cytoplasm; large oval granules containing elongated crystalloid and smaller granules; extensive smooth endoplasmic reticulum; few mitochondria
	Mast cell		14–16 μm; non-segmented nucleus; abundant cytoplasm; many large granules; few mitochondria
	Megakaryocyte		35–160 μm; irregularly shaped nucleus; abundant cytoplasm; fine granules
	Platelet		1.5–3.5 μm; non-nucleated; granular cytoplasm
	Erythrocyte		7.2 μm; non-nucleated; no organelles

NK, Natural killer cell.
(From Mak and Saunders, 2006, with permission from Elsevier.)

Box 7.3 Generation of macrophages and dendritic cells from myeloid and lymphoid precursors



Note that dendritic cells *(DCs)* can be derived from both sources.
NKT, Natural killer T cells; *PMNs*, polymorphonuclear cells; *RBCs*, red blood cells.

antigen to naive T cells and induce an entirely new adaptive immune response with all its features, including memory (see previous section). There are several subsets of DCs but, in essence, there are two major groupings: resident, non-migratory DCs which are located in the secondary lymphoid tissues such as the spleen and the lymph node and migratory DCs, which arise in the bone marrow and migrate to the tissues where they are available as sentinels to register pathogens and other danger signals and transport antigen to the secondary lymphoid tissues where they activate the adaptive immune system (T and B cells). Tissues at the front line of attack, e.g. lungs, skin, gut mucosa and conjunctiva, are rich in DCs.

Further information on naming the cells of the mononuclear phagocyte system are available online at https://expertconsult.inkling.com

Tissue-resident DCs (cDC1) constantly traffic as precursors from the bone marrow to the tissues and then migrate to the secondary lymphoid organs. If they do not detect 'dangerous' antigens, but only self-antigens, they do not activate the adaptive immune system. They therefore in fact maintain immune tolerance by sustaining the population of endogenous Tregs, i.e. it seems that, although DCs have the powerful potential to rapidly activate adaptive immunity, as in the case of a viral infection, their homeostatic (and probably main function) is to maintain tolerance.

Other granulocytes

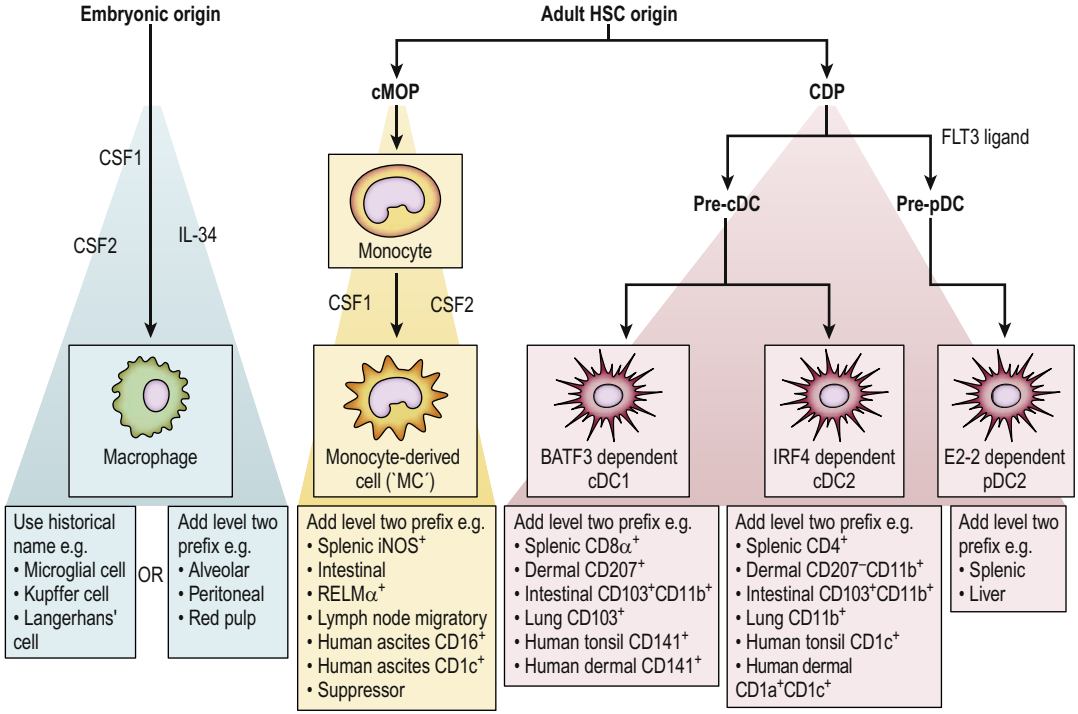
Mast cells/basophils/eosinophils (see Box 7.2) are all part of the granulocyte series of cells. Basophils (life span: 2–3 days) are the circulating equivalent of mast cells (life span: long-lived and can proliferate in the tissues), which occur only in the tissues but have separate lineages. Mast cells and eosinophils also have separate lineages and basophiles and eosinophils may derive from a common granulocyte–monocyte precursor.

Mast cells derive from the yolk sac during embryogenesis but adult mast cells come from late fetal erythro-myeloid progenitors, which populate the fetal liver via an integrinβ7+ progenitor cell (Li et al., 2018). Two types of mast cell are described: the connective tissue mast cells and the mucosal mast cells on the basis of their granule

NAMING THE CELLS OF THE MONONUCLEAR PHAGOCYTE SYSTEM

Several types of DC have been identified on the basis of cell surface markers (CD markers). Two main subsets are recognized: plasmacytoid DC (pDC) and conventional DC (cDC) (eFig. 7.3). Both are derived from precursors in the bone marrow, the myeloid-derived precursor cells (adult haematopoietic stem cells (HSCs)), which also generate precursor cells for the macrophage lineage. In addition, there is an embryonic yolk-sac derived precursor for resident tissue macrophages. HSC-derived DC precursors (pre-cDC) circulate in the blood and also seed the tissues under homeostatic conditions where they provide a surveillance function for pathogens and danger signals. They also proliferate and respond rapidly to danger and can become inflammatory DCs. Some of these are derived from monocytes, which can become either tissue inflammatory macrophages (see earlier) or become Tip DCs (TNF-α, NO-producing DCs) which are active in antigen presentation and generation of the immune adaptive immune response.

Studies on DC lineage and phenotype have mostly been performed in mice and humans. DC subsets and markers are defined by several different cell surface markers, which are variably expressed on different DC subtypes. In this classification, two discrete subsets of conventional DC are recognized (cDC1 and cDC2) (eFig. 7.3) (from Guilliams, et al., 2014).



eFig. 7.3 Two levels of nomenclature may be used for classifying mononuclear phagocytes: (a) ontogeny-based (level one nomenclature; *yellow boxes*), (b) function, location and/or morphology-based (level two nomenclature; *blue boxes*). This yields three main groups of cells: namely, common dendritic cell (*DC*) precursor (*CDP*)-derived DCs, embryonic-derived macrophages and monocyte-derived cells. DCs are further subdivided into classical type 1 DCs (cDC1s), cDC2s and plasmacytoid DCs (pDCs) because their development depends on distinct sets of transcription factors and because they arise from discrete committed precursors. In the lower part of the figure, we have added some examples to illustrate how our approach can yield a unifying nomenclature without losing flexibility. Level one nomenclature also includes unambiguous and widely accepted historical names. Level two nomenclature can include surface markers that are used to identify the cells, the functional specialization studied or information on cell localization. Examples of level two nomenclature are provided; however, in many cases, level one should be sufficient to adequately define a population, except when a novel function and/or relevant marker is required to discern a particular cell subset. We suggest the use of 'MC' as an abbreviation for monocyte-derived cells. *BATF3*, Basic leucine zipper transcriptional factor ATF-like 3; *cMoP*, common monocyte progenitor; *CSF1*, colony-stimulating factor 1 (also known as M-CSF); *CSF2*, colony-stimulating factor 2 (also known as GM-CSF); *FLT3*, FMS-like tyrosine kinase 3; *HSC*, haematopoietic stem cell; *IL-34*, interleukin-34; *iNOS*, inducible nitric oxide synthase; *IRF4*, interferon-regulatory factor 4; *RELMα*, resistin-like molecule-α. (From Guilliams, et al., 2014.)

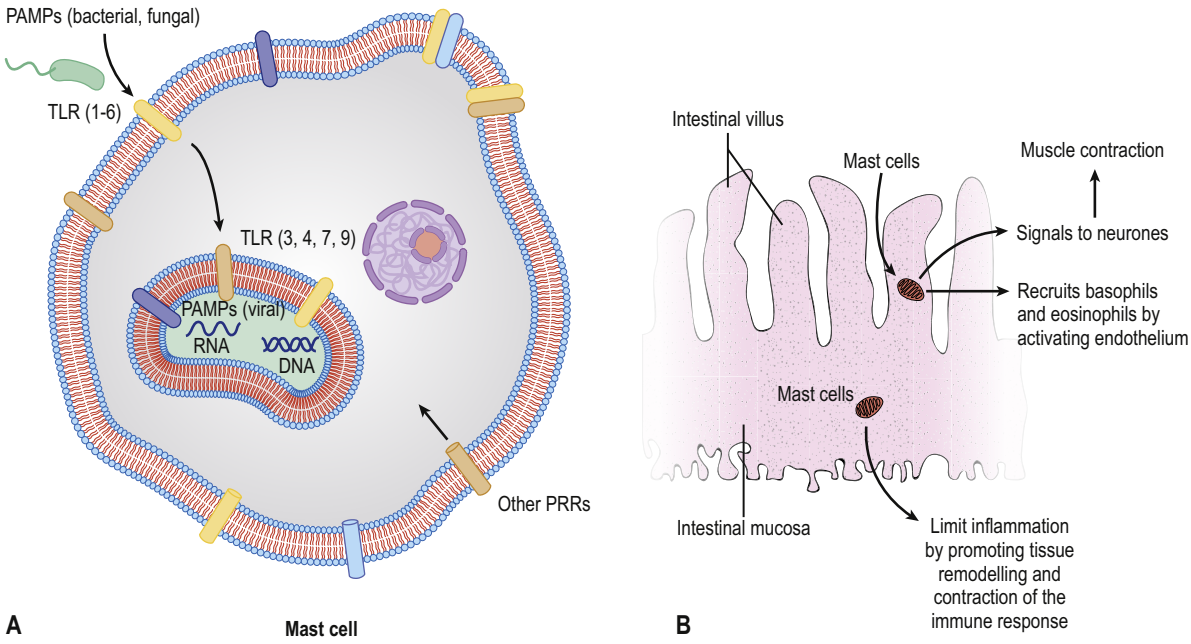


Fig. 7.1 Pathogen-recognition receptors *(PRRs)* occur on many cell types including mast cells, which use them to regulate innate immune responses. In (**A**) the different types of PRR on mast cells are shown, as well as other unique receptors that drive immune responses by the mast cells determined by a level of tissue specificity, which act in several ways including cell chemotaxis, vascular permeability increase and mucin production. In (**B**) location of mast cells in the lining of the gut is demonstrated. *PAMPs*, Pathogen-associated molecular patterns. (From St John and Abraham, 2013.)

proteases, their susceptibility to degranulation by immunoglobulin E (IgE) and their requirement for different maturation signals (see later). Mucosal mast cells lie in connective tissue adjacent to epithelial surfaces and connective tissue mast cells lie along the walls of endothelial (blood vessel) barriers and both are thus important innate immune cells, contributing to the initial response to external pathogens or endogenous danger signals by releasing mediators and increasing vascular permeability (Fig. 7.1). Mast cells thus have further important roles in many aspects of immunobiology including an important role in engineering the contraction of the innate immune response.

Eosinophils account for about 0.1% of circulating leucocytes. Their numbers are elevated in chronic allergic disorders, in both the circulation and the tissues. They are also produced in large numbers in response to parasite and helminth infections and contain a panel of particularly potent parasite-specific proteases. Like mast cells, they have high-affinity immunoglobulin E (IgE) receptors and are probably effectors of tissue damage in allergic disease in human disease, including asthma and chronic allergic conjunctival disorders such as atopic keratoconjunctivitis and vernal keratoconjunctivitis (see p. 456). Interestingly, eosinophils are major components of allergic disease and have been reported to be prominent in fungal antigen-driven allergy. In this respect, eosinophils, like mast cells and basophils, are classic biomarkers of type 2 immune responses (see p. 425) and are associated with alternatively activated (M2) macrophages. As such they promote not only immune homeostasis but also have a more general role in homeostasis including metabolic homeostasis (by promoting glucose tolerance and preventing fat deposition) and regulation of autophagy/ageing.

NK cells are an infrequent but important constituent of innate immune cells, although they are technically lymphoid cells. Their main function is to provide defence against virally infected cells and tumour cells, which they perform in an antigen-independent manner. Classically, they are considered not to generate immunological memory, which is the domain of the adaptive immune system, and therefore they are not contributors to long-term protective immunity. They express the surface antigen CD56 at both high and low levels. In humans they express the KIR (killer inhibitory receptor) family of receptors and in mice a range of Ly-49 antigens, which allow subsets of NK cells to be identified and which vary between strains of mice. KIR and Ly49 are one of the numerous receptors on immune cells expressing an immunotyrosine inhibitory motif (ITIM). NK cells use inhibitory receptors to prevent them responding to healthy cells that express MHC class

I (self) antigens (see p. 427) but respond to stressed cells that may have low levels of MHC class I (absent self-antigen) and kill such cells using cytotoxic mechanisms.

NKT cells are a further set of innate immune cells of lymphoid origin that also express a single invariant T-cell receptor that binds CD1 antigens (combined with glycolipid antigens, see p. 427), and thus like NK cells they are essentially involved in innate immune responses to microbial lipid, particularly from Gram-negative microorganisms expressing the glycolipid, glucuronyl ceramides.

The lymphoid system and acquired immunity

The two most important features of the acquired immune system are exquisite antigen specificity and immunological memory. The specificity of T cells (through the T-cell receptor (TCR) for peptide MHC complexes) and B cells through immunoglobulin antibodies for antigens is generated during T- and B-cell maturation by two recombination-activating genes (RAGs) that encode enzymes involved in the shuffling of the VDJ segments of immunoglobulins and TCRs. These define lymphocytes. In contrast to cells of the myeloid system, which remove debris and organisms by mechanisms that have limited specificity (macrophages phagocytose broad ranges of organisms using 'pattern recognition receptors' (PRRs) to recognize PAMPs; see p.381), each clone of lymphoid cell responds to a single antigen (in theory). T cells respond to antigen by proliferating and releasing cytokines, while B cells respond by maturing to plasma cells and producing antibodies. RAG deficient mice do not display adaptive immunity.

T cells

T cells (for thymus-derived) are lymphoid mononuclear cells that recognize antigen in conjunction with self-antigen (self-MHC, see page 427). They reside in the lymphoid tissues and circulate in the blood stream and lymphatics. T helper (Th) cells respond to antigen in association with MHC class II self-antigen, while T cytotoxic (Tc) cells respond to antigen combined with MHC class I antigen (see p. 427). T cells release cytokines, which are required for T-cell and B-cell proliferation and differentiation and for innate immune cell activation. T lymphocytes express surface markers (molecules detectable by specific monoclonal antibodies) characteristic of their phenotype. Thus, Th cells are described as $CD4^+$ cells and Tc cells are known as $CD8^+$ cells.

What are CD numbers? Further information is available online at https://expertconsult.inkling.com

This general categorization of T cells into two subtypes (CD4 and CD8 T cells) was found to be insufficient to explain the numerous different functions of T cells. After many years of controversy over whether a T-cell subset existed that suppressed (regulated) the immune response, the discovery of a T cell that was required to prevent spontaneous autoimmune disease indicated that regulatory T cells (Tregs) are present in normal healthy individuals.

T cells, particularly $CD4^+$ T cells, are defined by the cytokines they produce and the transcription factors they use to signal in response to stimulation. Cytokines are multifunctional short-acting, short-range mediators of cellular activities, released by T cells and other immune and non-immune cells (see p. 408). Cytokines are distinguished from other mediators as the molecules of 'intercellular communication'. In addition to $CD4^+$ Tregs, $CD4^+$ T cells now include several subsets such as Th1, Th2, Th17, IL22-secreting $CD4^+$ T cells and it is likely that this T cell specialization will be increasingly recognized (Fig. 7.2). Subsets of cytotoxic T (Tc) cells and B cells (including B regulatory cells) with defined roles in pathogenesis of disease have been described.

Further subsets of T cells occur, such as the γδT cells (which possess a T-cell receptor (TCR) with a γ–δ dimer, rather than the αβ TCR present on conventional T cells) and NKT cells (which combine properties of innate immune NK cells but possess an invariant αβ T-cell receptor).

B cells

B cells are mononuclear lymphoid cells that are specialized for the production of antibody and they reside mostly in B cell follicles in the lymph nodes, spleen and mucosa-associated lymphoid tissues (MALT, see page 422). There are five types (isotypes) of antibody (IgG, IgA, IgM, IgD and IgE) and B cells require 'help' from T cells to function (T follicular helper (T_{FH}) cells) (see Fig. 7.2).

During a primary immune response, IgM antibody is initially produced by activated B cells and expressed on the B cell surface as a surface receptor (sIgM). In secondary immune responses, the B cells switch to producing sIgG (isotype switching), often with higher binding capacity (affinity) for the antigen, a process termed affinity maturation. During allergic immune responses, a further isotype switch occurs from sIgG to sIgE. Secretory IgA forms part of the mucosal immune system, being present in large amounts in surface secretory fluids including tears. IgA is the most abundant immunoglobulin in the immune system. sIgD is co-expressed with sIgM on B cells as they leave the bone morrow. IgD may be a maturation marker of B cells indicating an activation state. IgD is present in very low amounts in the normal circulation.

B cells recognize and bind antigen via surface immunoglobulin (sIg), which is in effect the B-cell receptor. Antigen binds to sIg in the afferent phase of the secondary immune response; in contrast, antigen binds to the secreted form of immunoglobulin (antibody) in the effector phase of the response. Secondary afferent (cognitive) interactions are antigen-specific, but effector functions are not (see later).

WHAT ARE CD NUMBERS?

CD means cluster of differentiation: this term refers to a molecule that has a defined structure, that can be recognized by a group or cluster of monoclonal antibodies, and that identifies a specific lineage or differentiation stage in the cell. CD numbers are continually being defined and the 10th Human Leukocyte Differentiation Antigen (HLDA) workshop took place in conjunction with the Australasian Society of Immunology in December 2014. Over 400 CD numbers were designated at this meeting. There is general consensus that the designation of CD numbers is driving innovation and discovery in many fields, but especially immunology, which, when combined with developments in flow cytometry, is revolutionizing the field. More importantly, its greatest value is in validating monoclonal antibodies, which has been a major problem in the field to date and led to contamination of the literature with false results.

The 11th HLDA meets in 2019. Further details are available at: www.hcdm.org. A sample of CD numbers is presented in eTable 8.1.

eTable 1 Some of the better-known CD numbers

CD No	Cell/molecule	Function/role
1	Thymocytes, Langerhans' cells	
3	T cells	
4	T helper cells	Adhesion
8	T cytotoxic cells	
11a	Leucocytes (αL chain of LFA-1)	Adhesion
11b	Leucocytes (αM chain of LFA-1)	Adhesion
11c	Leucocytes (αX chain of LFA-1)	Adhesion
19	B cells	
25	Activated lymphocytes, IL-2 receptor and macrophages	
28	T cells	Co-stimulation
40	Many cells	Co-stimulation
44	T and B cells	Activated memory cells
45	Common leucocyte antigen	
45RA	Naive T cells	
45RO	Memory T cells	
56	NK cells	
62E,L,P	E, L and P selectins	Adhesion
68	Macrophages	
69	Many cells	Short-term activation
95	Many cells	Fas
120	TNF-α receptor	
142	Tissue factor	Coagulation
152	Cytotoxic T lymphocyte antigen-4	Immunoregulation
169	Sialoadhesin	Macrophages
183	CXCR3	Chemokine
284	Toll-like receptor 4	Myeloid cells
332	FGF receptor	Growth factor
363	SIP receptor	Lymphocyte homing

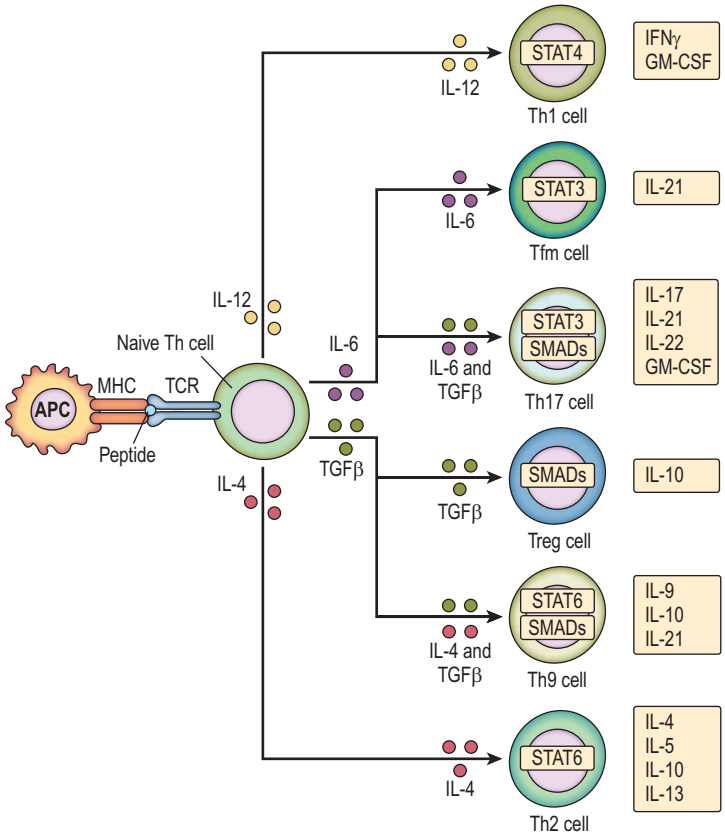


Fig. 7.2 The influence of cytokine environment on Th-cell differentiation. While the strength of T-cell receptor *(TCR)* signalling and the quality of co-stimulation have been demonstrated to influence naive CD4+ T-cell polarization, the cytokine environment largely dictates Th differentiation. The prototypical cytokines and their corresponding signalling pathways *(STATs, SMADs)* regulating each Th fate are depicted. Additional cytokine and signalling pathways influencing T-helper cell *(Th)* differentiation exist but have effects on the development of multiple Th subsets. For example, IL-2, through the STAT5 signalling pathway, is important during Th1, Th2, Th9, and T-regulatory cells *(Treg)* development but inhibits T-follicular helper cells and Th17 differentiation. *APC*, Antigen-presenting cells; *GM-CSF*, granulocyte–macrophage colony-stimulating factor; *MHC*, major histocompatibility complex; *STAT*, signal transducer and activator of transcription. (From Kaplan, 2015.)

B-cell subsets also exist such as marginal zone B cells and B1 cells, which like γδ T cells and NKT cells lack the diversity typical of αβ T cells.

Blurring the margins between innate and adaptive immunity

Some cells do not fit into this simple scheme separating innate from adaptive immunity. These includes NK cells, NKT cells and lymphoid tissue inducer (LTi) cells, all of which belong to the recently rebranded set of innate lymphoid cells (ILCs) (Fig. 7.3). γδT cells and some of the subsets of B cells described previously also probably belong here.

Innate lymphoid cells

ILCs lack both αβ and γδ T-cell receptors, a feature that suggests they could be included in the myeloid system, but these cells also have lymphocyte characteristics, expressing the common leucocyte marker CD45 and the IL-7 receptor (CD127). ILCs were originally described in the intestine and later found to occur in the blood stream and in tissues such as the skin. Three groups of ILC have been identified that secrete cytokines normally associated with CD4 T cells: Group 1 secreting IFN-γ and TNFa (Th1-like), Group 2 secreting IL-4, 5 and 13 (Th2-like) and Group 3 secreting IL-22/IL-17 (Th17-like) (Fig. 7.3). In addition, a fourth group of regulatory ILCs has also been described that produce the suppressive cytokine IL10 but do not express FoxP3.

ILCs develop from HSCs during which a subset of myeloid/lymphoid precursor cells transiently express the transcription factor promyelocytic leukaemia zinc finger (PLZF encoded by the gene *Zbtb16*), which sets them on the ILC differentiation pathway. Interestingly, NK cells and LTi cells do not express PLZF although they are included within the class of ILCs (see Fig. 7.3). ILCs are increasingly recognized as mediators of disease: for instance, ILC2s are involved in atopic skin disease through presentation of skin lipid autoantigens and bacterial lipid antigens through the non-conventional antigen-presenting molecule CD1a (Starling, 2018).

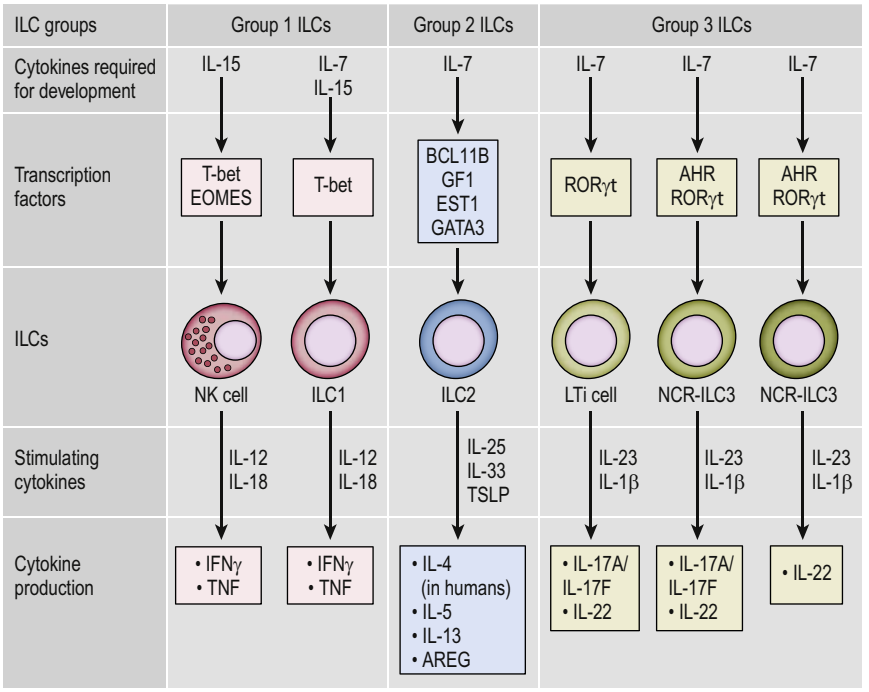


Fig. 7.3 Classification of innate lymphoid cells. Three subgroups of innate lymphoid cells (ILCs) have been defined on the basis of their expression of key transcription factors and production of cytokines. ILCs can be classified as 'cytotoxic' ILCs (*bona fide* natural killer (NK) cells) or as 'helper' ILCs (ILC1s, ILC2s and ILC3s). However, this dichotomy is debatable because some ILC1s can also be cytotoxic (via TRAIL expression). Group 2 ILCs are referred to as ILC2s, whereas group 1 ILCs comprise NK cells and are referred to as ILC1s, and group 3 ILCs comprise lymphoid tissue inducer (LTi) cells as well as cells that are either positive or negative for the natural cytotoxicity receptor (NCR) NKp44 in humans and NKp46 in mice. ILC1s are defined by their expression of T-bet and production of interferon-γ (IFNγ); ILC2s are defined by their expression of GATA3, BCL11B and GFI1 and by their production of IL-4, IL-5, IL-13 and amphiregulin (AREG); and ILC3s are defined by their expression of RORγt and by their production of IL-17A, IL-17F and IL-22. IL-7 is required for the development of ILC2s and ILC3s, whereas both IL-7 and IL-15 are required for the development of ILC1s. Each subset is also activated in response to stimulating cytokines: IL-12 and IL-18 for ILC1s; IL-25, IL-33 and thymic stromal lymphopoietin (TSLP) for ILC2s; and IL-23 and IL-1β for ILC3s. This figure represents a schematic classification of ILCs and, as such, does not indicate all of the transcription factors and immune mediators expressed by ILCs or represent all of the cytokines that are required for their development and activation. (From Ebbo et al., 2017.)

Natural killer cells

NK cells are circulating granulocytes but are classified within Group 1 ILCs (Fig. 7.3) and are particularly effective against tumour cells and virus-infected cells. Previously called null cells because they lack any of the specific lymphocyte markers including a T-cell receptor, they recognize and kill virus-infected cells when activated for instance by the cytokines IL-12 or IL18. Similarly, they have receptors for antibody allowing them to kill antibody-coated cells. NK cells are characterized by inhibitory receptors (KIR), which bind MHC class I, thus preventing them from killing uninfected healthy cells. In contrast, NKT cells are specialized lymphoid cells that only express a single type of TCR that, like ILC2 cells, responds to lipid antigen presented via the molecule CD1a on antigen-presenting cells (thus they are known as invariant NKT cells).

Lymphoid tissue inducer cells

All ILCs apart from NK cells depend on IL-7 for development. LTis are a subset of Group 3 ILCs (see Fig. 7.3) and have characteristics of both NK cells and lymphocytes but are specialized to produce lymphotoxin-α and -β (LTα, LTβ), which are required for the development of lymph nodes and other secondary lymphoid tissue. LTis also secrete IL17 and 22.

Immunological memory might not be exclusive to adaptive immunity

The two hallmarks of adaptive immunity are specificity generated through MHC genes and memory whereby lymphocytes produce an accelerated and heightened response to reinfection or rechallenge with the same antigen. Recently, myeloid cells (macrophages and DC) have also been shown to generate an accelerated and stronger response to rechallenge because of chromatin structural re-arrangement, a phenomenon termed 'trained Immunity'. This is particularly important for allografts where the antigen (alloantigen) persists and continuous recruitment and activation of macrophages to the graft risks graft failure. In fact, targeting these proinflammatory

macrophages (M1) and promoting suppressive macrophages leads to better graft survival (Braza et al., 2018. Immunity). It has been suggested that the effectiveness of BCG vaccination in tuberculosis is based on activation of macrophages with heightened trained immunity (Koeken et al., 2019).

Initial response of the host to injury (the innate immune response)

The acute inflammatory response is the host's initial reaction to challenge. During this response, invasive pathogens are removed by cells of the innate immune system brought to the site of injury by changes in tissue components such as blood vessels and extracellular matrix. Meanwhile antigen from degraded microorganisms is transported to lymphoid tissues to activate the adaptive immune system.

The acute inflammatory response

The acute inflammatory response goes through three phases:

- tissue damage and the acute early response
- the delayed cellular response and phagocytosis
- resolution of the inflammation and tissue remodelling.

The acute early phase has several components:

- tissue damage and release of mediators
- vascular changes
- leucocyte activation and adhesion
- leucocyte emigration.

Tissue damage and the release of mediators

The response to tissue injury (physical, chemical or mediated by microorganisms) is immediate. Reactions occur at several levels, both locally and systemically. Immediate local reactions include the release of tissue factors and chemoattractants (chemokines, see p. 411) from damaged tissue and microorganisms. Vessels are also damaged, inducing venous stasis and the leakage of plasma components; platelet and leucocyte activation with intravascular clotting occurs; plasma/serum transudation and exudation lead to tissue fibrin deposition and activation of serum components such as complement (see p. 411).

There are several classes of inflammatory mediators derived from both inflammatory and damaged tissue cells:

- vasoactive amines (e.g. histamine and serotonin)
- cytokines and chemokines
- lipids (e.g. prostaglandins, thromboxane and leukotrienes)
- free radicals (see Chapter 4, Box 4.6, p. 200 and Box 4.8, p. 204)
- neuropeptides (e.g. substance P, vasoactive intestinal peptide)
- endothelium-derived mediators (endothelin, nitric oxide, prostacyclin, platelet-activating factor, etc.)
- plasma-derived mediators (e.g. complement, kinins and clotting cascade peptides)
- leucocyte-derived mediators (e.g. granule proteases, phospholipase A_2)
- bacterial products (e.g. endotoxin, proteases and chemotactic factors including formylated peptides).

Table 7.1 **Neutrophil antibacterial agents**

Class	Agent
Free radicals/gases	Hydrogen peroxide
	Hypochlorite
	Chloramine
	OH radical
	Nitric oxide
Enzyme	Proteinase 3
	Collagenase
	Elastase
	Azurocidin
	Cathepsin G
	β-glucuronidase
	Myeloperoxidase
	Lysozyme
Peptide	Defensin
	β-lysin
	Vasoactive intestinal peptide
Ion binders	Lactoferrin
	Calprotectin

During the first 20 minutes to 48 hours, there is a progressive increase in polymorphonuclear cell infiltration into the damaged tissue, initiated by ILC3 and γδT cell secretion of IL-17, which accumulate around foci of injury or bacterial proliferation (a process known as 'swarming', which requires leukotriene B4 and integrins at the site of injury). Degranulation of these cells leads to high tissue levels of several proteases, cytokines and cationic proteins. Neutrophils contain some of the most powerful antibacterial agents, including defensins, which have similarity to defensins from other species including plants and insects (Table 7.1). Defensins, and the related molecule cathelicidin (LL37), are leucocyte-derived low molecular weight mediators of non-oxidative bacterial killing and are also produced by epithelial cells such as intestinal and conjunctival mucosal cells. Defensins also direct antiviral activity working on both the virion and the host cell. Among the proteases released by neutrophils during acute inflammation are extracellular matrix-degrading enzymes, which are also released by activated or injured tissue parenchymal cells and are involved in tissue remodelling. More than 24 of these zinc-dependent endopeptidases, known as matrix metalloproteinases (MMPs) (Table 7.2), have been identified in humans and include the transmembrane proteins that contain disintegration and metalloprotease domains (ADAMTs) (Fig. 7.4). ADAMs have a particular role to play

Table 7.2 **Matrix metalloproteinases (MMPs) and their substrates**

MMP	Interstitial collagens	Basement membrane	Elastin	Other proteins
Collagenases				
MMP-1	Types III, I, II, VII, X	Laminin Entactin		
MMP-8	Types I, III, II	Fibronectin		L-selectin
MMP-13	Types II, I, III	(±) Proteoglycan		
Stromelysin				
MMP-3		Fibronectin Laminin	±	EGF-like proteins
MMP-10		Entactin Proteoglycan	±	Plasminogen
Stromelysin-like				
MMP-7		Fibronectin	+	Plasminogen
MMP-12		Laminin Entactin Proteoglycan	↓ ++	↓ Angiostatin; α_1-antitrypsin
Gelatinases				
MMP-2	Types I, VII, X, XI	IV/V Fibronectin	++	
MMP-9		Laminin Entactin Proteoglycan	++	
Membrane-type				
MMP-14 MMP-15 MMP-16 MMP-17 MMP-24	Types I, III, II	Fibronectin Laminin Entactin Proteoglycan		
Furin-recognition site				
MMP-11				

in communications between cells by cleaving proteins such as CD44 and Notch, both of which regulate immune responses. The MMPs self-activate and cross-activate each other in a cascade-like fashion, thus permitting maximal tissue degradation as required. Their effects are counteracted by naturally occurring inhibitors, termed tissue inhibitors of matrix metalloproteinases (TIMPs). There are four types of TIMP: 1, 2, 3 and 4, TIMP-4 only occurring in the mouse. MMPs are also inhibited by conventional antiproteases such as α_2-macroglobulin, tissue factor pathway inhibitor 2, and a more recently described plasma membrane inhibitor, RECK (reversion-inducing cysteine-rich protein with Kazal motifs), which appears to be downregulated in cancer cells, thus allowing them greater metastatic potential.

However, much of the cell and tissue damage is mediated by free radicals, particularly hydrogen peroxide (H_2O_2) and superoxide anions (see Chapter 4, p. 200), which are released as part of the respiratory burst (Box 7.4). Interestingly, flavonoids and adenosine, both important molecules in ocular physiology, inhibit the respiratory burst.

Free radicals may also combine with reactive nitrogen species (nitric oxide) released from inflammatory cells (Box 7.5).

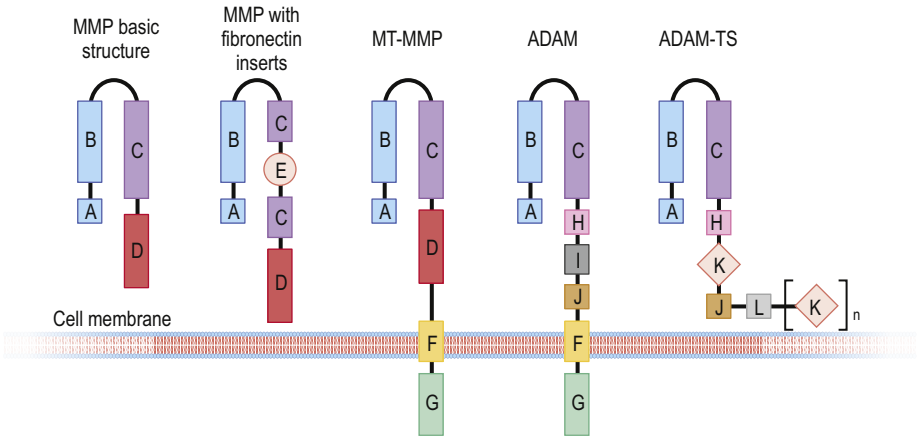


Fig. 7.4 Domain structure of metzincin proteases *(MMPs)*. Each of the domains is shown as rectangles identified by letters, as follows: *A,* signal peptide; *B,* prodomain; *C,* catalytic domain; *D,* hemopexin-like domain; *E,* fibronectin type II insert; *F,* transmembrane domain; *G,* cytoplasmic tail; *H,* disintegrin domain; *I,* cysteine-rich domain; *J,* EGF-like domain; *K,* thrombospondin type I-like repeat; *L,* spacer region. (From Klein and Bischoff, 2011.)

Dramatic vascular changes occur in inflammation

The immediate cause of inflammation is release of plasma into the extravascular space and the instantaneous coagulation of proteins with activation of inflammatory mediators. Plasma release (vascular leakage) is caused by changes to the blood vessels induced by the inflammatory mediators, particularly reactive nitrogen species (RNS), especially in conditions such as the ischaemia–reperfusion damage of stroke and vascular occlusion. The immediate response of the vascular endothelium is to undergo retraction and this is associated with transient vasoconstriction. The major vascular response, however, is vasodilatation mediated initially by nitric oxide (NO).

The initial vasoconstriction is mediated by several locally released compounds, particularly endothelin, released by pericytes and smooth muscle cells to act on the endothelium. Release of gaseous NO to induce second stage vasodilatation is also a local event. NO is synthesized by specific enzymes from the amino acid arginine and has widespread physiological and pathological effects, some of which are related to its role as a free radical (see Boxes 7.4 and 7.5). This is accompanied by an increased blood flow, the opening of capillary channels and the leakage of plasma into the extracellular space. This in turn leads to an increase in the tissue osmotic pressure, thus attracting further fluid build-up in the tissues (oedema). In response to this, there is an increase in lymphatic drainage from the injured site, thus reducing the tissue swelling and at the same time increasing the flow of antigenic material to the draining lymph nodes (see later). These vascular changes vary in degree with the severity of the tissue injury and protective measures are often in place such as the expression of caveolin-1 in microdomains of the vessel wall associated with transient receptor potential cation channel 4 (TRPC4), which mitigates the increased vascular permeability.

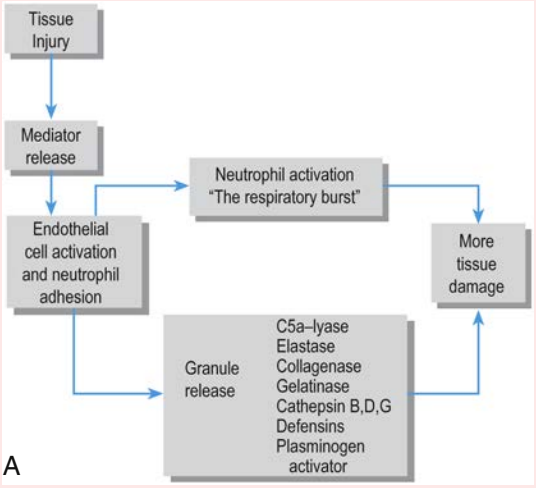
The vascular endothelium also undergoes significant functional and morphological changes during the inflammatory response. Whereas the normal endothelium presents a non-adhesive surface to circulating cells such as platelets and leucocytes, during inflammation the endothelium becomes much more adhesive, an effect achieved by the expression of specific adhesion molecules on its surface. There are three major classes of adhesion molecule, the selectins, the integrins and the cell adhesion molecules (CAMs), each with different functions (see later) (Fig. 7.5 and Table 7.3). In addition, the endothelium may undergo a marked morphological transformation, changing from a flat resting cell to a large protruding cell with multiple cytoplasmic organelles. These inflammation-associated vascular changes occur almost exclusively in the postcapillary venule and, in this respect, this region resembles the high endothelial venule cells in the lymph node, which are specialized for leucocyte adhesion (see later).

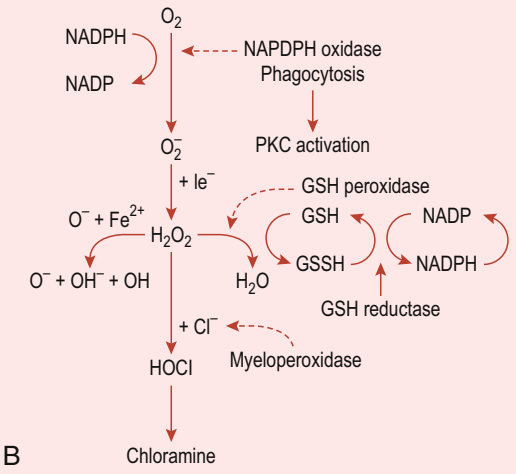
Activation of leucocytes causes them to adhere to the endothelium

Neutrophils are attracted to the site of inflammation through a series of discrete events that occur during margination and extravasation of the cell from the

Box 7.4 The neutrophil respiratory burst

The neutrophil respiratory burst describes the activation of neutrophils and their utilization of oxygen during the inflammatory cascade. The stimulus for this response is the release of mediators from injured tissue cells. Tissue damage is caused by free radical release and tissue proteases (see also Chapter 4, p. 200).





The central component of the respiratory burst is H_2O_2, which is metabolized through several pathways, some of which cause further tissue damage (e.g. superoxide and chloramine) while others reduce it to water. Tissue damage is therefore dependent on the levels of reduced glutathione *(GSH)* in the milieu. In addition, during the respiratory burst, immune cells activate a $Na^{(+)}/H^{(+)}$ membrane exchanger protein, which is required to control intracellular pH and cell volume.

Hypochlossrous acid (*HOCl*) is a short-lived, highly reactive oxidant that is lipophilic and membrane permeant. It binds proteins and renders them more susceptible to proteases. Chlorinated proteins are more immunogenic and may provide a link between the innate and the acquired immune systems by acting as damage-associated molecular patterns (DAMPs).

NADP, Nicotinamide adenine dinucleotide phosphate; *NADPH*, reduced NADP; *PKC*, protein kinase C.

vessel. These involve rolling, loose attachment, firm adhesion and then extravasation/migration of the cell through the intercellular junction. Each of these steps is mediated by the reciprocal expression of adhesion molecules and their respective ligands on the surface of leucocytes and the endothelium (see later) (Table 7.3). During the later stages of the response (24–72 hours) when other inflammatory cells are involved (monocytes and lymphocytes), similar adhesion mechanisms are involved but with different sets of molecules. Thus, the coordinated expression of adhesion molecules appears to regulate the nature of the inflammatory cell exudate.

Adhesion of leucocytes to the endothelium thus involves a series of molecular events

- Selectin–ligand (S–L) interactions occur during the initial rolling phase of leucocyte endothelial cell interactions. These commence as low-strength interactions and are enhanced by the upregulation of selectins on the endothelium by inflammatory cytokines such as interleukin-1 (IL-1) and TNF-α or by contact with an activated T cell.
- Leucocyte activation by chemokines is mediated in part through upregulation of specific chemokine receptors (see later), which induces polarization and firm adhesion of the cell to the endothelium.

Box 7.5 **Nitric oxide**

Nitric oxide was originally described as endothelium relaxing factor because it was found to be the agent released by the endothelium that was responsible for inducing autocrine vasodilatation in response to insult. Nitric oxide is a gas produced by the activity of the enzyme nitric oxide synthase (NOS) on interaction of the amino acid arginine with oxygen:

$$O_2 + \text{L-arginine} \xrightarrow[\text{NOS}]{} \text{Citrulline} + \text{NO}$$

There are at least three isoforms of nitric oxide synthase, endothelium-derived (eNOS), neuronal (nNOS) and inducible (iNOS). The iNOS is released from inflammatory and other cells, particularly macrophages, and is involved in both immunoregulation and tissue damage through its interaction with superoxide radicals released from activated neutrophils. In its latter role it may also function as an antibacterial agent by damaging bacterial cell membranes. The prostanoids are co-released with NO through induction of cyclo-oxygenase-2.

NO interacts with superoxide anions to produce per-oxynitrite, which is believed to be directly involved in membrane damage (see equation). NO has direct effects on many other proteins such as the important zinc finger regions of many enzymes where it nitrosylates free cysteine SH groups with ejection of the Zn moiety from the configured protein. These damaging effects to parenchymal cells may also be the basis for T-cell apoptosis (cell death), leading to downregulation of the immune response.

$$O_2^{\bullet} + NO^{\bullet} \rightarrow ONOO^- + H^{\bullet} \rightleftharpoons ONOOH \rightarrow HO^{\bullet} + NO_2^+ \rightarrow NO_3^- + H^+$$

Thus, NO produced constitutively in small amounts has a physiological role involving guanyl cyclase and increases in cGMP, while in large amounts it may be cytotoxic, producing depression of mitochondrial respiration, metal enzyme damage with consumption of both Zn and Fe ions and DNA damage.

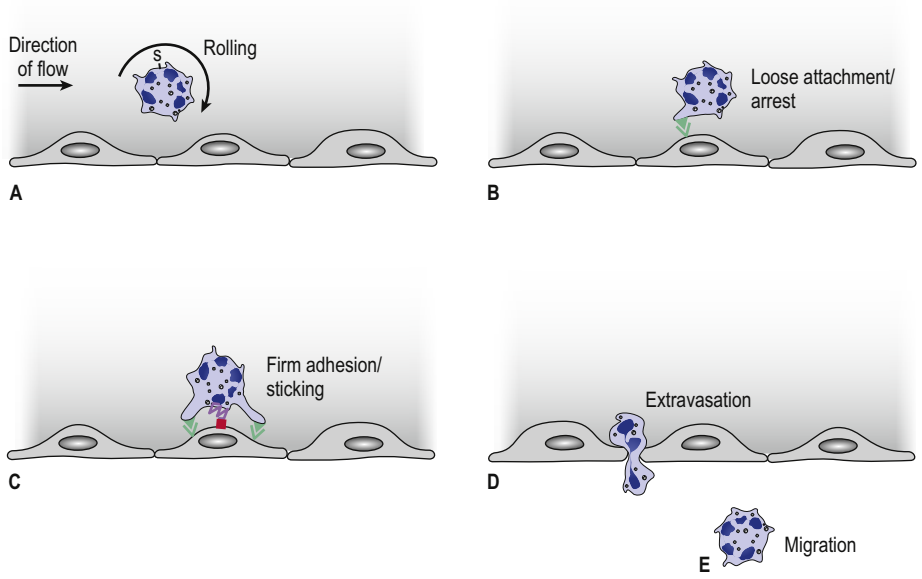


Fig. 7.5 Sequence of events in leucocyte adhesion: rolling (**A**), loose attachment (**B**), firm adhesion (**C**), extravasation (**D**), leading eventually to migration (**E**) of the cells within the tissue towards the site of inflammation.

Table 7.3 Endothelial adhesion molecules and their ligands on leucocytes

	Endothelium	Leucocyte
Loose attachment	P-selectin	PSGL-1, PTX3
	E-selectin	PSGL-1, ESL, CD44
	PSGL, GlyCAM	L-selectin
Slow rolling	ICAM-1	LFA-1/PSGL-1
	E-selectin	PSGL-1, ESL, CD44
Arrest/firm adhesion	ICAM-1	LFA-1
	VCAM-1	VLA-4
	Hyaluronan	CD44
Crawling	ICAM-1	MAC-1
Extravasation	ICAM-1, ICAM-2	LFA-1, MAC-1
	VCAM-1	VLA-4
	CD99	CD99
	PECAM-1	PECAM-1
	JAM-A, -B, -C	LFA-1, VLA-4, MAC-1
	CD99L2	?

The cell behavioural sequence of loose attachment, slow rolling, arrest and firm adhesion, crawling along the inner vascular wall, and finally extravasating into the tissues is utilized by neutrophils that preferentially, if not exclusively, leave from the postcapillary venules.
?, Unknown or opposing data; *CD99L2*, CD99 antigen-like protein 2; *ESL1*, E-selectin ligand 1 (also known as GLG1); *GlyCAM*, glycosylation-dependent cell adhesion molecule; *ICAM*, intercellular adhesion molecule; *JAM*, junctional adhesion molecule; *LFA-1*, lymphocyte function-associated antigen 1; *PECAM*-1, platelet/endothelial cell adhesion molecule 1; *PSGL-1*, P-selectin glycoprotein ligand 1; *VCAM-1*, vascular cell adhesion protein 1; *VE-cadherin*, vascular endothelial cadherin; *VLA-4*, very late antigen 4.

- Integrin–CAM interactions induce spreading of the adherent leucocyte on the endothelial cell surface and thus prevent its detachment.
- Extravasation of leucocytes through the endothelium is mediated by expression of PECAM-1 (CD31) on both the leucocyte and the endothelium, possibly through a 'zipper' mechanism in which disassembly of interendothelial tight (occludin) and adherens junctions occurs with the expression of several junctional adhesion molecules (JAMs; see Table 7.3). CD99 (single chain glycoprotein-1), expressed on both leucocytes and endothelium, is also important in transendothelial migration of monocytes.
- Migration of leucocytes through the tissues is the final stage and is induced through binding of chemokines selective for each cell type, which activate signalling pathways in the translocating cell inducing the motor machinery (actin–myosin cytoskeleton) to propel the cell forwards up the chemotactic gradient (Fig. 7.6).

Certain molecules are specific for each type of leucocyte–endothelial cell interaction; for instance, E- and P-selectins mediate the attachment of polymorphonuclear leucocytes to endothelial cells, while vascular cell adhesion molecule (VCAM) preferentially mediates T lymphocyte–endothelium binding. Both of these interactions have been reported in inflammatory tissue in the eye from cases of sympathetic ophthalmia, a form of autoimmune posterior uveitis.

Leucocytes migrate into the tissues by chemotaxis

Chemotaxis is the directional movement of a cell up an increasing concentration gradient of a specific substance

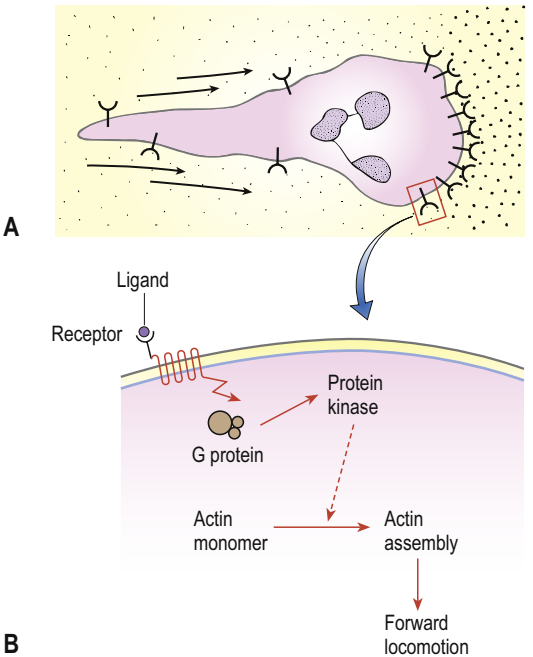


Fig. 7.6 Neutrophil migration up a chemical gradient. (**A**) Cells sense the concentration gradient and polarize towards it with a wide leading edge and a trailing tail. Receptors for the chemoattractant cluster in the membrane of the leading edge, increasing the signal at this end of the cell. (**B**) Diagram of the seven-coil transmembrane receptor (the same basic structure as rhodopsin) that mediates the intracellular signal via G protein links. This activates a protein, tyrosine kinase, to initiate actin assembly and induce forward movement of the cell.

(chemoattractant). Many of the mediators released in the earliest stages of tissue injury are chemoattractants for inflammatory cells and the highest concentrations are present at the site of injury/infection. Both thrombin and the fibrin cleavage product fibrinopeptide B lead to leucocyte chemotaxis from the time of onset of vessel leakage and fibrin formation. Prokaryotic peptides released from bacteria, such as formyl-methionine-leucine-phenylalanine (F-Met-Leu-Phe), are powerful neutrophil and monocyte chemotactic agents. Activated complement components (see p. 411) have an important role as chemoattractant agents in the neutrophil/monocyte response. Other important chemoattractants include interleukin-8 (IL-8), a cytokine (chemokine) released from tissue cells including the retinal pigment epithelium (RPE), tumour necrosis factor-α (TNF-α) and platelet-release compounds such as platelet-activating factor (PAF), transforming growth factor-β (TGF-β), platelet-derived growth factor, platelet-derived endothelial cell growth factor, and many others. Lipid mediators, such as the leukotrienes, are also important neutrophil chemoattractants, and certain cytokines, such as monocyte chemotactic protein (MCP, renamed CCL2) and the macrophage inflammatory proteins (MIP-1α, MIP-1β, renamed CCL3; see below) are selective for mononuclear cell chemotaxis. Several cytokines (soluble immune secreted molecules acting on cells) have multiple effects and act as chemokines (inducing movement through chemotaxis) and receptors for different chemokines are present on different cells, thus regulating not only the numbers of cells that migrate into the tissue but also the type of cells and thus the quality or degree of the inflammatory exudate (see p. 386). Similarly, cells that are considered classic chemokines (CCL2, CCL3) also affect other cell functions such as activation and adhesion.

Cells such as neutrophils and monocytes 'sense' a chemical gradient of these attractants and migrate up the gradient by using specific cell surface receptors clustered preferentially towards the leading, polarized edge of the cell (Fig. 7.6). These receptors (e.g. the C5a receptor) are composed of seven transmembrane segments (in a manner similar to the transmembrane spanning segments of rhodopsin; see Chapter 4, p. 178) that possess a cytoplasmic connection to a G protein-linked second messenger system. This activates the intracellular machinery (actin–myosin motor) required for forward movement. Studies have also shown that activation of G protein receptors occurs in waves, thus enhancing the overall effect of the gradient.

Phagocytosis and removal of damaged tissue and microorganisms

Resolution of the inflammation requires the removal of dead microorganisms and necrotic tissue by phagocytic cells (polymorphonuclear leucocytes and macrophages). Even in the absence of microorganisms, altered (damaged) self-proteins are recognized by cells of the innate immune system and phagocytosed. In the eye this is classically seen with lens-induced uveitis, in which denatured lens crystallin proteins are released into the anterior chamber in cases of traumatic and hypermature cataract. In the latter, engorged macrophages may block the outflow channels and produce a phacolytic glaucoma (see Chapter 9, p. 405).

Phagocytosis, particularly of bacteria, is facilitated by certain molecules of the innate immune system known as opsonins, which are present in plasma and bind to the surface of microorganisms when they are released into the extracellular space. One of the complement components, C3b (see later), acts as an opsonin. C3b is deposited on ageing cells and circulating bacteria and through binding to the complement receptor of the immunoglobulin superfamily (CrIg) expressed on liver Kupffer cells (yolk sac derived macrophages), a complex of immune adherent complement proteins on the opsonized cells is removed. Interestingly, different actin cytoskeletal structures are constructed to phagocytose different types of particle, e.g. IgG-coated versus complement-coated particles.

Molecules of the adaptive immune system also promote phagocytosis, particularly antibodies that bind avidly to the foreign antigen by specific interaction of their antigen-binding site (the Fab portion of the molecule) but are non-specifically removed by binding of their Fc portion to the phagocyte cell surface (see section on antibodies, p. 406). Cells that express high levels of surface receptors for C3b and Fc are termed 'professional phagocytes' and include neutrophils and macrophages. Studies in genetically targeted knockout mice suggest that the Fc pathway takes precedence for phagocytic activation.

Microorganisms activate innate immune cells

Invading microorganisms release factors that attract leucocytes to the site of inflammation and also induce those same leucocytes to engage in attempted removal of the offending agent. This requires activation of the immune cells. How is this achieved? As indicated in the introduction, this was previously considered to be a non-specific process of pathogen–host cell interaction, i.e. there was one general mechanism whereby the innate immune cells of the body reacted to all the different types of microorganisms. However, it has long been recognized that responses to different organisms vary greatly, some being virulent or lethal while others are harmless, and this is partly because of the way different classes of microorganisms activate the initial innate immune cells they encounter. They do this via ligands (PAMPs, see page 389), expressed broadly on classes of microorganisms (e.g. virus, bacteria, fungi) and which bind to class-specific pathogen-recognition receptors (PRRs) (see p. 389). There are several classes of both membrane and soluble PRRs, including the Toll-like receptors (TLR) (Table 7.4), carbohydrate-binding (C)-type lectins, retinoic acid-inducible gene-1 (RIG-1)

Table 7.4 **Toll-like receptors: mediators of inflammation with relative selectivity for different classes of microorganisms**

TLR	Localization	Pathogen-derived agonists	Endogenous agonists	Synthetic agonists
TLR1 and TLR2	Extracellular	Bacteria: peptidoglycan, lipoproteins, LTA Fungi: zymosan	–	Pam_3Cys
TLR2 and TLR6	Extracellular	Bacteria: lipoproteins	Veriscan	MALP2
TLR3	Intracellular	Viruses: dsRNA	mRNA	Polyl:C
TLR4	Extracellular	Bacteria: LPS Viruses: RSV fusion protein Fungi: mannan Protozoa: glycoinositolphospholipids	Saturated fatty acids, β-defensins, oxLDL[a], amyloid-β[a]	Lipid A derivatives
TLR5	Extracellular	Bacteria: flagellin	–	–
TLR7 and TLR8	Intracellular	Viruses: ssRNA	Self-RNA	Imiquimod, R-848
TLR9	Intracellular	Bacteria: CpG DNA Viruses: CpG DNA Protozoa: CpG DNA, haemozoin	Self-DNA	CpG-ODNs
TLR11	Extracellular	Uropathogenic bacteria Protozoa: profilin-like molecule	–	–

[a]Amyloid-β and oxLDL bind to CD36 and a TLR4–TLR6 heterodimer.
CpG-*ODNs*, CpG-containing oligodeoxynucleotides; *dsRNA*, double-stranded RNA; *LPS*, lipopolysaccharide; *LTA*, lipoteichoic acid; *MALP2*, mycoplasma macrophage-activating lipopeptide 2; *oxLDL*, oxidized low-density lipoprotein; *polyI:C*, polyinosinic-polycytidylic acid; *RSV*, respiratory syncytial virus; *ssRNA*, single-stranded RNA.
(From Mills, 2011).

helicases, nucleotide-binding oligomerization domain protein (NOD)-like receptors (NLRs), the scavenger receptors and some soluble PAMPS such as the collectins (including complement proteins) and acute phase proteins. Some PRRs also recognize damage self-antigens (DAMPS, see pages 381 and 389). The components (PAMPs, DAMPs) recognized by the PRRs include pathogen cell wall material, bacterial DNA and proteins, viral DNA and RNA, and lipoproteins from microbes as well as endogenous ligands such as self-DNA (if it is in the wrong place, such as the cytoplasm) and particulate material such as uric acid and cholesterol crystals.

Lipopolysaccharide is a classic microbial product that complexes with a serum protein lipopolysaccharide-binding protein (LBP) and binds to the microbial co-receptor CD14. This then complexes with TLR4 and initiates the signalling cascade (involving the adapter protein myeloid differentiation factor 88 (Myd88)) through several intermediaries transmitted on a nuclear factor-κB (NF-κB) core to activate the macrophage or DC (Fig. 7.7). The activated DC then primes the T cells through presentation of processed antigen in the presence of the necessary cytokines such as IL-12, IL-1 and/or IL-23 generated through the inflammasome.

What is the inflammasome? Further information is available online at https://expertconsult.inkling.com

Do ILCs behave like lymphocytes or innate cells?

ILCs unlike T cells are considered tissue resident cells that are adapted to their specific tissue microenvironment and respond to local stimulation by cytokine. NK cells recognize self and non-self ligands on virus-infected cells, tumour cells and host cells that have increased levels of stress proteins. MHC class I and MHC class I-like molecules (see later) are involved in these interactions. Using receptors such as Ly49H (in mice), natural cytotoxicity receptors and a further receptor termed NKG2D, NK cells recognize a range of virus-infected cells such as cytomegalovirus, myxoviruses and influenza viruses. NKG2D is a receptor for tumour proteins and the MHC class I molecules that they recognize in combination are known as MICA and MICB.

NKT cells, in contrast, express a T-cell receptor (see p. 431) but unlike normal T cells they only express one form of the receptor, which binds glycolipids on a range of organisms such as *Leishmania donovani*, Gram-negative glycosyl ceramides and similar molecules on organisms such as *Plasmodium* and *Trypanosoma* in the context of the non-classical MHC class I molecule, CD1d. Two important ocular microorganisms, *Pseudomonas* and *Staphylococcus*,

PRRS AND THE INFLAMMASOME

TLRs are not so much involved in the phagocytosis and clearance of microorganisms (see section above) as in the maturation of antigen-presenting cells for induction of the adaptive immune response through T-cell activation; for this, different TLRs activate different types of DC (see eFig. 7.3). In contrast, macrophages also express other PRRs, including scavenging receptors, which are primarily involved in phagocytosis. Two main classes of scavenging receptors occur in humans: SR-A, which binds molecules on ageing damaged cells, such as oxidized low-density lipoprotein, as well as some microorganisms; and SR-B, which occurs on other cells as well as macrophages and binds many different types of microorganisms through polyanionic interactions. TLRs are, for the most part, cell surface receptors, although some, such as TLR3 and TLR9, bind intracellular molecules such as viral genomic messenger RNA and nucleotide degradation products such as CpG molecules. In addition, a further set of molecules deals with intracellular pathogen products, the NOD 1 and NOD 2 proteins (nucleotide-binding oligomerization domain family), which like the TLRs have a leucine-rich repeat domain. These proteins are also known as CARD (caspase recruitment domain) proteins and bind unique peptidoglycan bacterial molecules. They are also involved in autoimmunity, especially the inflammatory bowel disease Crohn's disease, which is linked to certain types of ocular inflammation. Activation of an 'inflammasome' complex (see Fig. 7.7) and caspase-1 leads to production of IL-1 and IL-18, which drive the inflammatory process. Interestingly, these molecules have recently been implicated in the pathogenesis of age-related macular degeneration.

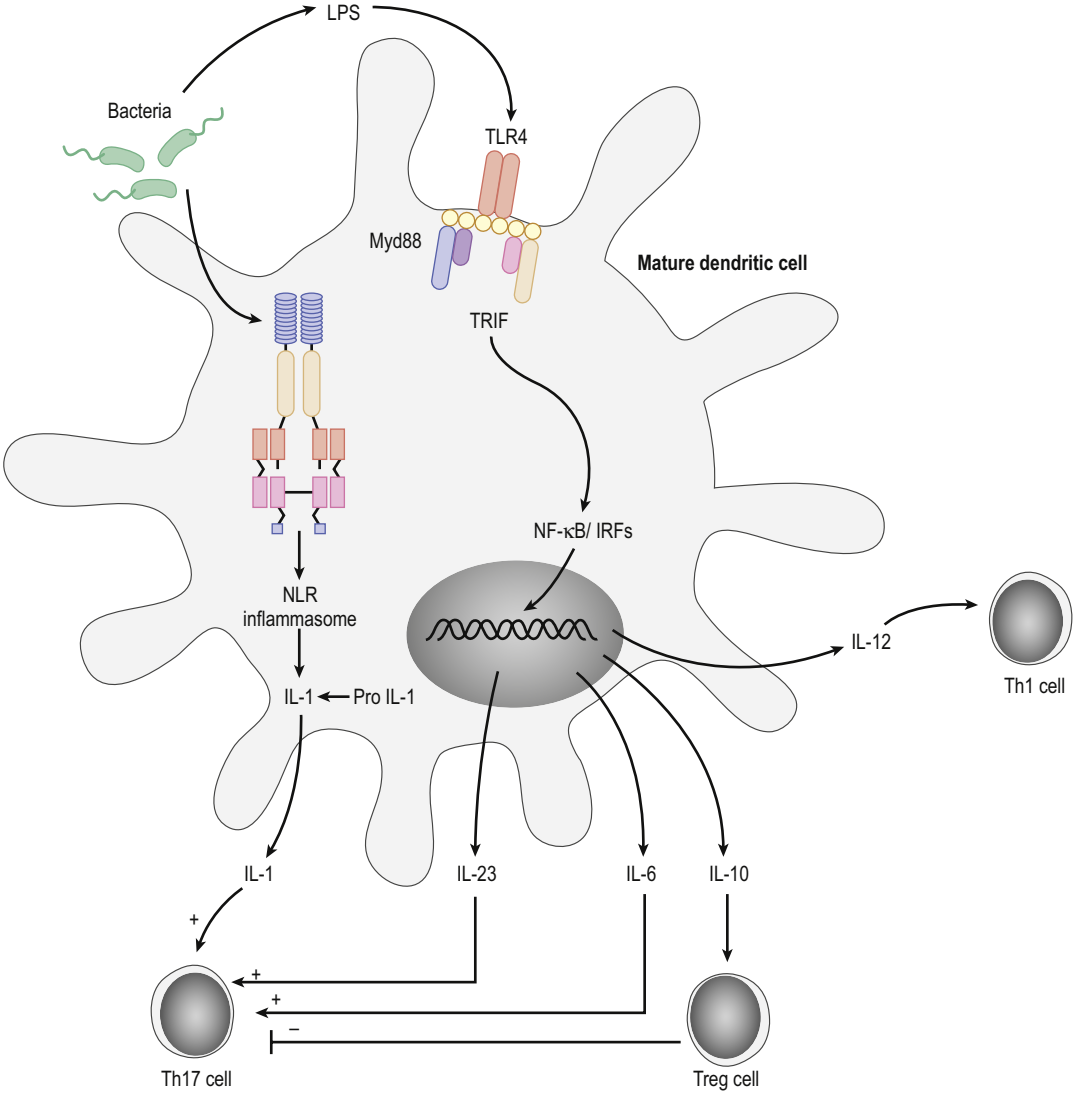


Fig. 7.7 Toll-like receptors *(TLRs)* act on myeloid and other cells by activation of transcription factors but they do not do this directly. Instead they do so through linker proteins (adaptors), which fit them to their particular pathway. All TLRs except TLR3 use the adaptor Myd88 and some TLRs use more than one adaptor, such as TRIF, TIRAP and TRAF. These generate cytokines through two main pathways, one involving NFκB, which leads to transcription of many pro-inflammatory cytokines and the other IRF3, which predominantly leads to production of INF-α. Both signalling pathways also lead to inflammasome activation with production of the central pro-inflammatory cytokines IL-1β and IL-18. Ultimately, these molecules assist in the activation of T cells, during antigen presentation. (From Mills, 2011.)

may also be detected by NKT cells. The cytolytic function of NK and NKT cells occurs via the release of granule contents such as perforin, granzyme and other proteoglycans (see later under cytotoxic T cells).

γδT are not ILCs but are participants in the immediate response to challenge. They colonize the skin and gut epithelium and recognize small alkylamines and pyrophospho-monoesters on microbes and tumour cells. The latter are a major component of mycobacteria. In addition, they recognize heat-shock proteins and have been implicated in inflammatory diseases such as Behçet's disease and the ocular inflammation associated with it.

Effector cells in the inflammation response

Much of the tissue damage in the early stages of the inflammatory response is caused by release of tissue-degrading enzymes such as MMPs (see Table 7.2) from professional phagocytes (macrophages and neutrophils) and as such they are considered to be important as non-specific effector cells in the inflammatory response. Activation of complement during the early phase of the response also provides the materials for non-antibody-dependent complement-mediated lysis of tissue cells, via the membrane attack complex (MAC) of complement and other cells such as NK cells. Macrophages, recruited by T-cell cytokines, also play a major role as effector cells in the adaptive immune response, in addition to antigen-specific cytolytic roles of T cytotoxic (Tc) cells and NK cells. Furthermore, macrophages that have developed trained immunity (see page 392) not only contribute to the ongoing inflammation but further activate newly recruited T cells in an amplifying cycle on increasing inflammation.

Conversely, lymphocytes also play a significant part in the initial acute inflammatory response. Activated γδT T cells release pro-inflammatory cytokines such as IL17. In addition, conventional αβT cells, play a part even in 'sterile wounds', in which microorganisms are not involved, in the overall response. Lymphocytes enter into the site of inflammation at the same time as monocytes and act as bystanders ready for activation by APCs primed with antigen. This may be from degraded organisms or denatured tissue proteins (altered self). In most cases, however, no detectable adaptive immune response occurs locally during the acute inflammatory response unless there is activation of tissue-resident memory T cells (see p. 426), specific for a previous pathogen or antigen.

Resolution of inflammation

Resolution of inflammation occurs when the foreign antigens have been completely destroyed and removed ('cleared') and tissue architecture is restored. Macrophages are critically important in the resolution phase of an inflammatory reaction. Effector pro-inflammatory M1 macrophages (see page 385 and p. 433) are involved in clearance of microorganisms and in the process generate considerable collateral damage ('friendly fire') and are replaced by alternatively activated macrophages, which promote healing (M2 macrophages) either by *de novo* recruitment or by re-programming of inflammatory macrophages. Such cells express high levels of arginase-1, characteristic of inhibitory immune cells, such as myeloid suppressor cells (MSCs) with high potency in inhibiting effector T cells.

Resolution of inflammation involves fibrosis, which is also mediated via alternatively activated M2 macrophages and regulatory macrophages through mediators derived from other cells in the inflammatory milieu (Fig. 7.8). In addition, new vessel formation (angiogenesis), restoration of epithelial surfaces by cell migration and proliferation (see Chapter 3, p. 136) and remodelling of the extracellular matrix by initial deposition of sulphated glycosaminoglycans and hyaluronan deposition, induced by fibroblasts, represent stages in a precisely orchestrated wound-healing process. For instance, angiogenesis is initiated in quiescent endothelial cells by the expression of protease activity on the cell surface and release of growth factors from surrounding inflammatory cells such as fibroblast growth factor (FGF), platelet-derived growth factor (PDGF), platelet-derived endothelial cell growth factor (PDEGF) and vascular endothelial growth factor-A (VEGF-A). VEGF is a major player in the overall wound-healing response: it initiates vascular leakage at the onset of inflammation and promotes angiogenesis in the later stages. The VEGF-C isoform of VEGF activates a specific endothelial cell receptor (VEGF-R3) for induction of lymphangiogenesis, essential for transport of soluble and cell-associated antigen to the draining lymph node.

VEGF induction is itself under the control of hypoxia-inducible factor (HIF1α and β), a transcription factor generated by cells in hypoxic tissue, as occurs in a wound, but is also produced by both M1 and M2 macrophages in the absence of hypoxia. VEGF also plays a major role in the blinding condition, retinopathy of prematurity (see Chapter 2 page 123). However, VEGF-A attracts immunosuppressive cells such as $CCR2^+$ M2 macrophages, which secrete the cytokine CCL2, MSCs, and Tregs and Bregs, all of which promote further angiogenesis. Thus, wound healing, acute inflammation, innate immunity and adaptive immunity are all part of a coordinated response by the organism to remove the offending microorganism and restore tissue homeostasis.

The entire process is driven by changes in metabolism. Activation both of innate immune cells in the early stages of tissue damage and of adaptive immune cells in the generation of immunological memory involve a major shift from resting oxidative metabolism to energy-dependent aerobic glycolysis and consumption of ATP. Even in the absence of hypoxia, this activates HIF1a and VEGF production with its many consequences for wound repair and eventual resolution of inflammation.

Chronic inflammation

When the foreign antigen is not completely removed, the inflammatory response enters a chronic state characterized by mononuclear inflammatory cells such as monocytes and lymphocytes, often arranged in granulomas (see Chapter 9, p. 505). Failure to remove the foreign antigen may occur because of an inadequate initial response or because the pathogen effectively evades the immune system. Intracellular bacteria that evoke phagocytic destruction are a prominent

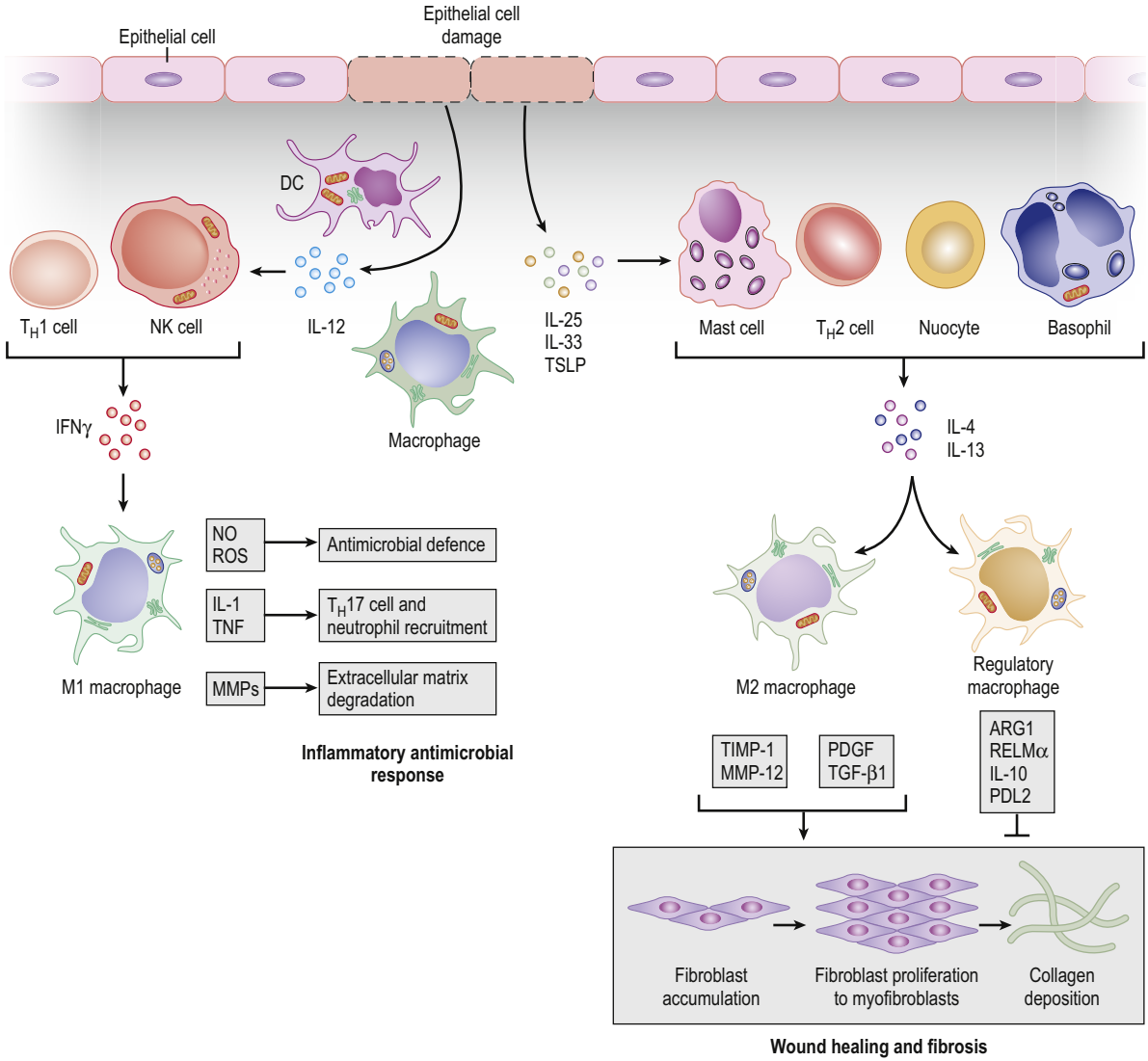


Fig. 7.8 The initial stages of the inflammatory response involve many early mediators including clotting factors, fibrinolytic products, cytokines such as IFN-γ, which programme macrophages towards a pro-inflammatory phenotype (M1). Th2-type cytokines (IL-4, IL-13), released from other cells such as mast cells, generate alternatively activated macrophages as well as regulatory macrophages, which promote healing (fibroblasts and collagen deposition) as well as new vessel formation. *DC*, Dendritic cell; *NK*, natural killer; *TNF*, tumour necrosis factor. (From Murray and Wynn, 2011.)

stimulus for induction of chronic inflammation. Persistence of innate immune cells such as neutrophils in swarms is a feature of many chronic inflammations and signals a failure to clear the foreign antigen.

However, even if the organisms are cleared, prevention of chronic inflammation requires the production of specialized pro-resolving mediators (SPMs) such as omega-3 fatty acid-derived lipoxins (see Chapter 4, p. 194), and other more recently described molecules such as resolvins, protectins, isoprostanes and maresins. These mediators are PUFA-derived (see Chapter 4 page 269) products such as eicosapentaenoic acid (EPA), docosahexaenoic acid (DHA) and the more recently recognized n-3 docosapentaenoic acid (n-3 DPA) (Fig. 7.9) and inhibit leucocyte migration and suppress effector innate and adaptive immune cells. CNS tissues such as the brain and retina are rich in these

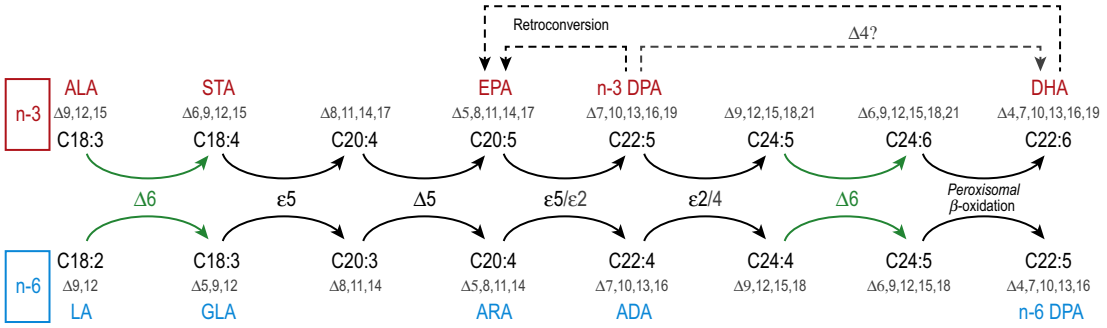


Fig. 7.9 Bioconversion pathways of n-3 and n-6 polyunsaturated fatty acid families. Δ, Desaturase; ε, elongase; *ADA*, adrenic acid; *ALA*, α-linolenic acid; *ARA*, arachidonic acid; *DHA*, docosahexaenoic acid; *EPA*, eicosapentaenoic acid; *GLA*, γ-linolenic acid; *LA*, linoleic acid; *n-3 DPA*, n-3 docosapentaenoic acid; *STA*, stearidonic acid. (From Gaetan et al., 2019.)

compounds and they are thought to have a role in providing neuroprotection in glaucoma and age-related macular degeneration (AMD).

In chronic inflammatory disorders, the acquired immune response also participates, but it too appears to be insufficient to remove the foreign antigen. This may be because the antigen has fooled the immune system and persists *in vivo* by parasitizing inflammatory cells, as in the worldwide blinding disease trachoma caused by the parasite *Chlamydia trachomatis* (see Chapter 8, page 482). Similarly, autoimmune disease (see p. 443) is chronic because 'altered self' antigen is continually present, acting as a danger signal, and induces persistent inflammation. In certain circumstances a low-grade lymphocytic activation occurs. Such lymphocytes may release cytokines which induce fibroblast activity, such as TGF-β and connective tissue-activating peptides (CTAP-1 to CTAP-6). CTAPs are low molecular weight compounds released from leucocyte and platelet granules during inflammation, which themselves undergo partial degradation to produce other pro-inflammatory peptides such as neutrophil-activating peptides (NAP-1 and -2), thereby sustaining the inflammatory response. If this response is excessive, subepithelial fibrosis may occur and produce conditions such as benign mucous membrane pemphigoid and subretinal fibrosis, both extremely debilitating and blinding diseases (See Chapter 9, page 506). CTAPs have been implicated in the fibrosis of Graves' ophthalmopathy through activation of the insulin-like growth factor receptor on orbital fibroblasts.

In many chronic inflammations, the balance of competing mediators and cells may decide the outcome of disease. For instance, in hepatitis caused by HepB or HepC virus, IL-22 produced by Th17 cells or by ILCs, or even by specialized IL-22-producing T cells, can have either a protective role promoting fibrosis or a pro-inflammatory role leading to liver failure. Thus, in chronic inflammation the distinction between innate and adaptive immune responses becomes blurred and such disorders manifest as a mixture of low-grade inflammatory activity with partial attempts at healing (fibrosis). This is well demonstrated in the subretinal neovascular membranes of chronic posterior uveitis.

The systemic response to acute inflammation: the acute-phase reaction

Although the acute inflammatory response is initiated at the site of tissue injury, systemic effects are produced in proportion to the level of tissue damage and virulence of the organisms. These effects are mediated primarily by cytokines acting in this situation at a distance and are known as the acute-phase response. These cytokines include the 'alarm' cytokines, IL-1, IL6, and TNF-α, released mainly through 'danger' signals, from macrophages activated by mast cell and platelet degranulation and/or directly by bacterial products such as endotoxin, peptidoglycan and degraded nucleotides through their PRRs (see earlier). During the acute-phase response, adhesion molecule expression on vascular endothelium is induced and initiates further rounds of inflammatory cell accumulation and cytokine release. T-cell trafficking is induced by raised core temperature (40°C) through a heat-sensing pathway in which heat shock protein-90 (HSP-90) induces expression of α4 integrins and promotes T cell adhesion to vessel walls and entry into the tissues (Lin et al., 2019). In addition, changes in vascular tone are caused by release of low molecular weight metabolic products including the prostaglandins PGI_2, PGE_2, PGD_2 and $PGF_{2\alpha}$ (vasodilatation), thromboxane A_2 (vasoconstriction), and leukotrienes C_4, D_4 and E_4 (smooth muscle contraction).

The effect of massive systemic cytokine release (such as occurs in the 'cytokine storm' of acute sepsis) is to induce a fever response by the direct action of IL-1 and IL-6 on the hypothalamic temperature control system and second, to induce hepatocyte gene transcription for several 'acute-phase reactants' such as C-reactive protein, serum amyloid

components A and P, α_1-glycoprotein, C3 and collectins (see later section on complement) in addition to mannan-binding proteins and haptoglobin. Fibrinogen, α_2-macroglobulin and α_1-antitrypsin are also synthesized and hypercoagulation/thrombocytosis occur.

While many plasma proteins rise in concentration, others such as albumin and transferrin fall. Clinically, the acute-phase response manifests as an elevated erythrocyte sedimentation rate (ESR) caused by more rapid settling of IL-6-mediated, fibrinogen-mediated rouleaux formation of red blood cells.

Many of the acute-phase proteins enhance existing innate defence mechanisms such as C-reactive protein, which acts as an opsonin and binds complement. Others act to inhibit the effect of inflammatory cytokines such as serum amyloid protein and IL-1. The acute-phase reaction has been described as innate immune 'brinkmanship' in that both the pathogen and the host are stressed but the overall outcome aims to limit systemic sepsis while avoiding disseminated intravascular coagulation (DIC), shock, organ failure and death. For instance, in Bunyavirus infection, a fatal outcome is more likely if interferon beta levels are reduced together with inhibition of TLR3 expression (Song et al., 2017).

Development of adaptive immunity and immunological memory

Normally, initiation of the adaptive immune response does not take place at the site of injury or penetration by foreign organisms. Instead, antigen is taken up by APCs at the site of inflammation and transported to the regional lymph nodes and/or spleen where it is presented to T and B cells. T and B cells specific for that antigen respond by undergoing a series of activation steps including cytokine production and proliferation (clonal expansion). Such T cells are known as effector T cells, which pass into the blood circulation and migrate ('home') back to the site of injury where CD8 T cells (Tc cells) kill infected cells, and CD4 T cells (Th cells) attract other pro-inflammatory phagocytic cells such as activated M1 macrophages. In the lymph node, specialized follicular dendritic cells (fDC) present antigen to B cells, which differentiate to become antibody-producing plasma cells. Some plasma cells migrate to the bone marrow but most remain in the draining lymph node follicles and germinal centres, producing high levels of specific antibody, which is released into the circulation (see later). T follicular helper (T_{FH}) cells are also instrumental in progressing B cells through the various stages of proliferation, differentiation to plasma cells, and antibody class switching (see later) via the action of IL4 and IL21.

Antigen recognition is made possible by antigen-presenting cells

Foreign antigen is *presented* to T cells by three types of 'professional' antigen-presenting cell: macrophages, B cells and DCs. However, antigen is *recognized* by specific T cells only after it has been processed and made presentable in an appropriate form to the T cell. Some antigens are recognized by T cells without processing, but they are very unusual. Antigen can also be presented by 'non-professional' APCs such as endothelium but the T cells may not become activated (see p. 443, discussion on tolerance).

Not all proteins induce an immune response and some are too small (haptens) but can become immunogenic if they are be combined with a carrier protein. There are differences in the type of antigen that each of the three cells can present. Macrophages and B cells usually recognize antigen through the immunoglobulin (Ig) molecule. Thus, macrophages and B cells can initiate an immune response only if the host has already been exposed to that antigen and has the capacity to mount a 'memory' response in the form of IgG. In contrast, DCs can process and present antigen to resting, naive T cells, i.e. cells that have not previously 'seen' the antigen. Accordingly, DCs are considered to be the cells that initiate immune responses to new antigens, while macrophages and B cells may be important in sustaining the response to antigen that persists in the tissue.

From the earliest stages of an inflammatory response, DCs at the site of injury (bone marrow-derived cells, mostly extravasated into the tissues from circulating monocytic precursors see p. 388) start to migrate in large numbers from the subepithelial layers into afferent lymphatics and thence to the regional lymph nodes. During this phase, they prepare the antigen by combining it with MHC class I and II molecules so that it can be presented as a complex to T cells. T cells will respond only if they possess the specific receptor (the T-cell receptor, TCR) for that antigen and if the antigen is sufficiently immunogenic, i.e. has the capacity to activate innate immune cells via PRRs (see p. 381). In the normal course of events, it is likely that the great majority of processed antigens never get as far as initiating a perceptible T-cell response.

T cells respond to antigen by clonal expansion

If the antigen is presented to the specific T cell that recognizes it in a suitable form and in the presence of the correct co-stimulatory signals, the T cell responds by clonally proliferating, i.e. it rapidly divides, producing many daughter cells, which are all exactly the same in their recognition of that antigen alone. This very dramatic response usually occurs in lymph nodes or in MALT and accounts for the enlarged lymph nodes and non-encapsulated lymphoid tissues such as tonsils and adenoids seen for instance after a viral

infection. It has been estimated that up to 20% of lymphocytes in an enlarged lymph node are specific for that virus.

The expanded T cells differentiate into various subtypes (Fig. 7.2) including T_{FH} cells, which migrate to the B-cell follicles in the lymph nodes and MALT (see p. 422) and release a range of cytokines that 'help' B cells to expand clonally and differentiate in an antigen-specific manner. However, the majority of activated T cells (T effector cells) enter the circulation and home to the site of injury where they assist in mounting the antigen-specific effector response that will eliminate the foreign antigen.

T and B cells participate in the effector response

Effector responses are those that actually mediate the immune response to clear the pathogen and restore homeostasis. Activated Th cells release cytokines that activate other cells in addition to B cells. These include:

- cytotoxic T cells (Tc cells) that recognize intracellular foreign antigen when it is presented on the surface of tissue cells complexed with MHC class I antigen (see later)
- macrophages that, when activated, remove foreign antigen and perpetuate the immune response by engaging in local antigen presentation at the site of inflammation (if this goes awry, conditions such as lepromatous leprosy and sarcoidosis can develop)
- B cells that are stimulated to full differentiation as plasma cells with considerable local specific antibody production (see later); soluble antibody is then available to form immune complexes and participate in further local antigen presentation and antibody-mediated cytotoxic reactions via NK cells (see later).

How does the organism deal with intracellular antigen?

In the above scenario, we considered how adaptive immunity is triggered when the foreign antigen is extracellular and is a target for phagocytosis. However, what happens if the antigen has already invaded and infected cells and persists as a viable entity intracellularly? This particularly applies to organisms such as viruses and protozoa, which infect many cell types and even bacteria such as mycobacteria, which evade the immune system by 'hiding' inside cells. Such cells may be killed by NK cells or by sensitized Tc cells recognizing viral peptide on MHC class I surface antigen as part of a memory response. The apoptotic and dying infected cells containing virus or other microorganisms are then 'cleared' (phagocytosed) by macrophages and DCs at the site of injury, which, after trafficking to the draining lymph node, further activate antigen-specific T cells through presentation of viral antigen (peptide) in conjunction with self MHC class I. Thus, Tc can be directly activated on appropriately conditioned DCs without obligatory help from Th cells.

Effector mechanisms

As mentioned previously (see page 381), the innate immune system as the first line of defence has a range of effector cells that remove damaged tissue and dead microorganisms (DCs, macrophages, NK cells, γδ T cells, ILCs). The adaptive immune system also has a variety of effector mechanisms, which it uses to rid the host of specific foreign antigen. These include antibodies and cells, but the adaptive immune system also uses non-specific mechanisms that are activated by antigen-specific cells and molecules, including complement and cytokines.

Antibodies

Antibodies are distributed in the endoplasmic reticulum, the Golgi apparatus and the surface of B cells and are secreted in large amounts form plasma cells into the extracellular fluids. Each antibody binds uniquely to a single antigen, which is usually a short sequence of an immunogenic molecule (an epitope). Most antigens are proteins, but carbohydrates and lipids can also act as antigens.

All five antibody isotypes have a similar basic structure

Antibodies are Y-shaped molecules composed of light and heavy chains (Fig. 7.10). Two identical heavy (H) chains linked by disulphide bridges to two identical light (L) chains, either a κ chain or a λ chain, form the basic structure of antibodies. Each chain is composed of a series of repeating homologous units, about 110 amino acids in length, comprising discrete immunoglobulin (Ig) domains. Many other molecules adopt a similar folded structure and are classed together in the Ig superfamily. Differences exist in the precise geometry of the molecules, as shown by crystallography, which has significance for antigen binding. There are five immunoglobulin isotypes IgM, IgG, IgE, IgA and IgD.

Further information on features and special characteristics of immunoglobulin isotypes is available online at https://expertconsult.inkling.com

Immunoglobulin isotypes not only vary on the basis of the number of immunoglobulin chains that make the definitive molecule but also the ends of the heavy chains show considerable and conserved amino acid variations. Similarly, the light chains exist in two isotypes, κ and λ, which vary in proportion (60:40 in humans, 95:5 in mice). In addition, surface membrane-bound immunoglobulin

IMMUNOGLOBULIN ISOTYPES AND THEIR CHARACTERISTIC FEATURES

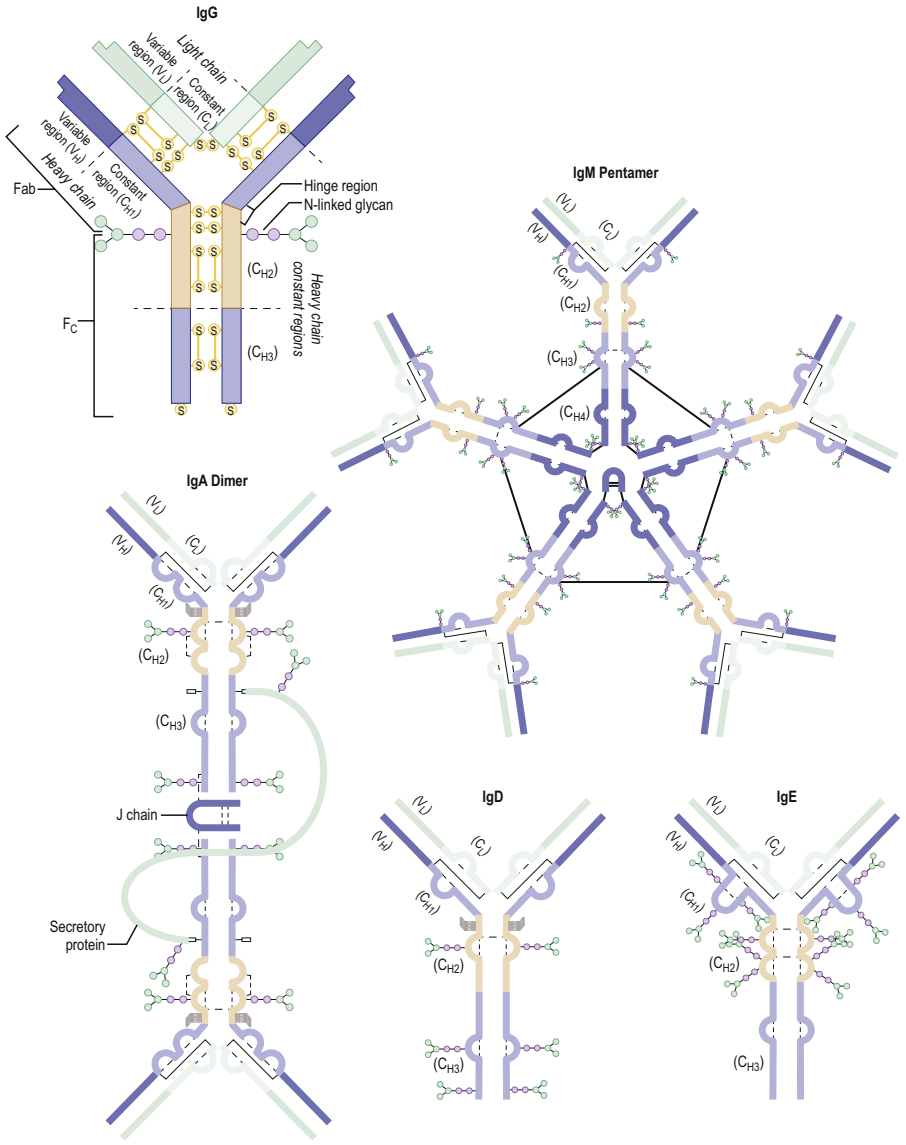
Five isotypes of immunoglobulin exist, denoted by their heavy chain (α, γ, δ, ε, μ); there are two light chains (κ , λ), which occur in a ratio of 60:40 in humans (eFig. 7.4). The Fc and C3 binding regions are responsible for the effector functions of antibodies. Three complementarity determining regions (CDRs) are nested within framework regions, in the V and hyper-V regions of the H and L chains, and receive contributions from both chains. Somatic mutations in the CDRs are responsible for the enormous antibody diversity and affinity maturation that occurs on repeated exposure to antigen. Although not directly involved in antigen binding, the framework region determines the folding of the molecules and thus the amount of complementarity-determining region (CDR) that is presented on the surface of the variable sections of the molecule for interaction with antigen (the antibody-binding site) (see Fig. 7.10).

The secretory forms of IgM, IgA and IgD have C-terminal extensions (tail pieces), which allow multimer formation by attachment to the J chain, while N-linked oligosaccharides bound via asparagine residues contribute greatly to the differences in the overall conformation. IgG comprises several subclasses: IgG_1, IgG_{2a}, IgG_{2b}, IgG_3, and IgG_4. Complement binds to the constant regions of the immunoglobulin molecules.

Limited proteolysis of antibody produces fragments: papain attacks the hinge region, producing single Fab fragments that bind one antigen molecule; pepsin attacks the second CH segment, producing $(Fab)_2$ fragments that can bind two antigen molecules (eFig. 7-4).

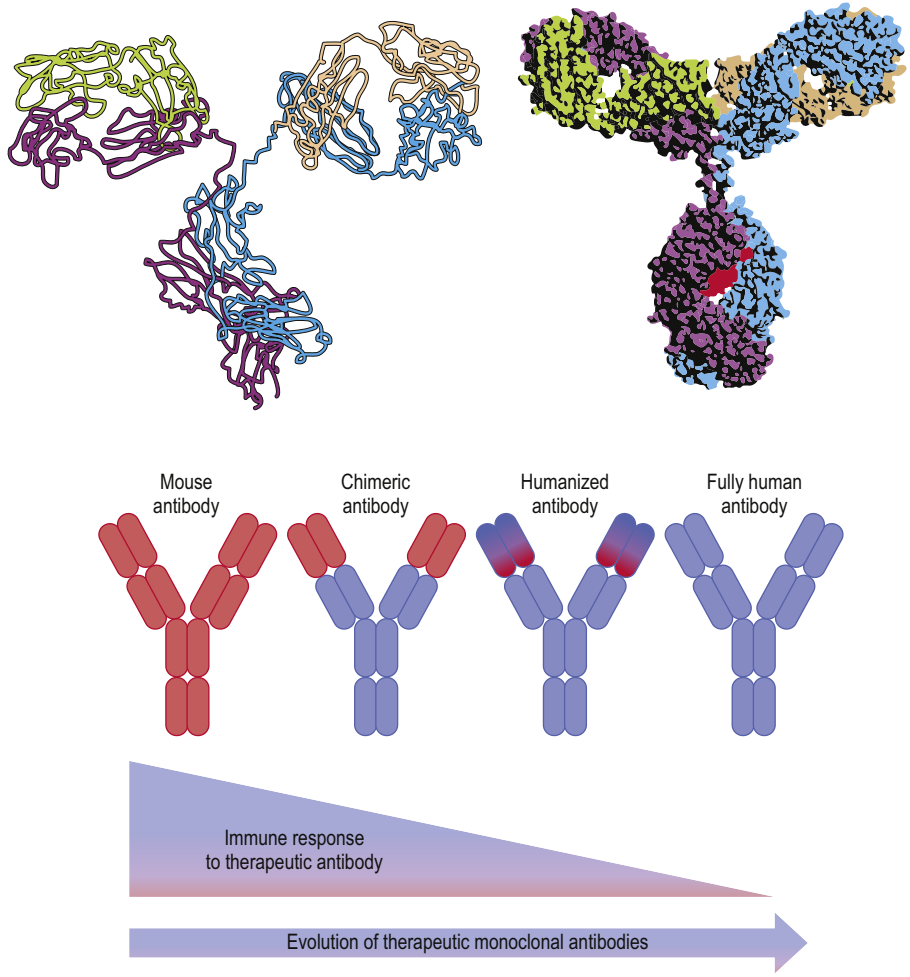
There are three components or segments to the CDR (CDR-1, -2 and -3) and the third (CDR-3) in both heavy and light chains is the most variable (see Fig. 7.10). The C regions on the heavy chains are globular structures attached to the binding region by a flexible rod-like hinge portion of the molecule (see Fig. 7.10 and eFig. 7.4). The hinge region has both rigidity (conferred by proline residues at the top of the CH rod) and flexibility as the result of a large number of glycine residues. The last CH domain of immunoglobulin has a transmembrane and cytoplasmic portion, involved in intracellular signalling. Each domain is composed of two layers of β-pleated sheets with three or four strands of anti-parallel polypeptides. Therapeutic monoclonal antibodies are used extensively for many diseases including eye diseases such as age-related macular degeneration (AMD, see Chapter 9, p. 27) in which a humanized (eFig. 7.5, lower panel) monoclonal antivascular endothelial growth factor (anti-VEGF) antibody is administered to patients with wet AMD by successive injections of the antibody into the vitreous chamber of the eye. Similarly, a humanized anti-TNFα antibody has been shown to be very effective in controlling non-infectious uveitis.

IMMUNOGLOBULIN ISOTYPES AND THEIR CHARACTERISTIC FEATURES—CONT'D



eFig. 7.4 Five isotypes of immunoglobulin exist, denoted by their heavy chain (α, γ, δ, ε, μ); there are two light chains (κ , λ), which occur in a ratio of 60 : 40 in humans. The Fc and C3 binding regions are responsible for the effector functions of antibodies (see below). Three complementarity determining regions (CDRs) are nested within framework regions, in the V and hyper-V regions of the H and L chains, and receive contributions from both chains. Somatic mutations in the CDRs are responsible for the enormous antibody diversity and affinity maturation that occurs on repeated exposure to antigen

IMMUNOGLOBULIN ISOTYPES AND THEIR CHARACTERISTIC FEATURES—CONT'D



eFig. 7.5 Antibody structure. The globular domains of the Fab and the Fc fragments are shown in the panel on *top right* and the anti-parallel pleated sheets on the *top left*. Development of engineered monoclonal fully humanized antibodies is shown on *lower panel*. (http://www.dsch.univ.trieste.it/~benedetti/antibody_catalysis.htm http://dict.space.4goo.net/dict?q=antibody. http://www.chiscientific.cn/service.aspx?ID=30)

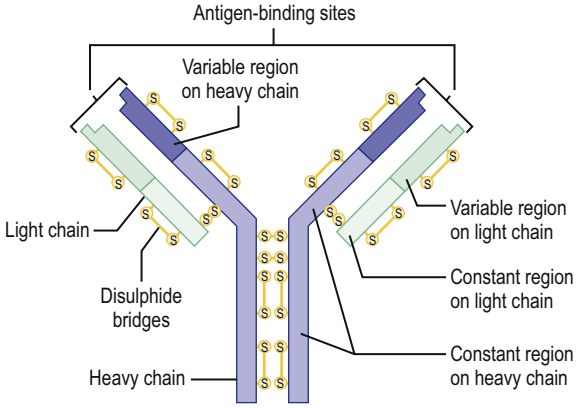


Fig. 7.10 Diagram of generic immunoglobulin structure showing the light chains, heavy chains and the variable and constant regions with the linking disulphide bridges. (Figure adapted from http://www.emc.maricopa.edu/faculty/farabee/biobk/BioBookIMMUN.html. Original image taken from Purves et al., *Life: The Science of Biology*, 4th ed. Sinauer Associates, www.sinauer.com, with permission.)

(sIg) is known as the B-cell receptor and contains a transmembrane region and an intracellular cytoplasmic domain.

Antibodies produced in response to the initial encounter with antigen are of the IgM or IgD isotype, while later antibody production (particularly that in response to rechallenge by the same antigen) is of the IgG, IgA or IgE isotypes. This *isotype switching* (see page 407) is regulated by the enzyme cytidine deaminase, activated through CD40L–CD40 interactions.

Antigen–antibody binding

The specific binding site on the antigen is known as the epitope. Antibody binding to carbohydrate or lipid antigens is strictly dependent on the structure of the antigen. Binding to protein antigens may, however, depend on either the linear sequence of amino acids or the three-dimensional conformation of the molecule. This can lead to the appearance of cryptic epitopes, neo-epitopes, in protein molecules, which are uncovered after changes to the three-dimensional conformation by partial hydrolysis. Neo-epitopes are considered important drivers of autoimmunity and the expression of neo-epitopes in cancer cells often fails to induce a strong enough immune response.

Further information on antibody epitopes and effector mechanisms is available online at https://expertconsult.inkling.com

As indicated previously, antigen specificity is determined by complementarity between the epitope on the antigen and the CDR. However, this is not exclusive and other neighbouring regions on the antigen (paratopes) and on the IgG molecule outside the hypervariable region may influence the final antigen specificity and avidity. In addition, antibodies as proteins may have other properties relating to their non-ligand (antigen) binding sites, for instance the recombination of the V_H and V_L chains may fortuitously produce a site that binds ADP or acts in a catalytic fashion; or it may have regulatory properties or assist in antigen presentation. Such antibodies may therefore have additional functions that are more closely ascribed to innate rather than adaptive immunity. Such antibodies are termed superantibodies.

Monoclonal and polyclonal antibodies

Any single antigen, especially large proteins, might have multiple antibody-binding regions (epitopes). Antibody responses may therefore be polyclonal, oligoclonal or monoclonal depending on the immunodominance of the antigenic determinants. Selection of cells from an immunized mouse under special conditions and fusion of those cells with an immortalized cell line or clone is a powerful technique for producing large quantities of antibody to a specific (single or mono) antigenic epitope.

Monoclonal antibodies (Mabs) have revolutionized diagnostic techniques in medicine today, particularly using flow cytometry in which antibodies specific to cell proteins are 'tagged' with fluorescent markers that identify even very rare cell types using narrow-wavelength lasers. Antibody therapeutics (biologics) are revolutionizing management of diseases such as cancer and autoimmunity, particularly in ophthalmology where anti-TNF-α therapy is used to treat uveitis and anti-VEGF therapy is used to block blood vessel growth 'wet' AMD. Some of these Mabs are chimeric antibodies, i.e. molecules that combine human and mouse antibody segments, while others are fully 'humanized' (see eFig. 7.4). Fragments of antibodies are sometimes as effective as the intact antibody and they possibly have better penetration into tissues as therapeutic agents as a result of their reduced size (Box 7.6). In addition, antibodies are being engineered to direct drugs and other therapeutic agents such as cytokines to sites of disease. Biologics as a term is not exclusive to antibodies but includes any biological large molecule therapeutic drug. Biosimilars resemble biologics functionally but are usually smaller (see Chapter 6, p. 376, for more details).

Cytokines are the effector elements released by cells during innate and adaptive immune responses

Almost all of the biological effects of T cells are mediated by cytokines. More importantly, T cells alter the characteristics of an immune response by releasing different cocktails of cytokines. In a feed-forward mechanism dictated by the cytokine milieu of the tissues, naive T cells differentiate into one or more of the several different T-cell types (Fig. 7.2).

EPITOPES AND ANTIBODY EFFECTOR MECHANISMS

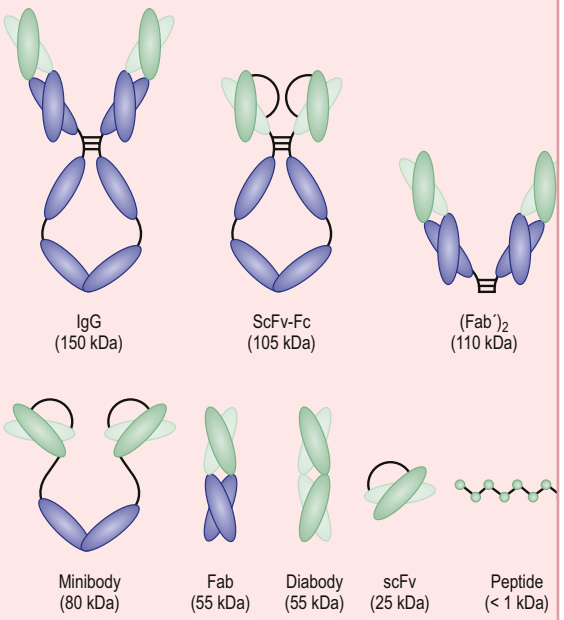
Epitopes on protein antigens can be overlapping or non-overlapping. Overlapping epitopes lead to competition for binding; in such circumstances antibodies to an overlapping region might sterically inhibit presentation of a peptide to T cells. This, however, would occur only with MHC class II–peptide complexes where the binding site for the TCR embedded in the MHC groove overlaps with a binding site for antibody. This mechanism has been suggested for the inhibition of experimental uveitis with an antigen-specific monoclonal antibody. Antibodies may also compete by allosteric mechanisms in which binding of the antibody alters the conformation of the molecule so that it does not bind a second antibody

B cells arise from precursors in the bone marrow as immature B cells expressing IgM, Igα and β and go through a series of transitional stages to become naive mature B cells in the spleen at which stage they express IgD. A small proportion of these cells become marginal zone B cells in the spleen and short-lived IgM-producing plasma cells, while the majority are available for interaction with specialized B-cell follicle (follicular) DCs to initiate antigen-specific immunoglobulin production through the B-cell receptor (BCR). The BCR, similar to the TCR, signals through cytoplasmic transcription factors such as Syk and Lyn, in addition to PI3 kinase and the phosphatase SHIP, to generate plasmablasts in the B-cell follicle. These cells leave the secondary lymphoid tissue to return to the bone marrow and other tissues, where they function as long-lived antibody-producing plasma cells. Isotype switching, i.e. from IgM to antigen-specific IgG, E and A, occurs while in the germinal centre of the B-cell follicle.

Box 7.6 Antibody engineering for therapeutic use

Monoclonal antibodies are produced by the fusion of antibody-secreting B cells with an immortalized myeloma cell line, engineered in such a way that only the B-cell/myeloma fusion cells survive, thus producing an immortalized antibody-producing cell. Some monoclonal antibodies have been produced continuously for decades.

The antibodies themselves can be engineered to produce fragments such as Fab fragments or single chain variable fragments *(scFv)* or even antibodies that combine a double set of variable fragments *(diabodies)* (see figure). In this way, the antibodies can be made more selective and potent.



Cytokines, in the broadest sense, are produced by a wide variety of leucocytes and tissue cells, particularly monocytes/macrophages, epithelial cells and fibroblasts. The response of tissues to invasion by viruses and bacteria is to produce cytokines; for example, virus-infected cells activate NK cells, which can be induced to release cytokines by innate recognition of viral double-stranded RNA (e.g. IL-1, IL18); bacterial lipid (endotoxin) is recognized by CD14 and TLR4 on monocytes and by complement, all leading to pro-inflammatory cytokine release, e.g. IL-12 and IL-23 (see page 411).

What makes a cytokine a cytokine?

Cytokines have the following properties:

- they are secreted by cells in response to a specific stimulus;
- they are short-lived and short-range molecules, acting on cells within their neighbourhood;
- they may have effects at a distance if they are liberated into the circulation in sufficient concentration;
- they are effective at very low concentration;
- they may secondarily induce cytokine release by the target cell;
- they may act upon many different cell types (pleiotropism) and may have multiple different effects on the same target cell;
- they may be redundant, may induce cytokine synthesis themselves and may alter the effects of other cytokines.

Cytokines have three general effects:

- they regulate the innate immune response;
- they regulate the adaptive immune response;
- they regulate the growth and differentiation of haematopoietic cells.

Cytokines now include several groups of molecules, such as interleukins, growth factors, colony-stimulating factors, transforming growth factors, interferons, tumour necrosis factors, chemokines and monokines (cytokines produced by monocytes). Considerable functional overlap exists between these groups. At least 41 interleukins have now been described, each with a range of actions. IL-1 is produced by almost all nucleated cells, including ocular cells such as the RPE, and is central to the initiation of most inflammatory and immune responses. Many of the more recently described interleukins have homology to IL-1, including IL-18, IL-33, IL-35 and IL-39. Some of these have anti-inflammatory effects, unlike IL-1 and IL-18 and the macrophage-produced IL-41, which are secreted on activation of the inflammasome (Fig. 7.11). IL-1 has extensive homology to fibroblast growth factor and may also be implicated in angiogenesis. IL-2 is the major T-cell growth factor and initiates release of cytokines from the cells upon which it acts (Fig. 7.2).

Cytokines function in a vast interconnected network of agonist/antagonistic feed-forward and feedback inhibitory and stimulatory loops that ensures fine control over the immune response, mostly to avoid the excessive collateral damage that would occur in an over-robust response to a pathogen. Much of this is achieved by through selective signalling pathways.

The amount of information at many levels has grown so much and so fast that bioinformatics and human protein databases are now essential repositories for data provision. For instance, the entire list of interleukins and their signalling pathways has been stored by the Reactome database. As of December 2017, the 63rd release of Reactome data included 10,996 human proteins involved in 11,426 events within 2179 pathways. Data on human biological pathways are freely available through the Reactome Web site (https://reactome.org). The value of this database is that it allows several levels of analysis from the multiple interacting pathways to the final 3D structural models with its molecular interactions of any protein on the database such as the interleukins.

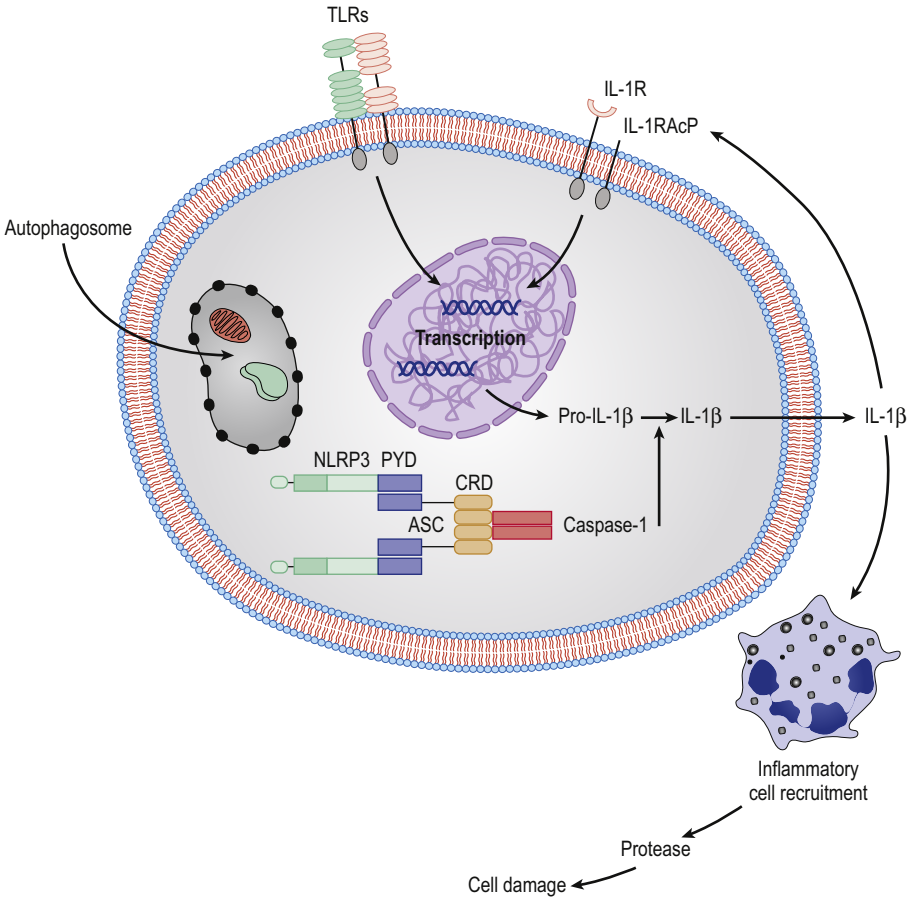


Fig. 7.11 IL-1β is transcribed through a process involving activation of PRRs such as toll receptors or via the IL-1 receptor itself, leading to activation of the inflammasome and induction of caspase-1 to generate IL-1β from its precursor pro-IL-1β. The process of autophagy regulates inflammasome activation and IL-1 production. *TLR*, Toll-like receptor. (From van de Veerdonk and Netea, 2013.)

Cytokines involved in specific immune reactions

The interleukins, interferons and tumour necrosis factors are central to the immune response and the character of the immune response is determined by the set of cytokines released. Activation of a naive T cell (a lymphoblast that has recently been presented with antigen) by different cytokines (in the presence of IL-2) released from the APC drives the T cell to differentiate into one or more of the T-cell subsets (see Fig. 7.2). Th1 cells secrete interferon-γ (IFN-γ). Th1 cells are involved in delayed-type hypersensitivity responses and tissue damage associated with granuloma formation (see Chapter 9, p. 505). This is achieved by activation of macrophages and NK cells, which release reactive oxygen and nitrogen intermediates (free radicals).

$CD4^+$ Th1 cells also activate CD8 cytotoxic T cells (by releasing IL2) as well as inducing Tregs. In contrast, $CD4^+$ Th2 cells release cytokines, which induce B cells to produce antibody and activate eosinophils, the effector cells in allergy-associated tissue damage. Th2-induced IgE avidly binds to mast cells, which cause immediate hypersensitivity. In mucosal tissues such as the gut lining, Th2-generated cytokines promote IgA for defence against mucosal antigens.

This simple outline (see Fig. 7.2), describing the generation of either Th1 or Th2 responses, is considerably modified during the event. For instance, IL-1, IL-2 and IL-4 can also activate macrophages directly, while IFN-γ is involved in the production of $IgG_{2\alpha}$ (the only immunoglobulin controlled by this cytokine). In contrast, activation of a naive T cell in the presence of the cytokine TGF-β leads to the generation of Treg cells, which suppress the immune response,

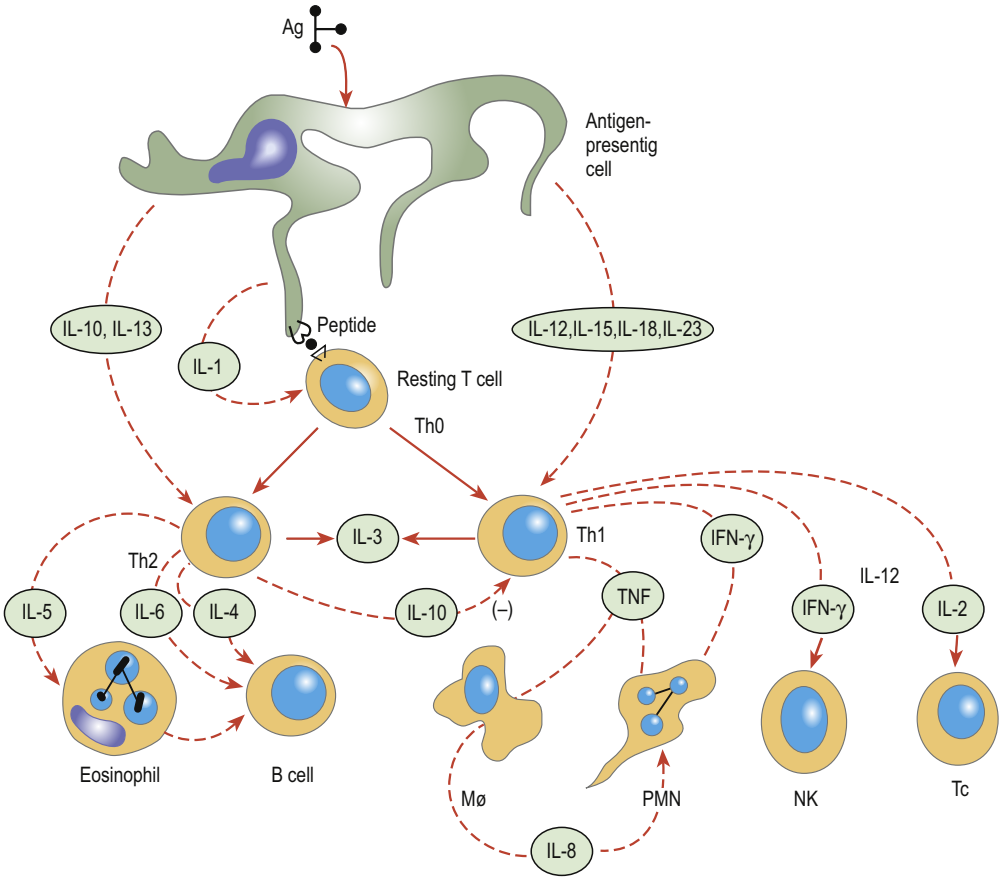


Fig. 7.12 Cytokine networks induced by immune responses. *Ag*, Antigen; *IL-1*, interleukin-1; *IFN*, interferon; *Mϕ*, macrophage; *NK*, natural killer cell; *PMN*, polymorphonuclear leucocyte; *Tc*, T cytotoxic cell; T_h, T helper cell; *TNF*, tumour necrosis factor; ● peptide. (Courtesy A. Abbas.)

unless IL-6 is also present in the milieu, in which case the strongly pro-inflammatory Th17 cell differentiates (see Fig. 7.2) driven by IL-23 production from the APC. IL-6 is produced by many tissue cells, including ocular cells such as the RPE, especially under stress.

Cytokines involved in lymphomyeloid cell maturation

The bone marrow is the powerhouse of our immune defence system. Several cytokines are involved in the growth and maturation of lymphomyeloid cell populations from stem cell precursors in the bone marrow. These include the colony stimulation factors, granulocyte colony-stimulating factor (G-CSF), macrophage colony-stimulating factor (M-CSF) and granulocyte–macrophage colony-stimulating factor (GM-CSF). In addition, several other cytokines with pleiotropic effects, including TNF-α and IL-1, have important roles in the maturation of these cells.

Cytokine receptors and cytokine receptor antagonists

T-cell activation is antigen specific but the cytokines released are not. However, cells targeted by cytokines require the appropriate receptor for that cytokine to respond (Fig. 7.12). In addition, the targeted cells require a cell-signalling mechanism to mediate the response.

Cytokines use a common cell-signalling mechanism involving a cytosolic protein NF-κB, which regulates genes encoding cytokines, their receptors and several other genes involved in the acute inflammatory response. NF-κB is released from its inhibitor (IκB) when the cell is stimulated by cytokine and enters the nucleus bound to the

transcription factors p65/p50. These initiate the changes in gene transcription with changes in function.

Further information on cytokine receptorsis is available at https://expertconsult.inkling.com

Cytokines, when released, do not have a free rein in mediating their activities. Cytokine receptor antagonists are well recognized, the prototype in this field being IL-1Ra. This protein is a naturally occurring competitive binding protein for the IL-1 receptor but fails to induce any of the signal transduction events of IL-1. This is important because the uncontrolled activity of IL-1 in large amounts produces severe side-effects similar to the acute-phase response. IL-1Ra recognizes the separate receptors for IL-1 on T and B cells. IL-1 can also bind to a 'decoy' receptor, IL-1RII, which fails to transmit the signal. Some cytokines play a bigger role in inflammation than others.

Lessons learned from treating patients with autoimmune inflammatory disease point towards the importance of certain selected cytokines as major effectors of inflammation. For instance, by selectively targeting TNF-α, remarkable recoveries in function can be achieved in patients blinded by uveitis or crippled with rheumatoid arthritis. Several other cytokines have primary roles in disease: for instance, IL-1 (and its family members) is involved in virtually every form of inflammation and IL-2 is very important for T-cell function. Effector T cells can make their own IL-2, but Tregs, which are essential for controlling the inflammatory response, are unable to synthesize IL-2 and depend on IL-2 from other sources, including T-effector cells (thus T-effector cells initiate their own regulation).

TGF-β is also a major regulator of T-cell differentiation (see Fig. 7.2) in addition to being much more generally involved in several cell biological processes such as fibrosis and angiogenesis. TGFb probably has a default role as a mediator of immune regulation but requires conversion from a latent to an active state. Interestingly, it is constitutively present in its active form in aqueous humour (see p. 460).

How many cytokines are there?

Cytokines include almost any secreted short range-acting small proteins and are produced by almost every cells. In addition, they have pleiotropic effects, i.e. they act on many cell types. As previously indicated, they include interferons, growth factors and interleukins. They differ from hormones, which are released by specific cells, enter the blood stream and act on specific target organs. A snapshot profile of some cytokines is provided through the following link.

Brief descriptions of some cytokines are available online at https://expertconsult.inkling.com

Chemokines and chemokine receptors

Cytokines that are specific for induction of leucocyte migration to the site of inflammation are termed chemokines and include some interleukins in addition to other cytokines with additional functions. Chemokines are small peptides (usually 8–15 kDa) and are classified in two main subsets based on a particular amino acid sequence involving two cysteine residues, –C–C– chemokines and –C–X–C– chemokines, in which the latter contains an intervening non-cysteine residue. However, many chemokines were discovered before the current nomenclature came into being and in many cases the older terminology is still in use (Table 7.5). Some chemokines do not belong to either category, having either only one C residue or having additional intervening amino acids, but they are unusual.

Regulation of inflammatory cell traffic to sites of inflammation is determined by the set of chemokines released in the tissues and the expression of specific chemokine receptors on different cell types. There is considerable redundancy in the system but, despite this, there is temporal regulation of specific leucocyte recruitment depending on which chemokines and chemokine receptors are active. Chemokines are released by tissue cells at the time of injury or antigen challenge and, because of their low molecular weight, percolate and diffuse through the tissues to line blood vessels and lymphatics, thereby promoting leucocytes to find a pathway to the site of inflammation or, in the case of the lymph node, to the site of T-cell activation. For instance, chemokines such as IL-8 (CXCL8) specifically attract neutrophils through the receptor CXCR1, while CCL19 and 21 attract T cells and dendritic cells using CCR7. In addition, there are several other receptors and chemokine-binding agents that act as decoys by binding the chemokine but failing to signal, thus effectively 'confusing' the cell and preventing effective migration. There are five such decoy receptors including D6 and DARC (Fig. 7.13) now known as atypical chemokine receptors (ACKRs 1-5) and their function appears to be to fine-tune leucocyte migration during the immune response.

Complement

The term 'complement' was coined to describe an activity in sera required for antibody-mediated lysis of bacteria but that was lost after heating to 56°C. Antibody itself was heat-stable and retained the ability to agglutinate the bacteria but could not kill them. This additional activity therefore complemented antibody-mediated cytotoxicity.

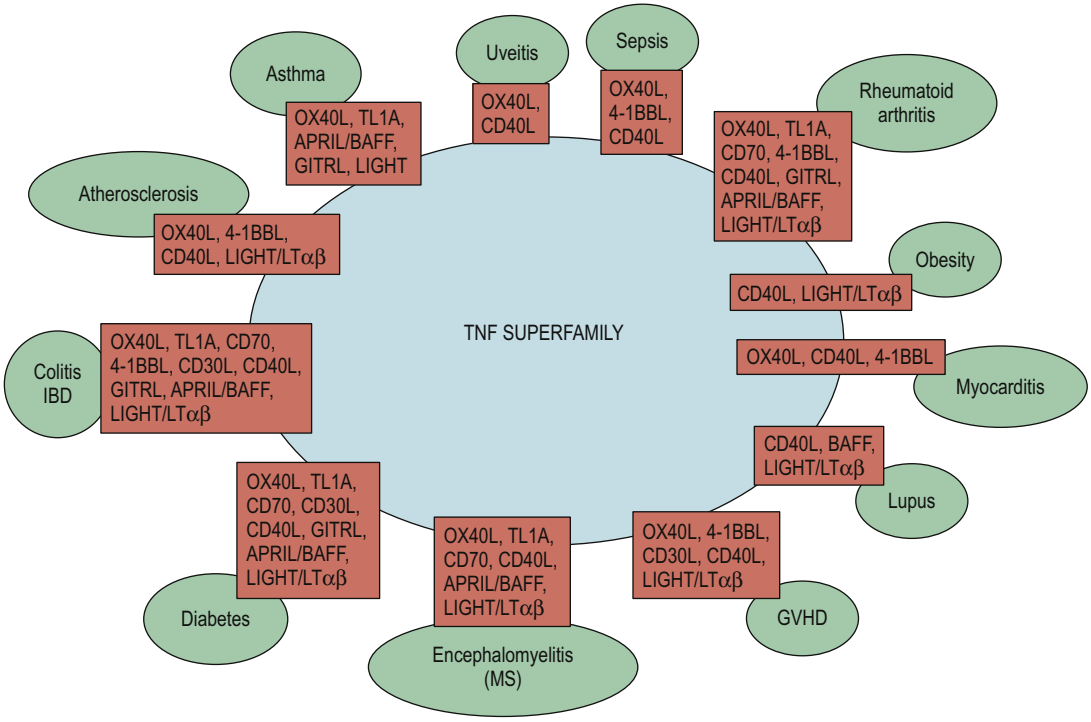
What is complement and what does it do?

Complement is a property of serum derived from sequential zymogen activation of a series of plasma proteins (a

CYTOKINE RECEPTORS

Five families of cytokine receptors are described based on structural motifs in the proteins: (1) the immunoglobulin superfamily (IL-1 and c-kit, IL-1Ra, IL-18, IL-33, IL-35, IL-37, IL-39); (2) a two cysteine/WSXWS or type 1 receptor family (binds IL-2, IL-3, IL-4, IL-5, IL-6, IL-7, IL-9, IL-11, IL-13, IL-15, IL25, IL27, IL31, GM-CSF, G-CSF); (3) a type II receptor (binds IFNs, IL-10, IL-9, IL-20, IL-22, IL-24, IL-26); (4) a TNF receptor, part of a larger family of receptors involved in apoptosis including Fas–Fas ligand (FasL) mechanisms, TRADD (TNF receptor-associated death domain), TRAF (TNF receptor-associated factor) and CD40 (eFig 7.6); and (5) a seven transmembrane helix (chemokine) receptor family (see p. 413). Some cytokines have more than one receptor and some of the structural motifs are shared between the receptors: for instance, the common γ chain of the IL-2 and IL-15 receptors. Binding of cytokine to the receptor initiates signal transduction pathways such as the Janus family kinases (Jaks) and the signal transducer and activator of transcription (STATs) proteins that act upstream of NF-κB (eFig. 7.7). Several proteins are included in the Jak–STAT families and these intracellular proteins are prominent targets for drug discovery programmes in attempts to control immune responses and Jak inhibitors are now being used in rheumatoid disease and being trialled in uveitis.

Receptors may be of low or high affinity in their ability to bind ligands and this encroaches somewhat on receptor specificity, allowing certain cytokines to compete for the same receptor. In addition, for certain multichain receptors, such as the IL-2R, one of the chains may be shared by several cytokines (the common γ chain) and therefore assembly of the appropriate receptor on the membrane may depend on the precise cytokine milieu presented to the cell. The local concentration of any particular cytokine will therefore have an influence on the final cell response. This cytokine redundancy is seen for instance with GM-CSF, IL-3 and IL-5. GM-CSF and IL-3 have widespread effects, whereas IL-5 is more restricted. By competing for the same receptor on different cell types, the amount and receptor-affinity of any one cytokine can determine the nature of the cellular response.



eFig. 7.6 A diagram of the tumour necrosis factor receptor superfamily with identification of the autoimmune diseases associated with specific molecules. Some, such as OX40/OX40L, are associated generally with organ specific diseases (such as myocarditis and uveitis) but not with systemic autoimmune diseases such as lupus. (From Croft et al., 2012)

BRIEF REVIEW OF SELECTED CYTOKINES AND THEIR RECEPTORS

The number of cytokines being discovered continues to increase and many of these play important roles in inflammation and its regulation. The following sections provide brief information on specific cytokines.

Tumour necrosis factor

Certain cytokines have a prominent role in tissue damage, one of which is TNF. TNF exists in both soluble and membrane-bound forms and can induce a variety of responses in cells, including activation of polymorphonuclear leucocytes, induction of MHC antigens and adhesion molecule expression, and prostaglandin synthesis. Two forms of TNF exist, TNF-α and TNF-β, the latter also known as lymphotoxin; both forms bind to the same cell receptor(s).

TNF-α is pro-inflammatory, pro-coagulant, cytotoxic and antiviral and modulates haematopoiesis. Its inflammatory effects act on both the acute-phase response and locally on cells at sites of inflammation. As such it is a mediator of endotoxic shock. TNF-α signals through two receptors (p55 and p75), which are members of a larger TNF receptor family involved in several aspects of the immune response. Many are associated with cell death such as Fas-FasL (apoptosis), also important in immune privilege. Each receptor forms a (p55 or p75) homotrimer before ligand binding. TNF-α is produced as a pro-protein with a long signalling sequence. When cleaved off, the molecule trimerizes and binds to the trimeric receptor. TNF receptors can be cleaved from the cell surface and persist in the extracellular space and the bloodstream, where they can act as competitive inhibitors of TNF itself. Humanized fusion proteins using TNF have been used in this way to prevent disease.

TNF is a major mediator of experimental and clinical uveitis and thus blockade of TNF-α ameliorates disease and preserves sight.

Tumour necrosis factor receptor superfamily

The tumour necrosis factor receptor (TNFR) superfamily (there are about 50 soluble and membrane-bound members) is a very important class of signal transduction molecules in the immune system. Included in this family are the TNF receptors 1 and 2 (TNFR1 and 2), the low-affinity nerve growth factor receptor (NGFR), CD40, CD30, Fas, and others indicated in the table below. Members of the TNF receptor family are generally transmembrane proteins, but many of them can also be secreted as soluble molecules, derived either by proteolytic cleavage from the membrane or by differential mRNA processing. The TNFR family is characterized by the presence of cysteine-rich motifs of 40 amino acids in the extracellular domain that are involved in ligand binding. The ligands for these receptors are type II transmembrane proteins with a 'jelly-roll' β-sandwich structure, many of which can also be secreted.

Some TNF receptor family members are associated with additional signal transduction molecules called TNFR-associated factors (TRAFs). These proteins contain zinc-binding domains that are thought to mediate the binding of the protein to DNA, leading to transcriptional activation. Signal transduction is initiated by stimulation of oligomerized receptor complexes upon ligand binding. The ligands, such as TNF, are often found in multimeric form and this multivalency enhances the induction of signalling. Other members of the TNFR family (most notably, the apoptosis-inducing molecule Fas) contain death domains (DDs) in their cytoplasmic tails. For example, the DD sequence allows the interaction of the membrane-bound Fas molecule with the signal-transducing protein FADD (Fas-associated death domain), which in turn delivers signals causing the cell to initiate apoptosis. It has been discovered that several proteins containing the DD sequence are not involved in apoptosis at all but actually promote cell survival. In fact, depending on downstream events, engagement of TNFR1 (which contains a DD sequence) can lead either to cell death or to cell survival.

The functional consequences of ligand engagement by members of the TNF receptor superfamily are very diverse. Fas binding induces apoptosis, which is important in the maintenance of immune self-tolerance. CD27, CD30, 4-1BB and CD40 signalling enhance the survival, proliferation and activation of B or T lymphocytes, often playing critical roles as costimulators or signal modulators. Signalling by other TNFR family members results in the activation of NF-κB in macrophages and the induction of an inflammatory response. The TNF receptor superfamily is thus central to many aspects of both innate and acquired immunity and plays fundamental roles in both cell death and survival. Different TNFR members are involved in different diseases (see eFig. 7.6). A greater understanding of this receptor family, as well as its corresponding family of ligands, holds great potential for the creation of novel drugs and therapeutics that could be used to control many aspects of autoimmunity and other immunopathologies.

TRANCE

TRANCE (TNF-related activation-induced cytokine, also known as osteoprotegrin ligand, or OPGL, and receptor activator for NF-κB ligand, or RANKL) was defined initially as an activator of osteoclasts in bone for bone turnover and of innate immune cells, macrophages and dendritic cells for the immune response. Deficiency of TRANCE leads to the severe bone disease osteopetrosis. It is now also recognized as a major ligand for the promotion of immunological tolerance on medullary epithelial cells (mTECs) in the thymus where it binds the receptor RANK and sustains tolerance to self-antigens during adult life once it is established in the fetus.

Interleukin-1

Interleukins constitute a subclass of cytokines: their name is derived from their ability to effect communication between leucocytes. However, they are derived from a range of cells

BRIEF REVIEW OF SELECTED CYTOKINES AND THEIR RECEPTORS—CONT'D

and have effects on many cell types, including leucocytes. IL-1 is released on cell death by many cells as part of the injury response and then amplifies this response by signalling via the inflammasome to release more IL-1 and IL-18. In macrophages and endothelial cells (its selective targets) it induces adhesion molecule expression and also promotes prostaglandin synthesis. IL-1 also activates bone cells, accelerates bone turnover and induces marrow stromal cells to produce G-CSF and IL-3. Two forms of IL-1 exist, IL-1α and IL-1β, each with its own receptor, but receptor usage is not highly restricted. Thus, IL-1α and IL-1β have broadly similar effects on cells. IL-1 is extremely potent and is counteracted by the IL-1 receptor antagonist, which plays a major role in regulating IL-1 activity.

IL-1 activates cells within the eye, particularly the RPE and the retinal vascular endothelium. It has been implicated in the pathogenesis of various forms of uveitis and has been shown to be uveitogenic experimentally and to have significantly greater damaging effects than TNF-α.

IL-1α and IL-1β are produced as pro-proteins without a secretory signal sequence and must be digested by caspase-1 (IL-1β) or other proteases (IL-1α) before they are activated. IL-1 (and IL-18) signal through Myd88 and the inflammasome.

Interleukin-2

IL-2 is the major cytokine involved in T-cell-mediated responses, both Th1 and Th2, and it induces NK cell activity; in both cases it can act in an autocrine manner. IL-2 also initiates activation of CD8 T cells. Activated T cells thus express the IL-2 receptor (CD25), which has, for instance, been detected on circulating T cells as well as lymphocytes from intraocular samples from patients with endogenous uveitis. Interestingly, Treg cells express the highest level of the IL-2 receptor and IL-2 is essential for Treg induction even though Tregs cannot produce this cytokine. Tregs express and require the transcription factor FoxP3 and, conventionally, Tregs are described as $CD4^+CD25^+FoxP3^+$ T cells.

The interferons

IFN-α and IFN-β are members of the type I interferons, of which there are many, while IFN-γ is the only type II interferon, a classification based on structural differences. IFN-α was the first cytokine to be identified, sequenced, cloned and introduced to clinical therapeutics. It is used in the treatment of multiple sclerosis, hepatitis C and certain forms of retinal vasculitis. It is produced by many cells in response to virus infection and in this regard is important in promoting viral clearance. The main source of constitutive IFN-α is a rare population of dendritic cells, the plasmacytoid dendritic cells (pDCs). Large quantities of this cytokine are produced by pDCs in response to virus infection, leading to maturation of myeloid dendritic cells with strong induction of antiviral cytolytic activity and B-cell isotype switching. IFN-α can also be produced by activation of DCs via the TNF receptor but with a gradual kinetic as opposed to the massive antiviral response. INF-α produced by TNF receptor stimulation may be immunoregulatory. IFN-β is produced mainly by fibroblasts.

Two interleukins (IL-28 and IL-29) have been described that have IFN-αβ-like activity and can induce antiviral activity in the absence of IFN. These cytokines have now been rebranded as IFN-λ, i.e. type III interferons, and like the type 1 interferons they act via the Jak-STAT pathway (eFig. 7.7).

IFN-γ is released from virally infected cells and has potent antiviral activity. It is also a major pro-inflammatory cytokine and induces MHC and other antigens on cell surfaces, thus promoting adaptive immunity via antigen presentation. It is the signature cytokine for $CD4^+$ Th1 cells. It has been shown to induce MHC class II on cells in the retina (which does not normally have many MHC class II^+ cells); these may be tissue-specific APCs. It also induces the innate immune response by activating macrophages to produce cytokines such as TNF-α, IL-1β, IL-12, IL-18 and IL-23 and to engage in cytotoxicity through release of reactive oxygen species and nitric oxide. Like TNF-α and IL-1, IFN-γ has potent uveitogenic activity. In addition, high levels of IFN-γ have been detected in aqueous samples from patients with acquired immune deficiency syndrome (AIDS) retinitis.

Production of interferons is under the control of interferon-regulating factors (IRFs), which bind to interferon-stimulated regulatory elements (ISRE) on the promoters of the IFN genes.

Transforming growth factor-β

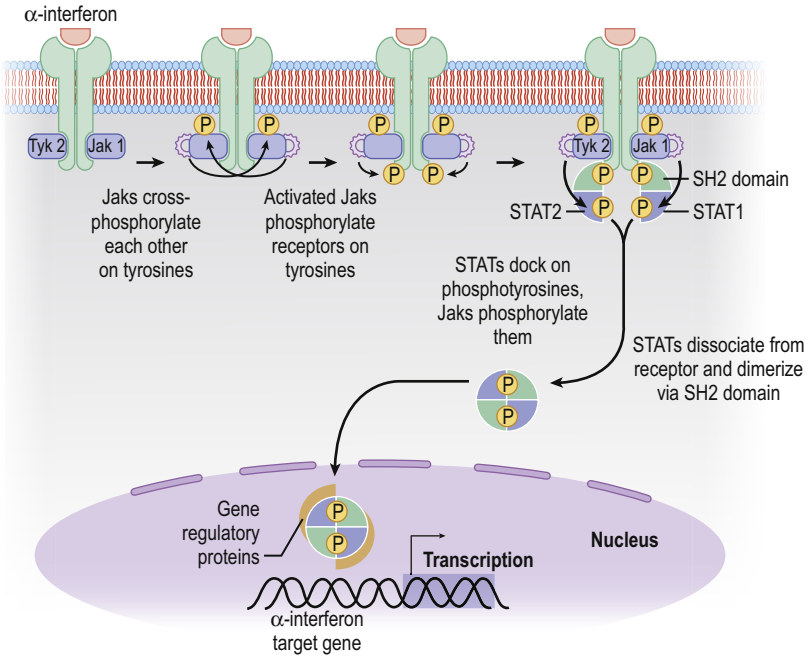
TGF-β is an important immunosuppressive cytokine. There are at least six types of TGF-β and many cell types elaborate this mediator, including cells within the eye such as ciliary body epithelium and RPE. TGF-β has been suggested to account for part of the immunosuppressive activity normally found within the eye (see p. 460). TGF-β2 appears to be the main isotype found within the eye and is normally secreted in a latent form; however, it is readily converted to the active form by enzymes such as plasmin, which would normally be present in an inflammatory exudate.

TGF-β promotes T-cell differentiation as well as angiogenesis and chemotaxis of several leucocyte types.

TGF-βs have widespread effects on cell adhesion, differentiation, proliferation, migration, maturation, activation and regulation, both within and on cells outside the immune system. *In vitro* studies are greatly affected by small changes in the TGF concentrations and many effects therefore require *in vivo* investigation. TGF-βs are essential, however,

Continued

BRIEF REVIEW OF SELECTED CYTOKINES AND THEIR RECEPTORS—CONT'D



eFig. 7.7 The JAK-STAT pathway is activated by many cytokines, in this example by the cytokine interferon-α.*STAT*, Signal transducer and activator of transcription. (From http://www.motifolio.com/5111157.html.)

because deletion of the TGF-β gene in mice is lethal for the embryo. Overall, TGF-β1 and TGF-β2 are anti-inflammatory and immunoregulatory, but TGF-β1 is pro-fibrotic. TGF-βs are produced by leucocytes and by parenchymal cells in the central nervous system, kidney and eye. TGF-β receptors comprise three chains, of which chains I and II bind to form a high-affinity receptor while III binds either of the other two chains in a regulatory, non-signalling role.

Interleukin-4

Interleukin-4, -5, and -13 are prototypical cytokines inducting Th2-type responses and are thus major players in allergic diseases including asthma and allergic skin and mucous membrane disorders such as allergic conjunctivitis. In addition, IL-4 and IL-13 are important inducers of alternative macrophage (M2) activation promoting wound healing, fibrosis and angiogenesis.

Interleukin-6

IL-6 has pro-inflammatory and fever-inducing activity, but recent evidence suggests that its major role may be to limit tissue damage. It has multiple and wide-ranging effects and participates in the acute-phase reaction (fever) as well as in haematopoiesis. IL-6 is produced by many cell types, both immune and non-immune, including ocular cells such as RPE cells. IL-6 may be instrumental in promoting a Th2-type response with preferential activation of B cells. However, experimentally, IL-6 appears to have a marked uveitogenic effect, similar in severity to that of endotoxin. In addition, IL-6 has been detected in the aqueous of patients with endogenous and postsurgical uveitis. IL-6 combines with TGF-β to promote induction of Th17 effector cells, Th17.

Like many cytokines, IL6 and the IL6 receptor (IL6R) can exist in soluble form and engage in 'trans-signalling'. IL-6, IL-11, IL-30 and IL-31 are grouped together because of similarities in their receptor and signalling pathways. In addition, the immunosuppressive cytokine IL-35 combines the components of IL-6 and IL-12.

Interleukin-8

IL-8 is regarded as a chemokine for neutrophils but also has some level of chemotactic properties for monocytes and lymphocytes. IL-8 is released by immune and non-immune cells but is somewhat more restricted than IL-6. RPE cells, corneal endothelium and stromal cells release IL-8 after appropriate stimulation. IL-8 is less uveitogenic than IL-1, which appears to be the main cytokine in this respect. IL-8 binds to the chemokine receptors CXCR1 and CXCR2.

BRIEF REVIEW OF SELECTED CYTOKINES AND THEIR RECEPTORS— A AND B CONT'D

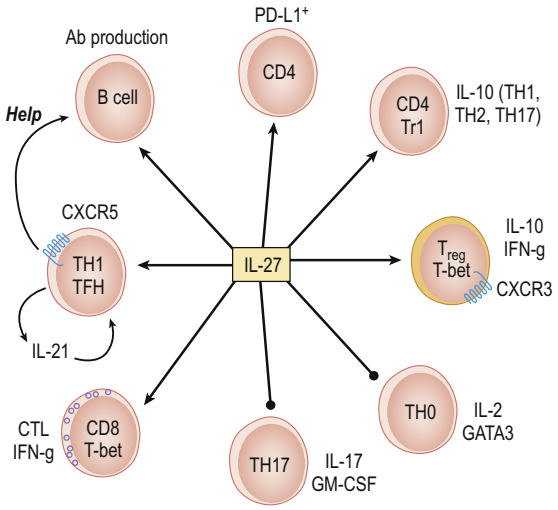
Interleukin-10

IL-10 is an 18 kDa cytokine produced by many cells. It is the prototype cytokine of the IL-10 family of cytokines, which consists of nine members: IL-10, IL-19, IL-20, IL-22, IL-24 and IL-26, which are also related to IL-28 and IL-29. The IL-10 family is predominantly an 'immunosuppressive' cytokine family and is involved for instance in immunological privilege of the eye and other organs. The effects of IL-10 can override those of many pro-inflammatory cytokines. IL-10 does this by inhibiting synthesis or secretion of TNF-α, IL-1, chemokines and IL-12 by macrophages. It also reduces MHC class II expression on APCs, thus downregulating specific and innate immune responses. Certain viruses such as Epstein–Barr virus release an alternative form of IL-10, raising the possibility of virus-induced immune suppression. The heterodimer IL-10 receptor is expressed mainly on haematopoietic cells.

While IL-10 is mostly produced by all leucocytes, IL-19, IL-20 and IL-24 are produced by myeloid and epithelial cells and are involved in antibacterial responses, wound healing and tissue remodelling. In contrast, IL-22 and IL-26 have similar functions but are produced by T cells, NK and NKT cells, as well as innate immune cells such as those in the gut. IL-28 and IL-29 are predominantly antiviral.

Interleukins 12, 17, 23 and 27

IL-12 rose to prominence as the major mediator and inducer of IFN-γ-producing CD4+ Th1 responses during the time that IL-2 was recognized as having a regulatory (inducer of Tregs) as well as a T-cell-activating role. IL-12 is produced by APCs, particularly mature dendritic cells and activated macrophages, and is strongly pro-inflammatory both for Th1 cells and NK cells. IL-12 synergizes with IL-18 to induce IFN-γ production by activated T cells and NK cells. IL-12 is a heterodimer with p35 and p40 chains. It shares the p40 chain with IL-23, in which the p40 chain combines with a separate p19 chain to form the IL-23 heterodimer. IL-23 is the main pro-inflammatory cytokine, while the p35 chain of IL-12 may have some regulatory activity. IL-23 is responsible for induction of IL-17-producing CD4+ Th17 cells, which are major players in the pathogenesis of autoimmune disease as well as adaptive immunity to many infectious agents, particularly fungi. Both IL-1 and IL-17 play major roles in recruiting neutrophils to sites of infection, e.g. *Staphylococcus aureus* skin infections. They may also be involved in tumour immunity. In contrast, IL-27, with p35-like and a p40-like chains, induces the IL-12Rβ chain on Th1 cells during the initiation of Th1 responses, but its main function is to put a brake on inflammatory responses (eFig. 7.8). IL-12, IL-18, IL-23 and IL-27 induce both primary and memory adaptive T-cell responses in a coordinated series of reciprocal interactions between the dendritic cells and T cells.



eFig. 7.8 IL-27 is produced by innate immune cells and modulates the behaviour of adaptive immune cells.

Interleukins 15 and 21

IL-2, IL-4, IL-7, IL-9, IL-15 and IL-21 share the common γ-chain of the IL-2 receptor. NK cells are absolutely dependent on IL-15. Both IL-15 and IL-12 synergize to activate NK cells. IL-15 also activates γδ T cells. IL-15 has similar functions to IL-2 on T cells. However, there are striking differences from IL-2; for instance, IL-15 is not produced by T cells but by bone marrow stromal cells. The IL-15 receptor has a unique α chain but uses the IL-2R β and γδ chains. IL-21 also activates NK cells and induces T-cell activation but, unlike IL-2 and IL-15, it inhibits B-cell proliferation.

Interleukin-18

Both IL-1 and IL-18 are released from pro-peptides in APCs on activation of the inflammasome and caspase-1. IL-18 has similar pro-inflammatory functions to IL-1 and IL-12 and induces production of IFN-γ and TNF-α in macrophages. Interestingly, IL-18-deficient mice show marked susceptibility to bacterial infections but have no impairment of response to challenge with ocular antigens. In contrast, IL-18 may have an anti-angiogenic role in diseases such as age-related macular degeneration.

Interleukin-32

IL-32 is a pro-inflammatory and pro-apoptotic cytokine involved in many autoimmune inflammatory conditions and in viral and mycobacterial infections and cancer in humans. It has not yet been found in rodents and there is no clear

Continued

BRIEF REVIEW OF SELECTED CYTOKINES AND THEIR RECEPTORS—CONT'D

evidence of a specific receptor, although it does appear to signal via integrins.

Interleukins 33 and 36

IL-33 and IL-36 are members of the IL-1 family and, like IL-32, have both intracellular and extracellular effects. IL-33 is released on cell form death from nucleus disintegration and mediates both Th2 and anti-Th1/Th17 immunosuppressive effects. IL-36 is associated with a psoriasis-like disease.

Interleukin-34

IL-34 binds to the CSF-1R (M-CSFR) and appears to be necessary for Langerhans' cell and microglia development, differentiation and activation, and may be important in the pathogenesis of neurodegenerative disorders.

Interleukin-37

IL-37 is the newest member of the IL-1 family and is an immunosuppressive cytokine. Like IL-1a and IL-33, IL-37 binds to the nucleus in a receptor-independent manner to exert its anti-inflammatory effects.

Interleukin-38

IL-38 binds to the IL-36 receptor (also a member of the IL-1 family) and inhibits IL-17 and IL-22 effects.

Interleukin-40

IL40 is produced by activated B cells and is particularly involved in the production of IgA in secretory fluids such as milk and tears.

Table 7.5 **Selected nomenclature for chemokines and their receptors**

Chemokine	Old name	Receptor
CXC family	**α family**	
CXCL1	Groα	CXCR2
CXCL2	Groβ	CXCR2
CXCL5	ENA-78	CXCR2
CXCL8	IL-8	CXCR1, CXCR2
CXCL9	HuMIG	CXCR3
CXCL10	IP-10	CXCR3
CXCL11	ITAC	CXCR3
CXCL12	SDF-1	CXCR4
CXCL16	SR-PSOX	CXCR6
CX3CL1	Fractalkine	CX3CR1
CC family	**β family**	
CCL2	MCP-1	CCR2
CCL3	MIP-1α	CCR1, CCR5
CCL4	MIP-1β	CCR5
CCL5	RANTES	CCR1, CCR3, CCR5
CCL7	MCP-3	CCR1, CCR2
CCL19	ELC	CCR7
CCL21	SLC	CCR7
CCL28	MEC	CCR10

(From Lalor et al., 2007.)

zymogen is an enzyme that is activated by a second enzyme, which itself has been activated by proteolytic cleavage, i.e. an enzyme cascade as occurs in kinin formation, coagulation and visual transduction).

Complement is a major component of innate immunity, i.e. the initial response to attack by pathogens, and in this sense can be regarded as a soluble PRR as opposed to the membrane-bound PRRs such as the TLRs (see p. 391). Two enzyme cascades exist, one initiated by antibody combining with antigen, the classical pathway, and a second initiated directly by bacterial surface components, the alternative pathway (Box 7.6). Thus, even at this level, a distinction between innate and acquired immunity exists. The classical complement pathway is in fact the major effector mechanism for humoral immunity (see earlier).

A third mechanism for complement activation has been described. This involves members of a family of molecules called lectins (lectins are non-antibody, non-enzyme carbohydrate-binding proteins, as in the selectins adhesion molecules, see page 398). Lectins that contain collagen-like domains are known as collectins; these molecules are of considerable importance in innate immune mechanisms against microorganisms. Collectins bind to the same receptor as the C1q component of complement (see later) and are thus able to activate the classical pathway. A major lectin in this pathway is mannan-binding lectin (MBL), which is linked to serine proteases MASP-1 and MASP-2. These enzymes are responsible for cleaving C4 and C2, respectively.

Complement has the following effects:

- It is involved in the initiation of the acute inflammatory response by release of certain peptides that act as chemotactic factors and induce vasodilatation with increased permeability (anaphylatoxin).
- It mediates antibody-dependent cytolysis by polymerizing on cell surfaces to form pores in the cell membrane.
- It solubilizes and removes immune complexes from the circulation.
- It induces phagocytosis by acting as an opsonin.

Complement proteins in the normal circulation are inactive and are maintained in this state by an elaborate system of inhibitors that not only inhibit activation of the various enzyme systems but also limit the response once activated.

The central axis of the complement pathway is the conversion of C3, activated by C3 convertases, to C5 convertase by the binding of C3b (Box 7.8). This leads to the sequential addition of a series of complement proteins that result in the membrane attack complex (MAC). Complement proteins are synthesized in the liver and in mononuclear phagocytes. In addition, there is some local production

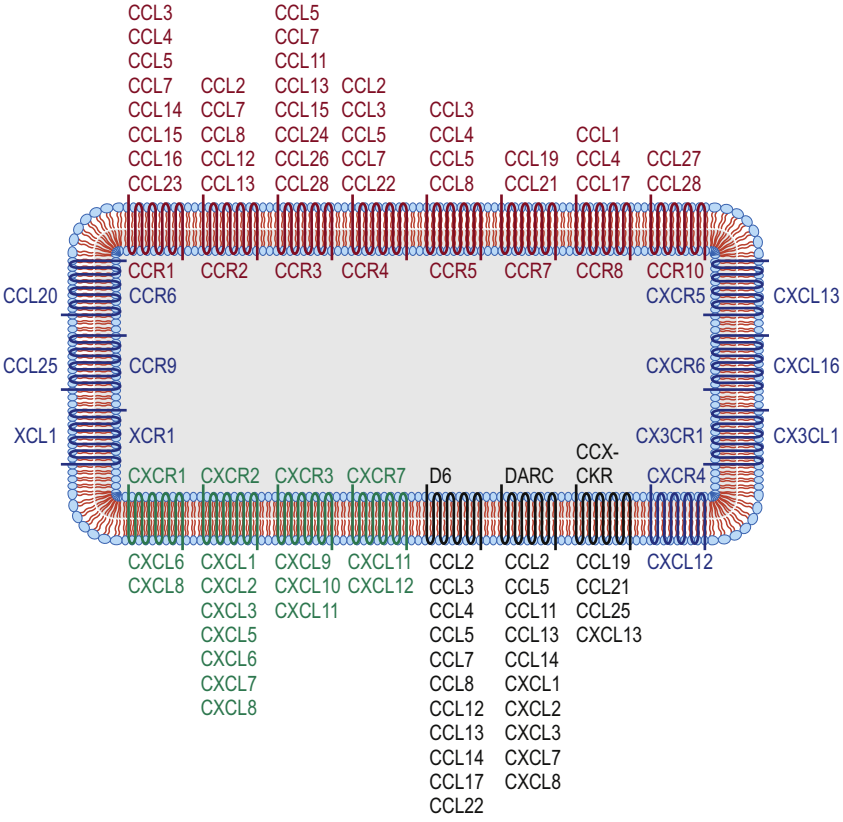


Fig. 7.13 Chemokines and chemokine receptors bind through multiple overlapping, and probably redundant, ligand–receptor interactions. (From Lazennec and Richmond, 2010.)

by parenchymal tissue cells, as occurs in the retina, for instance by retinal pigment epithelial cells.

The classical complement cascade

Activation of C1 is induced by binding to IgM or IgG, but only if the immunoglobulin has bound antigen in the form of an immune complex. Free immunoglobulin does not activate complement. Binding occurs to the C_H3 domain of IgM or the C_H2 domain of IgG, and requires at least two immunoglobulin molecules. Thus, a single molecule of IgM, which is a pentamer, is able to 'fix' complement, while several molecules of IgG, usually aggregated together, are required to achieve the same. Alternatively, cell surface-bound IgG can suffice. IgM is therefore known as complement-fixing antibody and it is this antibody that plays the major role in the initial response to pathogens as part of innate immunity.

C1 can be activated by antibody-independent mechanisms including contact with retroviruses, mycoplasma, or even polyelectrolytes such as DNA and heparin. These presumably act in a non-specific manner by virtue of their charge. Importantly, the acute-phase proteins, C-reactive protein (CRP) and serum amyloid protein (SAP), can also bind complement non-specifically. C1 is composed of three molecules: C1q, C1r and C1s. C1q is a collagen-like molecule with a triple-helix conformation, while C1r and C1s are serine proteases. The molecular complex comprises a tetramer of C1s and C1r with six or more C1q molecules (Fig. 7.14).

When two or more globular heads of C1q bind IgM or cell-surface bound IgG, C1r is cleaved to $C1r^{-}$, which then cleaves C1s to $C1s^{-}$; this activates C4 to C2 to undergo partial proteolysis and bind to form $C4b2a^{-}$. C4b is unstable because it contains an internal thioester bond and is rapidly inactivated by binding a water molecule to form iC4b. If, however, C4b is formed in close contact with a cell membrane, it can covalently bind to the surface and remain in an active state. In contrast, C2 is a single-chain molecule that binds to surface-bound C4a in the presence of Mg^{2+} ions. C2b, produced by $C1s^{-}$ -induced partial proteolysis of C2, diffuses away while the C2a binds to C4b.

The complex of $C4b2a^{-}$ contains the C3 convertase activity. The C4b component binds to the C3 molecule and brings the C2a moiety into close contact with C3, which it

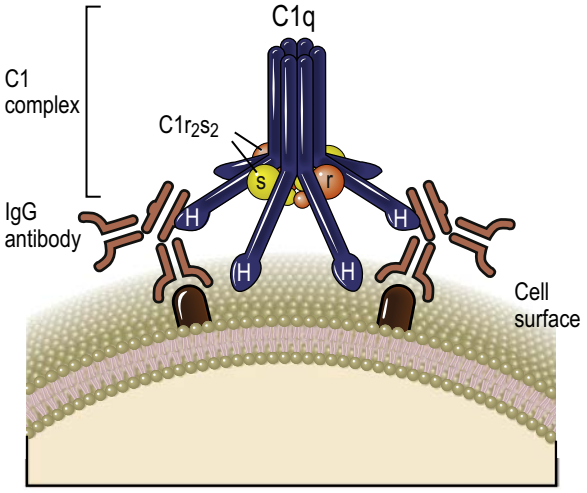


Fig. 7.14 Diagrammatic representation of the generation of the C1 complex comprising a tetramer of C1s and C1r with six molecules of C1q. (From Abbas, A.K. *Cellular and Molecular Immunology*. 7th ed. Elsevier, 2012, Figs 12-10.)

cleaves to C3a and C3b. C3 is a 195 kDa αβ heterodimer with an internal thioester bond similar to C4.

When partially cleaved, this molecule is also unstable and is rapidly inactivated to iC3b. However, if in contact with a cell membrane, it binds covalently in conjunction with C4b2a to form C4b2a3b^{-}, i.e. C5 convertase (see later).

C3a, C4a and C5a are small cationic peptides also known as anaphylatoxins that bind to specific receptors on basophils and mast cells. C3 cleavage products also have a role in antibody production by interacting with follicular dendritic cells (FDC) in the germinal centre (see p. 421).

The alternative pathway

The alternative pathway is triggered by low levels of C3b, which is spontaneously produced *in vivo* by proteolysis, and by C3(H_2O), which is also formed by spontaneous hydrolysis of C3. On normal cells C3b is rapidly deactivated to iC3b by innate regulatory mechanisms (see later), but on foreign surfaces C3b can remain active. C3b binds to factor B on the surface of the cells and is converted by factor D, a serine protease, to C3bBb. This contains the C3 convertase activity and requires a further protein, properdin, to protect it from proteolysis.

Deposition of C3b on foreign surfaces such as bacteria leads to further production of C3b, i.e. a positive feedback amplification loop occurs, which helps to eradicate foreign particles rapidly. C3bBb is combined to form C3bBb3b, which represents the alternative form of C5 convertase (see Box 7.8).

Collectin activation of complement

Collectins include several well-characterized proteins, such as conglutinin and MBL, which directly bind to the C1q receptor on cell surfaces and initiate such phenomena as C4-mediated red cell lysis. MBL directly activates the $C1r_2C1s_2$ tetramer in the absence of C1q. This may be mediated in association with a serine protease (MBP-associated serine protease, or MASP; see above). Collectins also have direct opsonin activity (see later section on macular degeneration).

Intracellular complement activation.

Resting T cells can also lead to assembly of C3 convertase induced by cathepsin L, which cleaves C3 to C3a and C3b. Cathepsin L is located to the ER of T cells and to ER derived secretory vesicles. An increase in this 'tonic' C3 convertase activity is noted when the cells are activated. It is suggested that the constitutive low level of intracellular C′ activation maintains cellular homeostasis and might apply to several cell types including myeloid and non-myeloid cells.

Cytolysis and the membrane attack complex

The MAC is formed by a set of complement proteins inserting themselves into the lipid bilayer and is possible because certain proteins within the complex have a lipophilic core. C5 binds loosely to C5 convertase on the cell membrane and is split into C5a, which diffuses away, and C5b, which complexes with C6 and then C7. The C5b,6,7 complex is highly lipophilic and burrows into the cell membrane. There it acts as a receptor for C8, an αβγ trimer whose γ chain is also lipophilic and similarly inserts into the bilayer. The C5b,6,7,8 complex is weakly cytolytic but becomes considerably more so when it binds C9, a serum protein that polymerizes to form the MAC with 12–15 C9 molecules per C5–9 complex. This forms a pore in the cell membrane, similar to the perforin pore of cytotoxic T cells and NK cells. The pore renders the cell permeable to small ions but not to proteins and is therefore thought to cause cell death by osmotic effects. It is also possible that the large influx of Ca^{2+} ions poisons the cell. The structure of the MAC is remarkably flexible as seen by cryo-electron microscopy (Fig. 7.15). MAC killing of bacteria is less effective than mammalian cells and requires the C5 convertase to remain surface bound as part of the MAC, i.e. the MAC needs to be assembled locally on the surface of the bacterium.

The effects of the MAC on the cell are dose dependent. Sublethal doses of MAC may activate the cell and induce a protective response against further attack and even cell proliferation. In addition, the MAC can lead to activation of

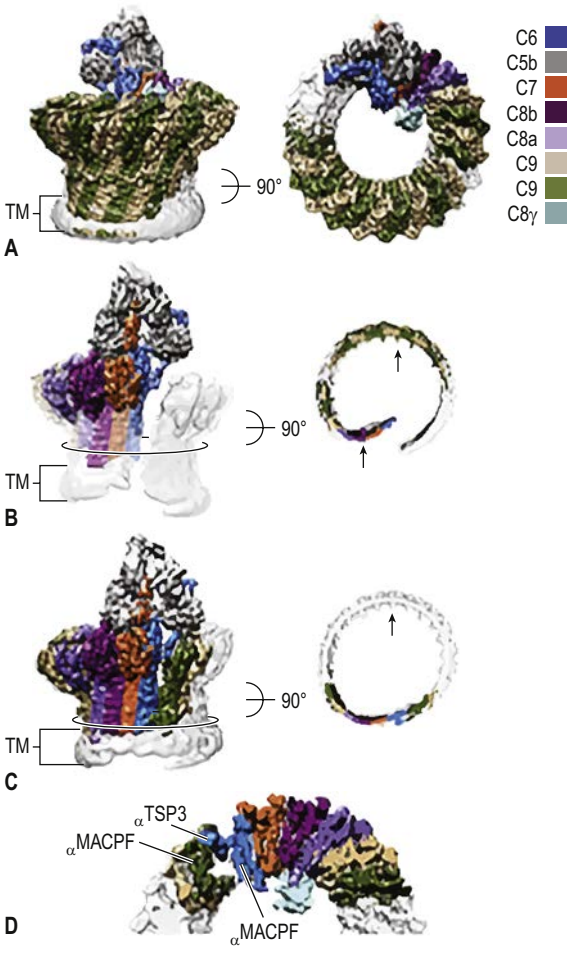


Fig. 7.15 CryoEM structures of the complement membrane attack complex *(MAC)* reveal conformational flexibility of the pore. (**A**) CryoEM density map for the open conformation *(grey transparent surface)* overlaid with subvolume reconstructions of the asymmetric, hinge and C9 oligomer regions *(coloured according to protein components)*. The transmembrane region is indicated *(TM)*. (**B**) CryoEM reconstructions shown rotated 180 degrees. (**C**) CryoEM map of the closed conformation *(grey transparent surface)* overlaid with the subvolume reconstruction of the asymmetric region *(coloured density)* in the same orientation as **B**. *Rectangle* indicates the cross-section shown in the *right panel*. (**D**) Closed conformation of the MAC slabbed through the core MACPF domains. (From Menny et al., 2018.)

NLRP3 and the inflammasome resulting in IL-1 and IL-18 secretion and, if sufficiently extensive as in sepsis, massive epithelial and endothelial damage with multi-organ failure. Genes activated in this manner are known as RGCs (response genes to complement) and several have been described.

Regulation of complement activation

Complement activation is an extremely powerful cytolytic mechanism that can be rapidly activated as a first line of defence against invasion by foreign organisms. It is also extremely effective in memory B-cell responses and antibody-dependent cytotoxicity. It can also swiftly remove potentially toxic immune complexes from the circulation. However, it is potentially extremely hazardous if randomly and uncontrollably activated and there are therefore several inhibitory mechanisms in place to regulate this system (Boxes 7.6 and 7.7). Several new complement inhibitors including the extracellular matrix protein decorin have been described (see Box 7.7).

Activation of complement can be induced by many cell types via receptors that exist on their cell surfaces. Furthermore, some of the biological effects of complement are produced by the cleavage products of complement activation. For instance, C3a and C5a are potent chemoattractants and anaphylatoxins and mediate early-phase responses in acute inflammation (see p. 393). Anaphylatoxins act directly via complement receptors on granulocytes, macrophages and mast cells to induce degranulation and the release of vasoactive mediators (Table 7.6).

Enzyme systems such as complement activation and its regulation do not function in isolation. For instance, the C1 inhibitor (see Box 7.10) regulates contact activation of coagulation as well as generation of plasmin in the fibrinolytic system. Thus, C1q joins other enzyme inhibitors such as α2 macroglobulin and α1anti-trypsin in general regulatory mechanisms.

Complement activation is usually incomplete: implications for age-related macular degeneration

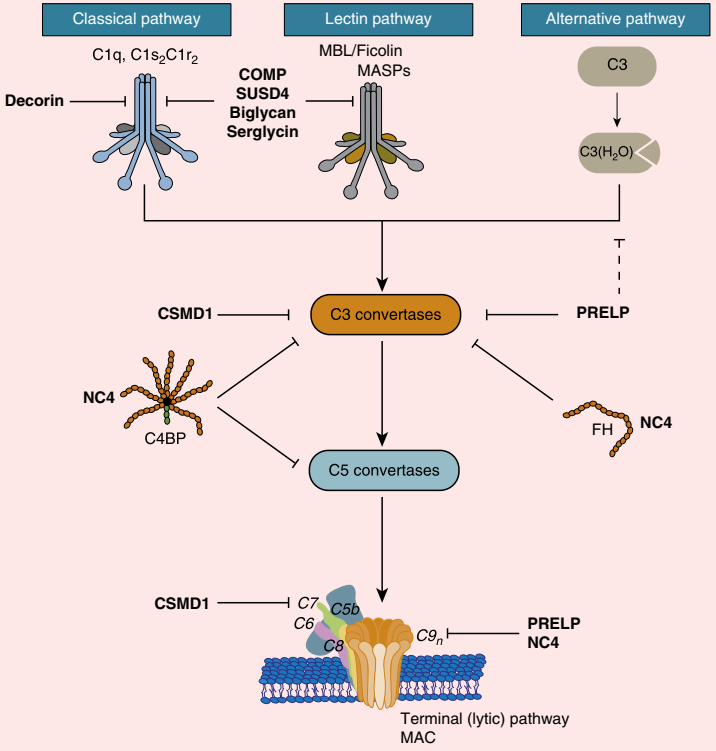
As shown in Box 7.6, the multistep complement activation cascade requires the correct conditions for activation of each step and, because of instability in many of the molecular intermediates, it frequently fails to proceed to formation of the full MAC, particularly if activated via the collectins. In addition, the process can be blocked by a range of natural and synthetic inhibitors (see Boxes 7.7 and 7.8) at many stages. What happens to these intermediates? It has been suggested that partial activation of complement may have a physiological role and that not all aspects of complement activation are harmful or even beneficial to the host overall. For instance, partially activated complement molecules may not cause lysis of the target cell; they may coat the cell and promote apoptosis; this is then followed by the silent (i.e. non-inflammatory) removal of the cell debris by scavenger macrophages. Such a mechanism may be occurring at some level of activity during the normal housekeeping actions of the resident macrophages. Indeed, it is likely to take place at sites where there is minimal cell turnover but where removal of cell debris is important, e.g. at the retinal pigment epithelium. Such a fine homeostatic mechanism may be susceptible to dysfunction: indeed, mutations in

Box 7.7 **Inhibiting the complement cascade**

The outline of the scheme for the complement cascade is presented in the figure. Sites at which novel complement inhibitors, such as CSMD1, SUSD4 and ECM proteins, might act are shown.

The complement cascade can be triggered by three canonical pathways (classical, lectin and alternative) that all lead to the assembly of C3- and C5-processing enzymes termed 'convertases'. These multiprotein complexes are responsible for the proteolytic activation of the central components C3 and C5 and the release of their respective bioactive fragments, C3a and C3b, and C5a and C5b and finally formation of a lytic membrane attack complex *(MAC)*, which disrupts membrane integrity (see main text).

The cascade is tightly regulated by several inhibitors that can be either soluble or bound to cell surface. The family of complement inhibitors has been extended in the recent years with new proteins containing complement control protein *(CCP)* domains as well as components of exracellular matrix *(ECM)*.



AMOP, Adhesion-associated domain present in MUC4 and other proteins; *C4BP*, C4b-binding protein; *CCP*, complement control protein; *CR1*, complement receptor1; *CR3*, complement receptor3; *CS*, chondroitin sulphate; *CSMD*, CUB and Sushi multiple domains; *DAF*, decay-accelerating factor; *DAMPs*, damage-associated molecular patterns; *DRAGO*, drug-activated gene overexpress protein; *DS*, dermatan sulphate; *ECM*, extracellular matrix; *EGF*, epidermal growth factor; *ER*, endoplasmic reticulum; *ERα*, oestrogen receptor α; *FACIT*, fibril-associated collagen; *FH*, factor H; *FI*, factor I; *GAGs*, glycosaminoglycans; *Gal-1*, galectin-1; *KS*, keratan sulphate; *LRRs*, leucine-rich repeats; *MAC*, membrane attack complex; *MCP*, membrane cofactor protein; *miRNAs*, microRNA; *MMP-13*, matrix metalloprotease-13; *NC4*, N-terminal non-collagenous domain 4; *PGs*, proteoglycans; *PRELP*, proline/arginine-rich end leucine-rich repeat protein; *SCR*, short consensus repeats; *shRNA*, short hairpin RNAS; *LRPs*, small leucine-rich proteoglycans; *SNPs*, single-nucleotide polymorphisms; *SO*, somatomedin B-like domain; *SUSD*, sushi domain-containing; *TNFα*, tumour necrosis factor α; *Treg*, regulatory T cells; *TSP-5*, thrombospondin-5; *TSP-3*, thrombospondin type 3; *UTR*, untranslated region; *VWD*, von Willebrand factor type D domain.

the complement inhibitory gene, complement factor H (*CFH-402His*), have been associated with a higher than normal risk of AMD and other alleles of CFH have been shown to confer resistance to AMD. In addition, there are additional factor H related (CFHR) genes that are also associated with AMD as well as a splice variant of CFH (CFHL-1) mutations in which genes are linked to AMD. CFH and CFHL-1 bind to Bruch's membrane (BM), which is where they exert their complement-activation inhibitory effect. This they do by binding to GAGs in BM for which the histidine residue in CFH is considered to

Table 7.6 **Receptors for fragments of C3**

Receptor	Structure	Ligands	Cell distribution	Function
Type 1 complement receptor (CR1, CD35)	160–250 kDa; multiple CCPRs	C3b > C4b > iC3b	Mononuclear phagocytes, neutrophils, B and T cells, erythrocytes, eosinophils, fDCs	Phagocytosis Clearance of immune complexes Promotes dissociation of C3 convertases by acting as cofactor for cleavage of C3b, C4b
Type 2 complement receptor (CR2, CD21)	145 kDa; multiple CCPRs	C3d, C3dg > iC3b	B lymphocytes, fDCs, nasopharyngeal epithelium	Co-receptor for B-cell activation Trapping of antigens in germinal centres Receptor for EBV
Type 3 complement receptor (CR3, Mac-1, CD11bCD18)	Integrin, with 165–kDa α chain and 95–kDa β_2 chain	iC3b, ICAM-1; also binds microbes	Mononuclear phagocytes, neutrophils, NK cells	Phagocytosis Leukocyte adhesion to endothelium (via ICAM-1)
Type 4 complement receptor (CR4, p150,95, CD11cCD18)	Integrin, with 150-kDa α chain and 95-kDa β_2 chain	iC3b	Mononuclear phagocytes, neutrophils, NK cells	Phagocytosis, cell adhesion?

CCPRs, Complement control protein repeats; *EBV*, Epstein–Barr virus; *fDCs*, follicular dendritic cells; *ICAM-1*, intercellular adhesion molecule 1.
(From Abbas, A.K. *Cellular and Molecular Immunology* 7th ed. Elsevier, 2012, Table 12.8, p. 284.)

be important. Complement inhibitor (iC3b) therapies are in clinical trial to prevent geographic (dry) AMD.

Thus, factor H provides a check on the complement cascade at two critical points, C3 and C5 induction, by their immediate precursors.

In contrast, defective complement activation may fail to clear opsonized pathogenic organisms properly and continued cell damage may occur to the detriment of the host (Fig. 7.16).

Cellular mechanisms of tissue damage

Tissue damage in cell-mediated immune reactions may be induced by a variety of cell types including macrophages, cytotoxic T cells and NK cells. In addition, macrophages are involved in the clearance of cell debris in acute inflammatory reactions and in antibody-dependent cytotoxicity.

The delayed-type hypersensitivity reaction is the hallmark of cell-mediated immunity

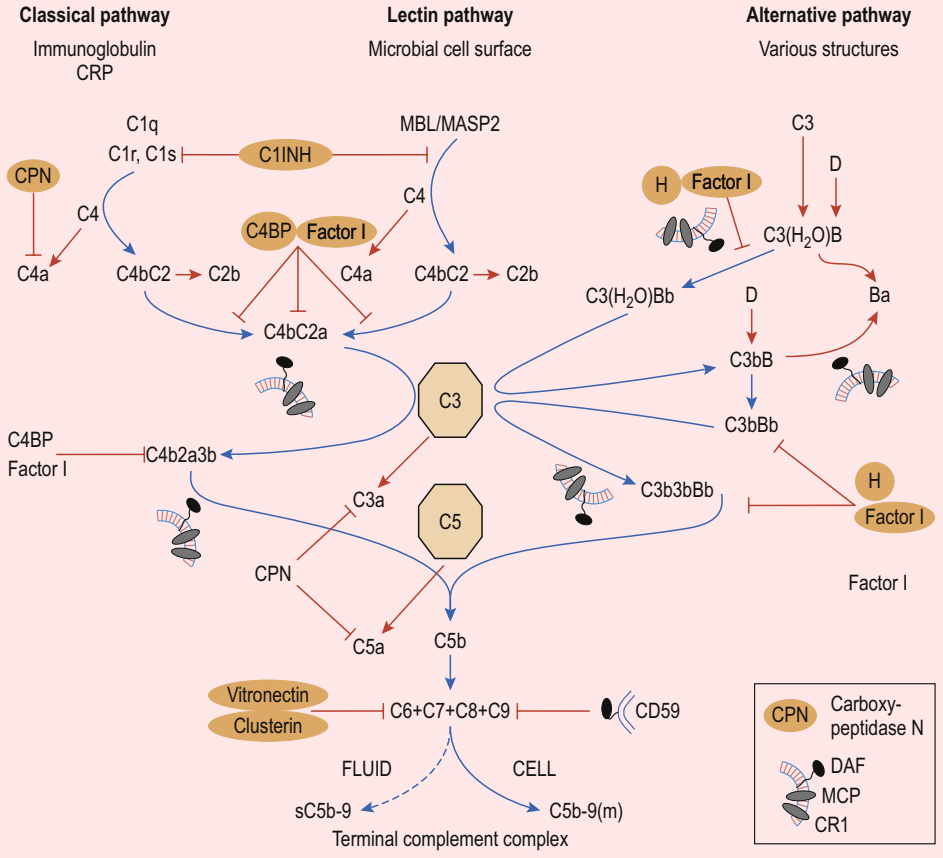
The delayed-type hypersensitivity reaction is a Th1-mediated reaction to foreign antigen and/or autoantigen and is characterized by the presence of granuloma in the tissues. These accumulations of cells contain a central core of macrophages around a vessel with T cells in the surrounding area. Fibrinoid necrotic material may be present in the centre with giant cells (fused macrophages) and epithelioid cells. Such lesions are typical of reactions to mycobacteria and also occur in less well-defined diseases such as sarcoidosis. Similar microgranulomas are typical of sympathetic ophthalmia (Fig. 7.17) and indeed of several chronic posterior uveitis syndromes as well as infectious disease such as viral keratitis. The innate immune system is prominently involved, for instance with NOD2 receptors activated by mycobacterial N-glycolyl mycobacterial muramyl dipeptide and TLR-9 by herpes simplex virus (HSV).

Such granulomas contain many types of T, B and macrophage-like cells, and also have a high content of dendritic cells. It is possible that these cell collections represent small extralymphatic lymphoid follicles known as tertiary lymphoid organs (TLOs) where extensive antigen presentation is in progress. Once the antigen has been removed, the granuloma subsides.

Inflammatory (M1-like) macrophages cause tissue damage: alternatively activated (M2-like) macrophages are more likely to promote healing

Macrophages cause tissue damage by releasing reactive intermediate metabolic products and tissue proteases, express high levels of adhesion molecules and chemokine receptors such as CCR5 and CCR2 and are generally recruited *de novo* from circulating monocytes. However, as indicated previously, macrophages are heterogeneous and not all macrophages behave in this manner. In the resting state, tissues contain resident macrophages that, in certain tissues, may have an immunoregulatory role (such as alveolar

Box 7.8 **Innate immunity: new insights in complement and collectins**



macrophages in the lung). In central nervous system tissue, including the retina, specialized resident tissue macrophages occur (microglia) (see Chapter 1, p. 40). These cells may be induced to express MHC class II antigen during inflammation but are still more likely to promote tissue homeostasis. Such cells express the mannose receptor, arginase and the fractalkine receptor (Cx3CR1) (see p. 385). These 'alternatively' activated macrophages are more likely to occur in parasitic infections controlled by cytokines such as IL-3, IL-5 and IL-13 and promote a predominantly eosinophilic response. In addition, there is more likely to be fibrosis than tissue damage with angiogenesis.

Cytotoxic T and NK cells kill cells by perforating the cell membrane

Killing by cytotoxic T cells is a multistep process. Initial recognition and binding of a target cell (e.g. an infected or mutated cell) is followed by damage to the cell membrane. This lethal insult induces apoptosis with DNA fragmentation and lysis of the cell. The cytotoxic cell then disengages to attack another target cell.

Membrane damage takes the form of pore formation similar to that induced by the complement MAC (see earlier). In this case, however, it is induced by perforin, a protein released from lysosomes of antigen-activated Tc cells that polymerizes in the cell membrane to form a leaky pore (Fig. 7.18). Perforin is released in association with a granule proteoglycan and serine proteases (granzymes), which are also likely to be involved in the lytic process. However, cell death is not automatically caused by osmotic lysis. Instead prelytic DNA breakup occurs as a result of activation of the apoptosis genes and killing occurs via caspase-3 and -7. Therefore, there are important differences between MAC-mediated cell lysis and Tc/NK cell killing.

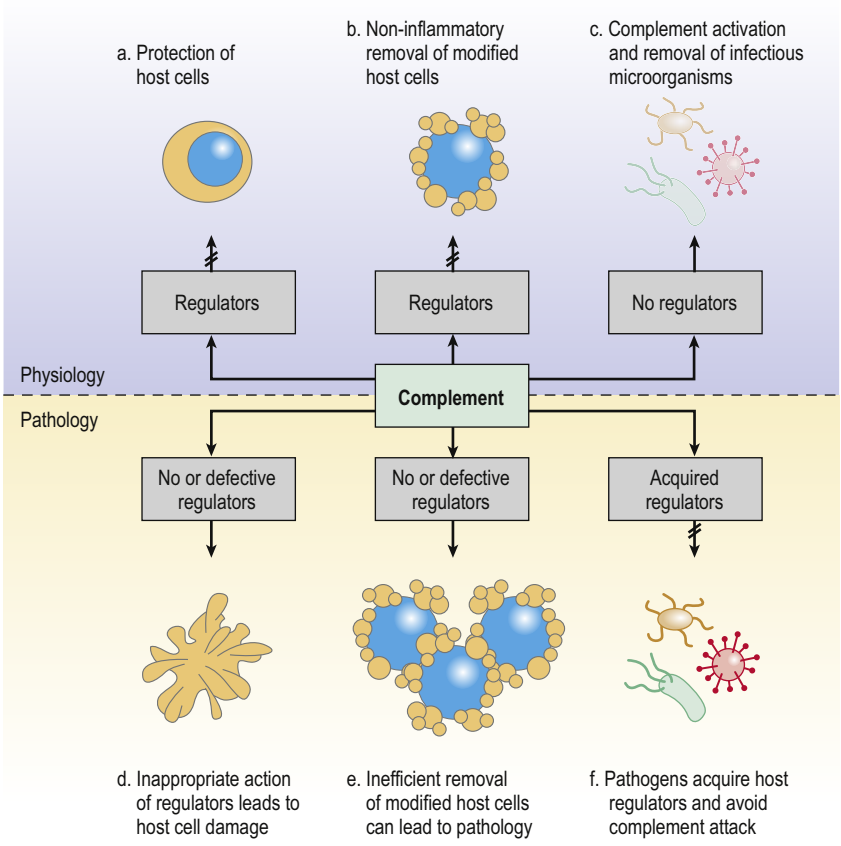


Fig. 7.16 Complement is a double-edged sword for the host. It can either have beneficial effects as in the *upper panel* where it can protect viable cells (*a*), remove aged or dying cells (*b*) or remove infectious organisms (c). If these mechanisms fail, healthy cells can be damaged (*d*), dead cell debris can accumulate and cause inflammation (*e*) or pathogens can use complement to evade the immune system (*f*). In the eye, mechanism (*d*) is thought to account for diseases such as age-related macular degeneration because of genetic faulty complement proteins such as complement factor H. (From Zipfel and Skerka, 2009.)

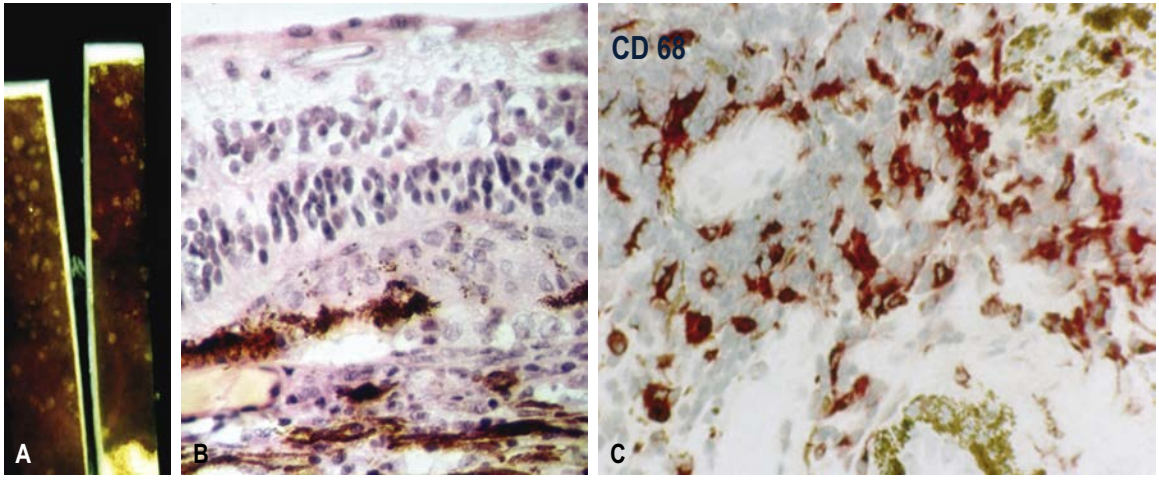


Fig. 7.17 (**A**) Macroscopic view of a case of sympathetic ophthalmia showing Dalen–Fuchs granulomatous lesions as *white 'excrescences'*; (**B**) macroscopic view of Dalen–Fuchs granuloma; (**C**) immunohistochemical view of activated (CD68+) myeloid cells in granuloma surrounding an occluded vessel. (Courtesy Prof. W.R. Lee.)

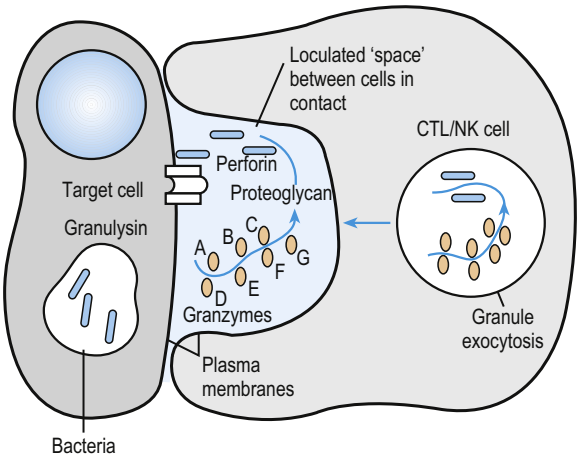


Fig. 7.18 Perforin killing.*CTL*, Cytotoxic T-lymphocyte; *NK*, natural killer cell.

Alternative mechanisms have also been shown for T-cell cytotoxicity. $CD8^+$ T cells may recognize non-peptide (e.g. lipid) antigen in the context of CD1. Thus, cells containing bacteria are directly recognized by $CD8^+$ T cells and the bacterial lipid induces the release of a novel enzyme, granulysin, which directly kills the intracellular bacteria.

Organization of the immune system

The immune system is designed to provide cells that can circulate freely through the tissues and organs of the body in such a way that they are readily available to mount a defence against foreign organisms at short notice. Immune cells are thus highly motile, normally quiescent cells travelling to and from the lymphoid organs; they can be readily activated if required.

Centralized antigen recognition mechanisms provide the most efficient means of rapid response because all the necessary requirements for cell activation can be concentrated at one site. This takes place in the lymphoid organs (primary and secondary lymphoid organs, SLOs). Some cells carry afferent information concerning possible breaches in the body's defences to the central lymphoid tissues (particularly DCs and other APCs), while effector cells (T and B cells) remove the invading organism (or arrange for this to be done by other cells such as inflammatory macrophages) and restore tissue homeostasis. Trafficking of cells to and from the tissues to the lymphoid organs requires specific receptors on the circulating leucocytes and the vascular endothelium in each tissue (chemokine and adhesion molecule receptors).

Tertiary outposts of local antigen presentation can be set up in sites where normal lymphoid tissues do not exist (such as the brain and eye) or where persistent antigen generates a chronic inflammatory response (frequent sites are liver and lung). Granulomas, the hallmark of chronic inflammation in the tissues usually forming at sites of small blood venules, represent one type of lymphoid cell accumulation where an antigen processing and presentation 'factory' might develop, but true tertiary lymphoid organs (TLOs) containing discrete T- and B-cell areas with a lymphatic, as well as a blood circulation, may also develop *de novo* at sites of persistent chronic inflammation.

Functional anatomy of lymphoid organs

During early development, the primary sources of lymphomyeloid cells are first the yolk sac and then the fetal liver. As the fetus matures, the bone marrow and thymus (primary lymphoid organs) take over. Ultimately, in adults, all newly forming cells derive from stem cells in the bone marrow. Stem cells are poorly characterized and their existence is based mainly on lack of evidence for cell differentiation, often from *in vitro* cell culture studies. For instance, haematopoietic stem cells have few lineage markers (i.e. they are Lin^-) and develop to common myeloid progenitors and common lymphoid progenitors. These then give rise to T, B and NK/NKT cells (via IL-7) or myeloid cells (macrophage/dendritic cells/neutrophils (M-CSF, GM-CSF, G-CSF)). B cells mature in the peripheral/secondary lymphoid tissues, particularly Peyer's patches (lymphoid tissue structures in the wall of the small intestine). T-cell precursors from the bone marrow colonize the thymus, where they undergo selection and lineage differentiation before being distributed to the SLOs.

Bone marrow stem cells produce all blood cells

Blood cells, including red cells and platelets, granulocytes, monocytes and DCs, are produced in the bone marrow in adults and released directly into the circulation. Sometimes, as in the case of DCs, they are released as late stage progenitor cells and undergo final differentiation in the peripheral tissues. B cells are also released directly into the circulation and circulate between the lymphoid organs and blood.

The marrow is a loose spongy stromal network whose cells, together with local macrophages, release growth factors (cytokines, see previous) that initiate differentiation of each cell type. IL-3, G-CSF, GM-CSF and M-CSF are particularly important. IL-1 and IL-6, released by stromal marrow cells, also participate in T-cell maturation, while IL-7 promotes B-cell development. The stroma is also the source of mesenchymal stem cells, multipotent cells that can give rise to several, but not all, cell types (cf pluripotent stem cells). However, although these cells are being promoted for treatment of a variety of ocular diseases including AMD, their precise nature is unclear and their mechanism of action uncertain. Stromal marrow cells are, however, essential to the survival of the host.

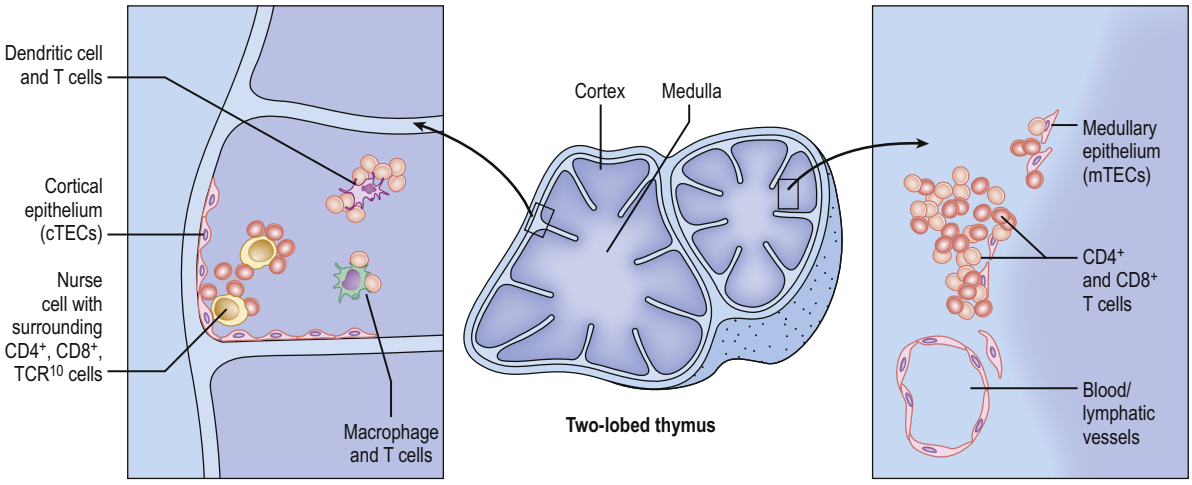


Fig. 7.19 The bi-lobed thymus. The central region shows the cortex and the medulla, while the cortical thymic epithelial cells *(cTECs)* and medullary thymic epithelial cells *(mTECs)* are shown to *left* and *right* in greater detail.

The thymus regulates T-cell development and maturation

T-cell precursors enter the thymic cortex (Fig. 7.19). Here they interact with cortical thymic epithelial cells (cTECs or 'nurse' cells), which encapsulate them in large numbers. At this stage they are still immature cells expressing cell surface markers for Th and Tc cells (CD3+, CD4+ and CD8+: the so-called double-positive cell.

An explanation of CD numbers is available online at https://expertconsult.inkling.com.

After some time, they are released from the nurse/cTEC cells and migrate through the cortex to the medulla, making contact with macrophages and DCs as they go. Many T cells die during this process (clonal deletion and/or apoptosis) but selected cells differentiate to express one or other T-cell phenotypic marker (CD4 or CD8). Maturation in the medulla also involves contact with the medullary thymic epithelial cells (mTECs) cells and thymic DCs, which express high levels of MHC class II antigen. The T cells then leave the thymus, enter the blood vessels and circulate to the lymph nodes and spleen.

cTECs and mTECs arise from a common TEC progenitor probably located at the interface between the cortex and medulla. mTECs are important in generation of Tregs as well as deletion of tissue-specific autoreactive T cells through the autoimmune regulator (*AIRE*) gene. Defects in this gene cause a condition in humans known as the autoimmune polyendocrinopathy candidiasis ectodermal dystrophy syndrome (APECED), which includes a low-grade form of uveitis caused by the failure to delete T cells reactive against IRBP (see p. 440). The medulla of the thymus contains several characteristic whorled bodies (Hassall's corpuscles) (Fig. 7.19), which are probably remnants of 'out-of-date' nurse epithelial cells. In adults, the thymus involutes but some T-cell maturation continues into adult life, both in the thymic remnant and extrathymically.

Lymph nodes are designed for antigen trapping

Lymphocytes and APCs from the tissues enter the cortex of the lymph node via the afferent lymphatics (Fig. 7.20) where they present antigen to T and B cells. Follicular dendritic cells (fDCs) in germinal centres of the lymph nodes and the spleen present antigen–antibody complexes to B cells in the B-cell areas, while conventional dendritic cells (cDCs) present antigenic peptides to T cells in the T-cell area. cDCs comprise two main types: migratory DCs (CD8α-) carrying antigen to the lymph node and resident, 'lymphoid' CD8α+ DCs, which may sample soluble antigen in the afferent lymph or receive antigen from migratory DCs. The eye (and the brain) connect directly with cervical lymph nodes.

Lymph nodes are organized in chains, the more peripheral node capturing soluble and cell-associated antigen from the tissues, which may be presented to T cells. T cells leave (egress) via the blood vessels and home to the site of inflammation containing the target antigen (e.g. virus-infected cells). Naive T cells (i.e. non-activated T cells) also egress through the efferent lymphatic and traffic to the next more central lymph node in the chain until they reach the thoracic duct and re-enter the circulation to begin another cycle of migration.

Antigen is carried in soluble form to the lymph node in blind-ended vascular lymphatics from the tissues and in cell-associated form inside migratory cDC. cDCs migrating from the tissues to the initial peripheral lymph nodes do not leave the node but present antigen to the LN T and B cells. However, naive unactivated T and B cells circulate between the lymphatics and the bloodstream as indicated and activated T cells specifically home to sites of inflammation.

The spleen receives antigen from all sources (lymphatics and blood)

The spleen is a central secondary lymphoid organ that acts as a trap for red cells and immune cells from the blood. The bloodstream communicates with the lymphatic system at the thoracic duct where recirculating lymphocytes enter. The spleen receives antigen-primed APCs from all regions, including those that may have bypassed their regional lymph node. For instance, APCs that encounter antigen within the anterior chamber of the eye may pass through the trabecular meshwork and aqueous veins (see Chapter 1, p. 24) to enter the conjunctival veins and eventually drain into the lungs and spleen. However, the spleen has no afferent lymphatics, i.e. it only receives lymphocytes from the bloodstream that are overwhelmingly naive, unactivated T cells. Aged immune and red cells are removed from the circulation by splenic macrophages and healthy cells re-circulate.

The spleen is organized like a lymph node with the addition of the red pulp (red cell area) and its venous sinuses (Fig. 7.21). T-cell and B-cell areas are separated, with the T-cell areas (white pulp) being particularly large and aggregated around arterioles (the periarteriolar lymphoid sheath, or PALS), surrounded by well-defined macrophage/dendritic cell zones (marginal zone).

Tertiary lymphoid organs

TLOs (see introduction to this section, p. 420) develop around sites of persistent chronic inflammation, spurred on by lymphoid tissue inducer (LTi) cells, similar to those which establish the development of primary and secondary lymphoid tissues which secrete lymphotoxin β (LTb). Chemokines such as CXCL13 and cytokines such as IL-17 derived from the inflammatory process provide the conditions for the further differentiation of lymphoid vessels and the T- and B-cell separation (Fig. 7.22).

The mucosal immune system

Initial contact of the host with exogenous antigen takes place at surfaces such as skin and mucous membranes. The skin provides an effective physical barrier to microorganismal contact unless penetration occurs, as with insect vectors and trauma. However, the mucous membranes are more easily breached by microorganisms and, not surprisingly, considerable lymphoid tissue is concentrated at these surfaces to deal with new antigens as they arrive. This is termed the mucosa-associated lymphoid tissue (MALT) and is specialized to some degree for each tissue, e.g. bronchus-associated lymphoid tissue (BALT), gut-associated lymphoid tissue (GALT) and conjunctiva-associated lymphoid tissue (CALT). In essence it comprises intraepithelial lymphocytes and APCs (especially DCs) and aggregations of lymphoid tissue in the mucosal layers. These are best exemplified by the Peyer's patches of the gastrointestinal mucosa and the tonsils in the oropharynx. Regional specialization occurs in various compartments of the mucosal immune system with specific homing receptors ('addressins', adhesion molecules) directing movement of cells to different mucosal sites (see later) and, like TLOs, are induced by innate lymphoid cells (ILCs) (see p. 391), such as LTis as well as a second set of ILCs that are NK-like IL-22 and (possibly) IL-17-secreting cells. In the lamina propria of the gut, $CD4^+$ Th17 cells and $CD4^+$ Treg cells abound and are important cells balancing the response of the host to commensal versus pathogenic microorganisms in the microbiome (see p. 383 and Box 7.1). In addition, there is a further subset of innate T cells, termed mucosa-associated invariant T cells (they express a single TCR) or MIT cells, that recognize small bacterial vitamin B-like molecules presented by MHC Class I related (MRI) molecules. Similar immune defence arrangements are likely to occur in all mucosal tissues but are most developed in the gut as a result of the constant exposure to microorganisms. Thus, there is a plethora of adaptive (T and B cells) and innate lymphocytes (ILCs, γδ T cells and mucosal-associated invariant T cells (MAIT cells)) that reside in the gut and other mucosal tissues and provide this first line of defence. Disturbance in this regulation accounts for many types of inflammatory bowel disease (IBD), some of which have associated systemic involvement leading to complications in other organs including the eye, as in HLA-B27-associated acute anterior uveitis. MAIT cells circulate in the blood and together with ILCs account for nearly 20% of circulating T cells.

As indicated later, MALT is specially adapted for antigen capture and presentation to specific lymphocytes that 'home' to that tissue. In addition, exposure of the host to antigen via the MALT frequently leads to tolerance to that antigen by mechanisms that are as yet poorly understood but involve Tregs (see p. 390).

The immune system as a police force

The immune system can be regarded as a surveillance system constantly checking for intruders. Cells of the innate immune system, such as neutrophils and monocytes, undertake part of this function but at a relatively non-specific level. The adaptive immune system has cells that mediate this function in a much more targeted way. The afferent limb of the immune system

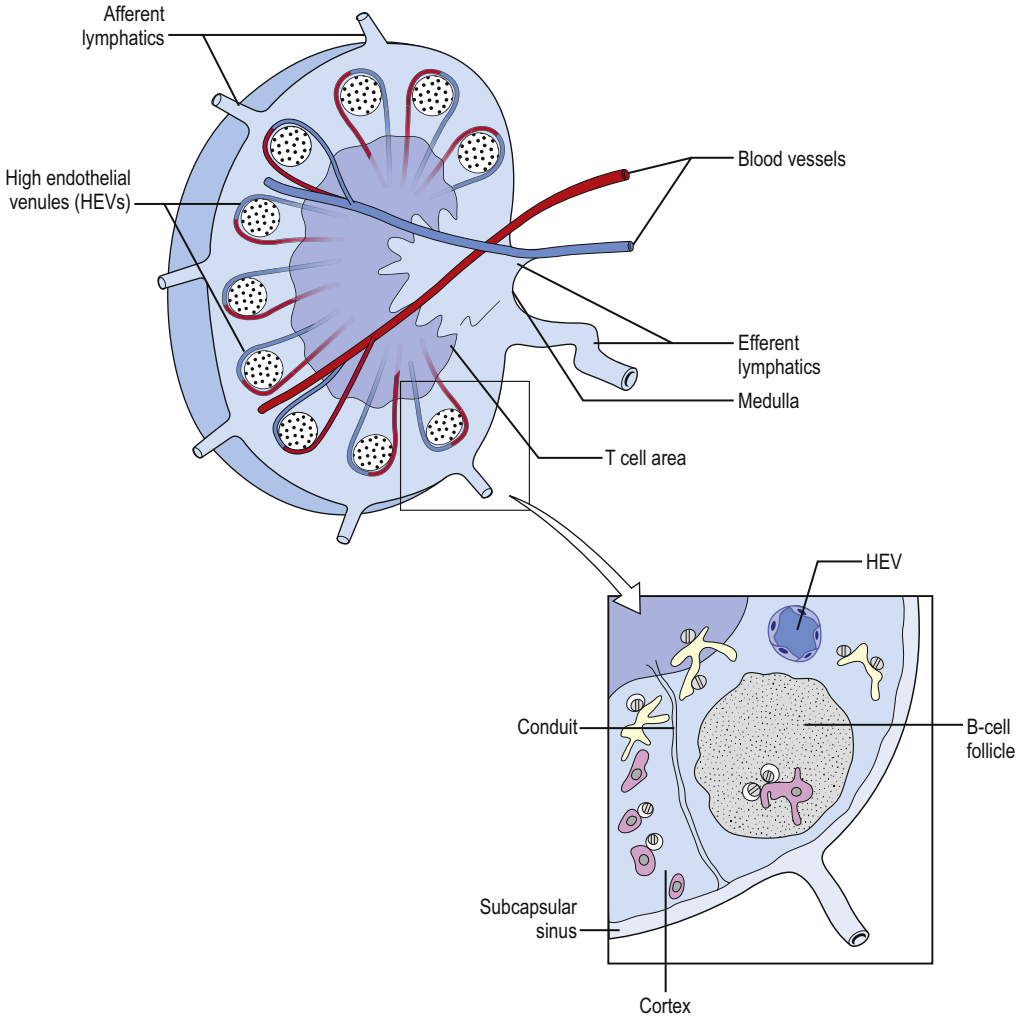


Fig. 7.20 Diagram of the lymph node. The T-cell area surrounds the B-cell follicle in the cortex; high endothelial venules *(HEVs)* are present in the T-cell area, mostly at the junction with the medulla. Conduits allow trafficking antigen-presenting cells to migrate into the node after entry through the subcapsular sinus.

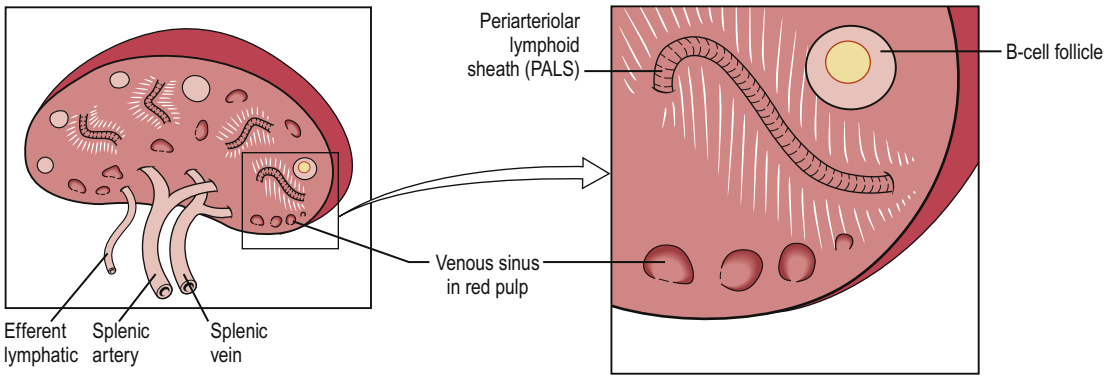


Fig. 7.21 Diagram of the spleen. The *white pulp* contains the T-cell area surrounded by the B-cell follicles; the *red pulp* is situated closer to the medulla and contains the venous sinuses with large numbers of red cells and macrophages. There are no afferent lymphatics to the spleen, all cells entering the spleen stroma via the bloodstream near the periarteriolar lymphatic sheath and leaving via efferent lymphatics.

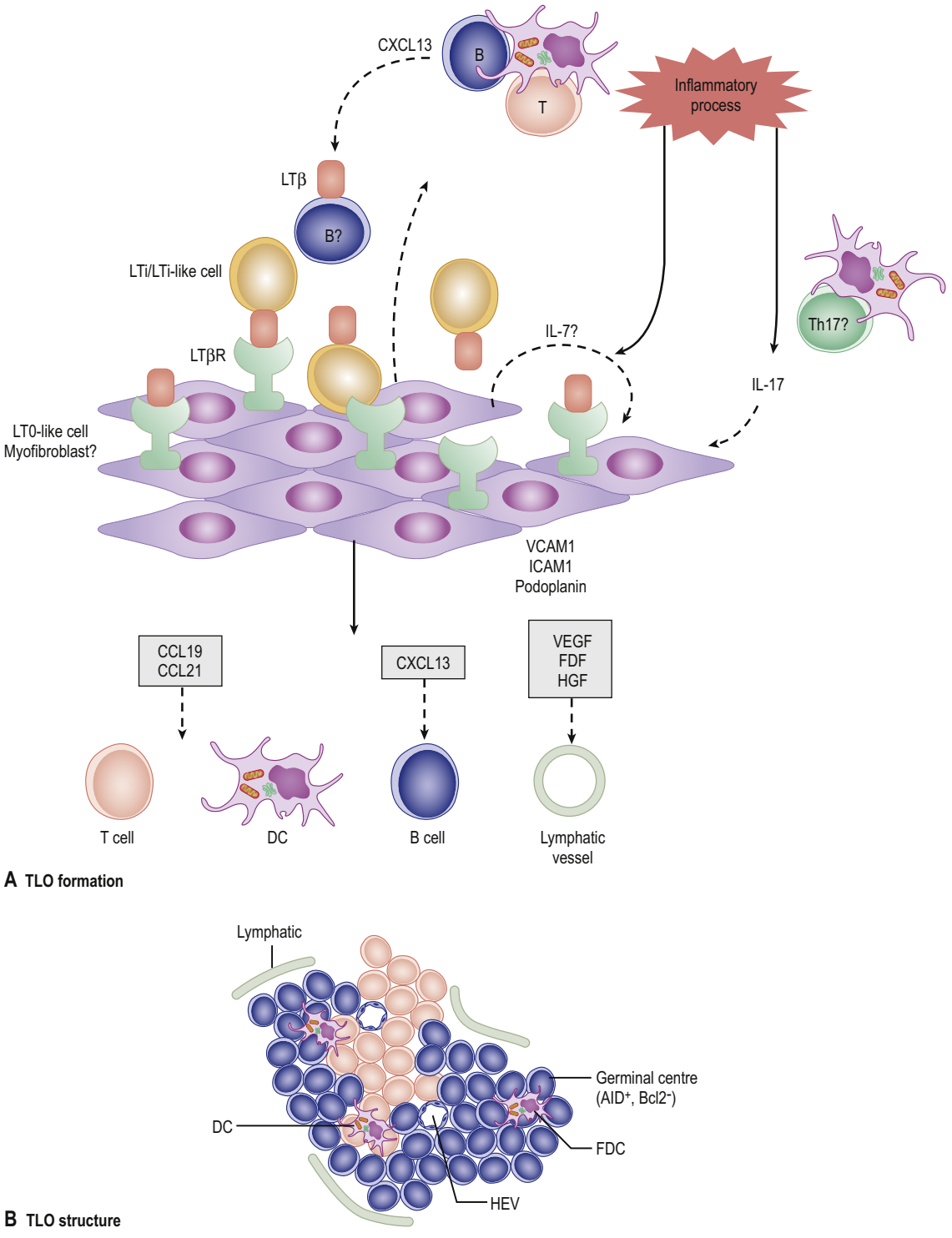


Fig. 7.22 Tertiary lymphoid structures develop during chronic inflammation and take on a form similar to secondary lymphoid structures such as the lymph node and Peyer's patches. In a sense they are outposts of the lymphoid defence system but often do not have the full characteristics of the structured lymph node. However, they can develop germinal centres and afferent and efferent lymphatics. The granuloma has similarities but is different since it contains a central core of antigen with myeloid cells and a large cuff of T cells with some B cells. (From Neyt et al., 2012.)

uses dendritic cells in this role, while the messengers sent out to do the 'dirty work' are the T and B lymphocytes.

Dendritic cells are the major surveillance cells of the afferent lymphoid system

DCs are derived from stem cell precursors in the bone marrow and proliferate/differentiate under the influence of cytokines such as GM-CSF, stem cell factor (Flt3 ligand) and TNF-α. They leave the bone marrow and circulate in various shapes and sizes (one of which is as very large 'veiled cells', like large floating jellyfish) in the bloodstream before entering the tissues, where they remain for a few days to weeks. In the skin, conjunctiva and peripheral corneal epithelium, they can be identified as Langerhans' cells (*bone fide* Langerhans' cells are derived from the embryonic yolk sac like microglial cells in the retina and brain and Kuppfer cells in the liver); in the stroma of these tissues, DCs are mostly migratory cDCs. They wrap their dendrites around tissue cells such as keratinocytes and epithelial cells. In the uveal tract, there is a rich network of DCs in association with tissue macrophages (see Chapter 1, Fig.1.21, p. 28).

Tissue conventional DCs (cDC2 and IRF4$^+$,CD4$^+$) constitutively migrate in the afferent lymph to the lymph nodes (LN) in response to appropriate stimulation (e.g. TNF-α); they interact with T cells in the T-cell areas of these tissues (interdigitating DCs). In contrast, cDCs (cDC1 and IRF8$^+$ Cd8$^+$) that have directly arrived in the SLOs from the bloodstream (e.g. from the bone marrow) without passing through other tissues become resident cDCs and take up antigen from cDC1 that have migrated from the tissues and undergone apoptosis in the LN. Other specialized DCs also reside in the LN that possess Fc receptors and serve as antigen traps because they have the ability to bind immune complexes. They occupy the B cell follicles in LN where they present antigen to B cells (follicular dendritic cells, fDCs). DCs that migrate or reside in the LN do not leave the LN; they are short-lived cells with a half-life of a few days and their numbers are replenished from the bone marrow.

In addition to cDC1 and cDC2, a third set of DCs, termed plasmacytoid DCs (pDCs, so named because they had some morphological similarities to plasma cells), arrive in the SLO from the blood and are programmed to produce constitutively INF-α, which is massively increased in response to viral infections. They may also be involved in tolerance induction (see p. 444).

Trafficking of cells to and from the lymphoid system depends on specific cell surface adhesion molecules

In contrast to DCs, T and B cells recirculate through the lymphoid organs many times. They do this by binding to specific regions of the lymphoid vasculature, the high endothelial venules (HEVs) (see Fig. 7.20), which are specialized for lymphocyte capture by the expression of specific adhesion molecules on their surface. About 25% of the lymphocytes passing through a lymph node will leave the bloodstream and any given naive lymphocyte can traverse each lymph node at least once a day. However, in practice, lymphocyte recirculation follows some patterns: B cells rarely recirculate, γδ T cells tend to reside in skin and most other T lymphocytes recirculate preferentially through their associated LN (each tissue has a site-specific LN). If they fail to meet cognate antigen, i.e. specific to that T cell's receptor (TCR), they remain unactivated naive cells and pass on to the next LN in the chain before finally reaching the thoracic duct, entering the blood stream and re-circulating. Activated lymphocytes, on the other hand, which egress into the thoracic duct and enter the bloodstream, 'home' to (more likely are 'captured' by) HEV-like structures that develop at sites of inflammation (e.g. in retinal vessels during the active phase of retinal vasculitis; see Chapter 9, p. 507). Activated T cells can circulate many hundreds or thousands of times before they are 'captured' (it has been estimated to take at least 16 hours from time of entry into the circulation to time of entry into the retina in mice). Receptors for specific adhesion molecules and chemokines are reciprocally expressed on activated lymphocytes, which assist in directing them to sites of tissue injury.

Further information on address labels for trafficking cells is available online at https://expertconsult.inkling.com

What turns a lymphocyte on?

Most circulating lymphocytes are resting. Most are naive cells (i.e. have not been activated by interaction with antigen) and have recently been released from the thymus or have passaged through the peripheral lymphoid tissues. Others are memory T and B cells that can be very long-lived. Memory cells are a residual population after most of the antigen-specific T cells have died once their effector duties are completed (i.e. getting rid of the pathogen). Both CD4$^+$ and CD8$^+$ T memory cells occur, some residing long term in the tissues (e.g. the lung parenchyma) where they can rapidly respond to re-challenge with the antigen (e.g. influenza virus). These cells are known as effector memory T cells. Other memory cells recirculate through the lymphoid tissues and are known as central memory cells. There is also a concept that memory cells arise de novo as a separate population during T cell activation. Depending on the local conditions and the cocktail or 'panel' of cytokines where the T cell is exposed to antigen presentation, further functional diversification is achieved by driving T-cell clonal expansion towards one or more of several subsets (Fig. 7.23).

Further information on T-cell differentiation by cytokines and chemokines is available online at https://expertconsult.inkling.com

ADDRESSINS AND CHEMOKINES IN LYMPHOCYTE TRAFFIC

Specific adhesion molecules are expressed on HEV in lymph nodes, where they are involved in physiological lymphocyte recirculation. These molecules are therefore sometimes referred to as 'addressins' (one address only!). For each addressin there is a corresponding lymphocyte 'homing receptor'. Circulation through typical lymph nodes is mediated by L-selectin on lymphocytes, which binds to the adhesion molecule GlyCAM-1 on the HEV. In Peyer's patch lymph nodes, HEVs express MAdCAM (mucosal addressin cell adhesion molecule) and VCAM-1 (see Table 7.3), which are the ligands for $\alpha_4\beta_7$ and $\alpha_4\beta_1$ integrins of lymphocytic microvilli. It is likely that each tissue has specific addressins that direct the circulation of surveillance lymphocytes through that tissue; in particular, most mucosal lymphoid tissue (MALT, CALT; see earlier) has addressins that ensure adequate trafficking of lymphocytes through that tissue during health and disease.

Some cytokines and chemokines facilitate the interaction of T cells with activated endothelium, such as macrophage inflammatory protein-1β (CCL4), which has been shown, for instance, to be particularly effective in mediating $CD8^+$ T-cell adhesion to the endothelium (see earlier section on chemokines). In addition, certain adhesion molecules are involved not only in endothelial cell interactions but also in mediating the close cell–cell contact required during antigen presentation (see later), particularly ICAM-1/LFA-1, ICAM-1/LFA-3, CD44 and its ligand cell surface hyaluronan, plus several others (see Table 7.3). In addition, chemokines and their receptors are intimately involved in the recirculation and trafficking of T and B lymphocytes and of APCs. In general, however, it is now clear that adhesion of leucocytes, either to each other or to the endothelium, during normal recirculation or as part of inflammation, is a complex molecular and tissue-specific process.

T-CELL DIFFERENTIATION AFTER ACTIVATION

Th1 cells, expanded by IL-2 and IFN-γ, induce a delayed-type hypersensitivity (DTH) response (type IV hypersensitivity in the classic nomenclature) (Table 7.7), which involves considerable macrophage activation, granuloma formation and tissue damage. It should be noted, however, that memory T cells do not release IFN-γ, while release of the chemokine RANTES (CCL5, see Table 7.5) by tissue cells appears to be important in the recruitment of memory T cells and macrophages to the site of granuloma formation. IL-12, IL-23 and IL-27 produced by macrophages and DC also play a major role in directing Th1/Th17 T-cell responses.

In contrast, Th2 cells, expanded by IL-2 and IL-4, promote the production of allergy and IgE antibodies. In addition, IL-4 inhibits Th1-cell activation and the delayed-type hypersensitivity reaction. IL-10 and IL-13 produced locally in tissues and LNs may condition APCs to induce a Th2 response. IL-2 and TGF-β combine to promote Treg formation, while IL-6 and TGF-β promote Th17 T cells.

Th1/Th17 (DTH) and Th2 cells are directed by particular sets of cytokines that determine the nature of the pathological event. Classically, four types are described (types I, II, III and IV hypersensitivity; see Table 7.7) but mixed responses also occur. Th1/Th17 and Th2 cells may initiate reciprocal downregulation of either response via regulated cytokine release. Chemokines and chemokine receptor expression also play a major part in determining whether a Th1/Th17 or a Th2 response is induced (see p. 390).

Table 7.7 Immunopathology of tissue reactions

Hypersensitivity response type	Immune effector process in tissue	Type of mechanism
I	Allergic reaction	Humoral (IgE)
II	Cytotoxicity	Humoral (IgG/M)[a]
III	Complement	Humoral (Ag/Ig)
IV	Macrophages	Cellular (T cells)[b]

[a]Cytotoxicity in this process is antibody-dependent.
[b]Type IV hypersensitivity is the classically described delayed hypersensitivity response, mediated by macrophages and memory T cells.

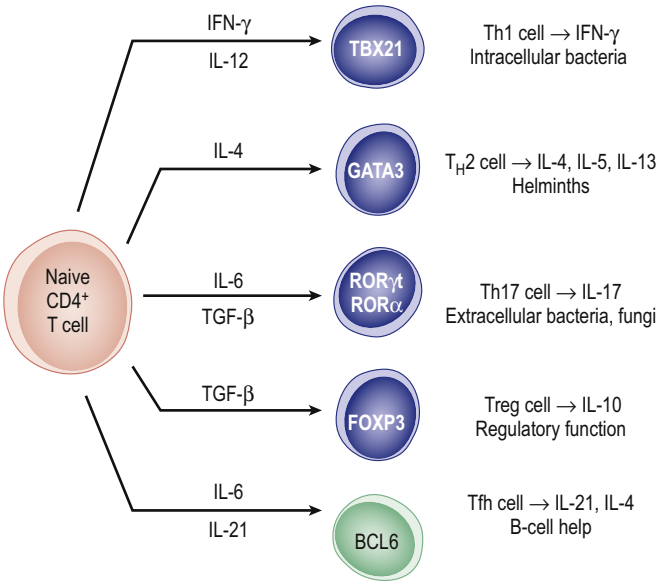


Fig. 7.23 The local microenvironment provides signals that permit differentiation of 'newborn' naive T cells emerging from the thymus. The cytokines signal via specific receptors and activate transcription factors specific for each cell type, which sets in motion a protein synthesis programme leading to production of factors specific to that newly differentiated T cell. *BCL6*, B-cell lymphoma 6; *FOXP3*, forkhead box protein 3; *GATA3*, GATA-binding factor 3; *RORα*, retinoid-related orphan receptor-α; *RORγt*, retinoid-related orphan receptor-γt; *TBX21*, T-box transcription factor TBX21; T_{FH}, follicular helper T; *TGF-β*, transforming growth factor-β; *Th1*, T-helper-1; *Th2*, T-helper-2; *Th17*, T-helper-17; *Treg*, regulatory T. (From Craft, 2012.)

Where are memory cells found?

Activated T-effector cells that exit the LN lose their expression of L-selectin and cannot therefore recirculate through the lymph node. Most of them home to the site of inflammation to promote removal of the antigen and then die *in situ* activation-induced cell death (AICD); as indicated previously, a small number of these cells have the potential to survive as memory cells, some of which re-enter the lymph–blood circuit as central memory cells while others remain long term in the tissues as effector memory cells. In contrast, memory B cells do not recirculate but continue to produce antibody of different isotype, with increasing affinity for the antigen, and release it into the blood via the efferent lymphatics. Memory cells for the mucosal lymphoid tissues have been shown to express a specific addressin, LPAM-1 (lymphocyte Peyer's patch adhesion molecule-1), or $\alpha_4\beta_7$ integrin, which binds to MAdCAM-1 in the GALT (see p. 422). IL-7, an essential cytokine for ontogeny of T cells in mice, is also required for induction of memory cells.

What turns a lymphocyte off?

Most immune reactions do not persist and it has been presumed that this is the result of effective removal of antigen. Indeed, the corollary is also probably true: that most chronic immunological diseases are the result of persistence of antigen in the tissues in a form that activates lymphocytes, i.e. on the surface of APCs. However, it is now recognized that switching off an immune response is an active process, involving a range of processes including AICD as well as direct action by subset of T cells known as T regulatory cells (Tregs).

What exactly are T regs and how do they work?

Lymphocytes that are not exposed to their cognate antigen in the appropriate form (i.e. on the surface of APCs and in the presence of co-stimulatory support mechanisms) enter a state of anergy and eventually die. The process of cell death or apoptosis during lymphocyte ontogeny is well established and is associated with the expression of specific 'suicide' or 'death' genes in the cell (see Chapter 4, Fig. 4.8 and Chapter 9).

Tregs also switch off inflammatory responses. Tregs are the most frequent cell type in the circulation and in the SLOs. Antigen-specific Tregs are induced as part of all adaptive immune responses to exogenous and endogenous (including self) antigens but these expand with a delayed kinetic compared with effector cells, allowing T-effector cells to finish clearing infection and other damage-inducing agents (DAMPs, see p. 381). Tregs are characterized by expression of high levels of the IL-2 receptor (CD25) and the transcription factor FoxP3, which is essential for their function. Tregs are generated constitutively in the thymus (natural or thymic Tregs) and inducible (iTregs) are generated in the periphery from naive $CD4^+$ T cells after presentation of antigen by APCs, e.g. after infection. Antigen-specific iTregs are more effective than thymic or natural Tregs but are relatively non-specific in their action (i.e. although

they are induced by specific antigen, their function is not restricted to suppressing antigen-specific T-effector cells, i.e. they have a broad-spectrum effect. IL-10 and TGF-β are important immunosuppressive cytokines secreted by Tregs. They are present in mice and humans and currently there is considerable interest in developing methods to induce them to high levels to prevent a range of autoimmune diseases.

$CD4^+$ $CD25^+$ T regulatory cells are only one set of suppressive T cells that exist; there are also other $CD4^+$T cells that produce high levels of IL-10 and/or TGF-β (Tr1 cells), CD8 T regulatory cells, which may be involved in aspects of ocular immune privilege and require NK T cells to mediate their activity (see p. 392), as well as B regulatory cells (Bregs).

Antigen recognition

The immune system recognizes antigen through molecular interaction with specific molecules, three of which are antibody (B-cell receptor, sIg), the TCR, and the MHC molecule. MHC molecules are highly polymorphic: i.e. although basically similar in structure, there are many different types (alleles) that occur in each individual, generated by small differences in nucleotide sequences in the MHC genes. Each TCR, antibody and MHC molecule has varying degrees of ligand-binding specificity, the MHC molecule being required to bind to a large variety of antigens, while antibody is specific for only one antigen. The interactions between antigen and antibody and between antigen and the TCR have, however, many similarities.

The great majority of immune responses are T-cell-dependent in that the final effector response, including B-cell responses, is achieved through an initial interaction between a resting T cell and an APC (Fig. 7.24).

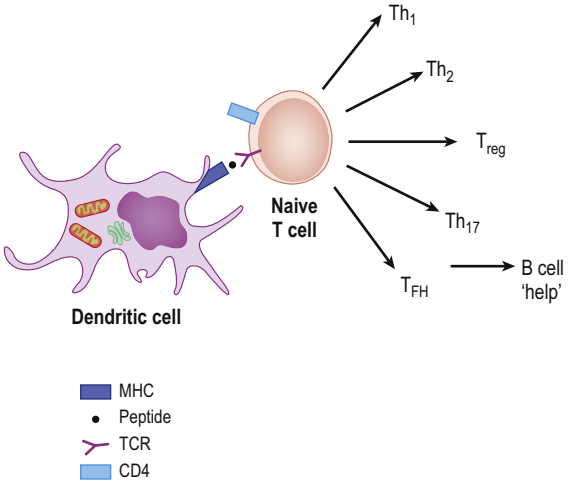


Fig. 7.24 Diagram of interaction between dendritic cells presenting peptide on the major histocompatibility complex *(MHC)* molecule to a naive T cell prior to T-cell differentiation.

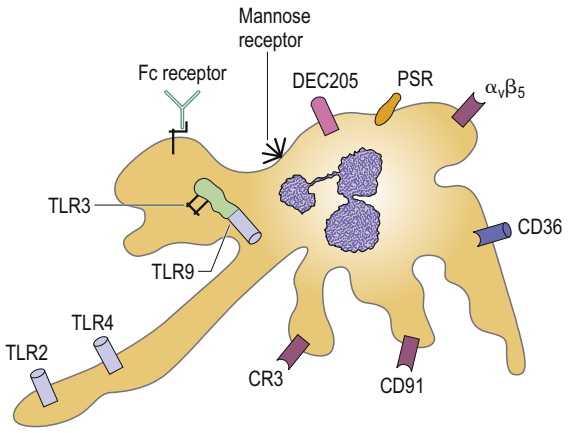


Fig. 7.25 Diagram of an immature dendritic cell showing some cell surface receptors for innate immune responses. *PSR*, Phosphatidylserine receptor; $\alpha_v\beta_5$, integrin; *TLR*, toll-like receptor.

The APC makes the antigen recognizable

The function of the APC is to capture antigen and to process it into a form that can be recognized by the T cell. The immature (veiled) DC with its numerous cell surface receptors is ideally designed for this (Fig. 7.25). It does this by partially digesting the antigen into short peptides and combining these with MHC molecules. Classically, intracellular antigens are complexed with MHC class I molecules for presentation to cytotoxic ($CD8^+$) T cells, while extracellular antigens are complexed with MHC class II molecules for presentation to helper ($CD4^+$) T cells. However, 'cross-presentation' of antigen from either the extracellular or intracellular pathway to the other pathway can also occur, i.e. both intracellular and extracellular antigens can be presented on either MHC class 1 or MHC class II (see later).

Each APC can present hundreds of thousands of MHC–peptide complexes on its surface. In addition, some non-protein lipid antigens, such as those in mycobacterial cell walls, can be presented atypically on MHC class I-like molecules, the CD1 receptor. Most recently, MAIT cells which are mostly found in the gut and comprise a substantial proportion of circulating T cells, also recognize antigens in an unconventional manner (see earlier). Most of the antigens

expressed on an APC surface are self-antigens, which promote tolerance (see p. 443). Immunogenic peptides occur at a level of 10–100 copies per cell.

Processing by APCs is under tight cellular control

Intracellular antigens that are likely to be processed and bound to MHC class I mostly comprise host cytoplasmic and nuclear proteins, digested during normal 'housekeeping' cellular repair work mediated through autophagy. These do not normally induce immunogenic responses. However, intracellular foreign material such as viruses in infected cells is similarly processed via an ATP-dependent structure (the proteasome) and bound to the MHC class I molecule in the endoplasmic reticulum. The proteasome is a constituent of all cells and accepts ubiquitinylated peptides and proteins for degradation (see Chapter 4, eFig. 4.5).

Further information on Antigen processing is available online at https://expertconsult.inkling.com

The CD8 and CD4 molecules help to stabilize the TCR:MHC peptide interaction in the MHC class I and MHC class II molecules, respectively. Defects in antigen processing can be associated with disease: for instance, uveitis in certain patients with the joint disease ankylosing spondylitis and severe rheumatoid arthritis is linked to abnormality in the genes coding for the proteasome and in *LMP-2* genes involved in MHC class I processing. In addition, there are specific genes (*TAP1* and *Tap2*) that generate tapasin proteins for editing peptides for optimal binding to the MHC binding site where they sit in the groove.

Further informaiton on peptide editing during antigen presentation is available online at https://expertconsult.inkling.com

Although the processing and presentation of intracellular and extracellular antigens have many similarities in detail, there are some fundamental differences. For instance, the term APC refers to cells that express high levels of MHC class II antigen and 'professionally' process extracellular antigen. The term is thus restricted to a few cell types such as macrophages, B cells and DCs. In contrast, the presentation of intracellular antigen to $CD8^+$ T cells on MHC class I has until recently been more difficult to explain. Most cells of the body express low levels of MHC class I although this can be upregulated in cells exposed to cytokines such as IL-1 and IFN-γ. However, effective antigen presentation requires additional co-stimulatory molecules, which are not normally present on healthy parenchymal cells. When antigen is presented in the absence of these molecules, the effect is to tolerize the T cell to the antigen, i.e. to render the T cell non-reactive.

This is fine for self-antigens but defeats the purpose when intracellular foreign antigen has to be dealt with. Ideally, foreign antigen should be presented when co-stimulatory molecules are also expressed in order to ensure an immunogenic, pro-inflammatory response, which will activate the T cells. This is achieved by cross-presentation of antigen by the APC. Tissue cells infected by virus (or other pathogens) are killed and APCs, especially DCs, phagocytose dead or dying infected tissue cells; the viral or bacterial antigens are then processed within the DCs for presentation on their cell surface MHC class I antigen together with co-stimulatory molecules (see later). Receptors involved in uptake of dead and dying cells (apoptotic cells) include $\alpha_V\beta_5$ integrin, CD91 and CD36 (scavenger receptors), CR3 (if coated with complement) and phosphatidylserine (see Chapter 4, p. 175 and Fig. 4.9, apoptosis pathways). Cross-presentation involves degradation of polypeptides in the proteasome (Fig. 7.26) and may be a major route for activation of CD8 T cells and activation of CD4 T cells on MHC class II antigens. In this way, $CD8^+$ cytotoxic T cells can be directly activated without requiring T-cell help and are therefore armed to induce killing in other infected tissue cells. Thus, DC can deal with both extracellular and intracellular foreign antigen and the response to self-antigen is minimized.

The antigen presentable: the MHC molecule as candy wrapper

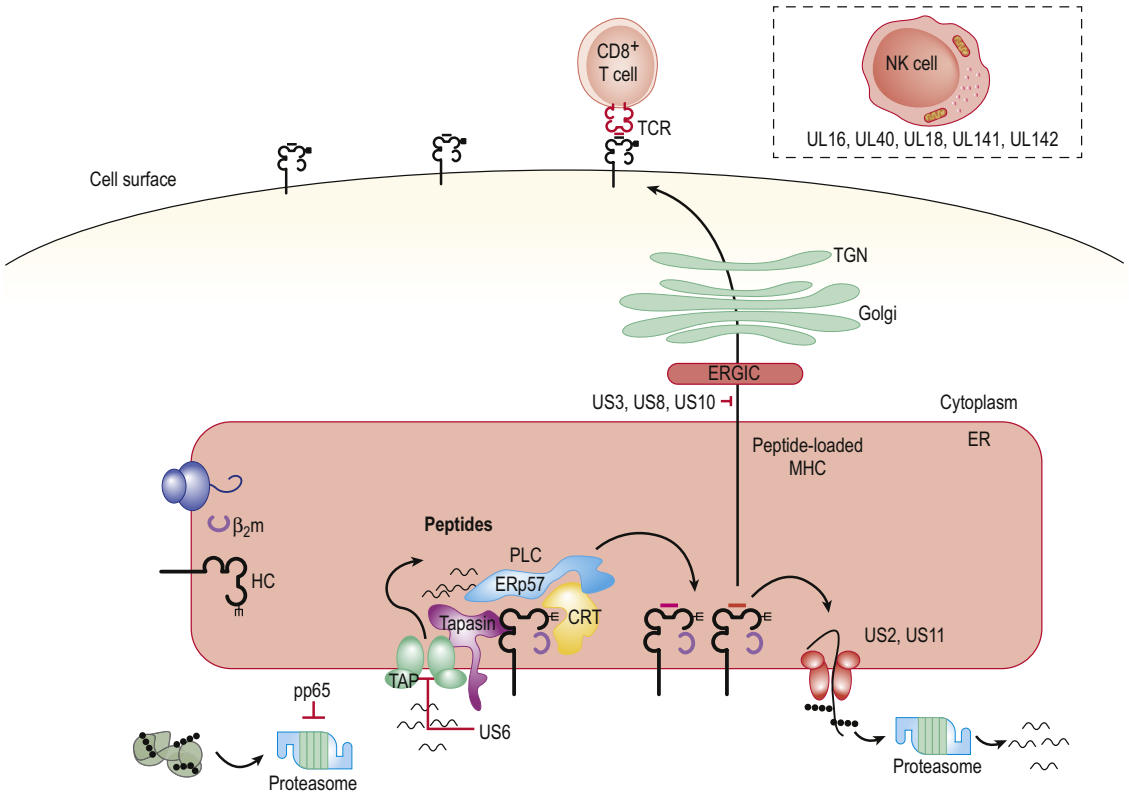
Presentation of the peptide by the MHC molecule to the TCR requires binding of the peptide to the groove in the MHC molecule and has some degree of specificity. Thus, only certain peptides will bind to each MHC allotype, of which there are many (see later). The specificity of the interaction between the MHC molecule and the peptide, however, is orders of magnitude less than that for the TCR, which is in turn considerably less specific than antigen–antibody binding. Preferential binding of certain peptides to different MHC molecules, however, is well established both for peptide length and sequence.

Further information on peptide binding to MHC antigens is available online at https://expertconsult.inkling.com

Differences also occur in binding of the TCR to MHC class I and II molecules. For instance, only a single MHC class I complex is required to present peptide to the $CD8^+$ T cell but two or more MHC class II–peptide complexes combine with two or more TCRs to activate $CD4^+$ T cells in the immunological synapse (see later). Binding of peptide to the TCR has many resemblances to antigen–antibody binding in terms of specificity but peptide-binding affinities vary greatly for the TCR (see section on T-cell activation, p. 431). In the SLO, T cells repeatedly 'sample' peptides presented on MHC molecules on APCs by making

ANTIGEN PROCESSING

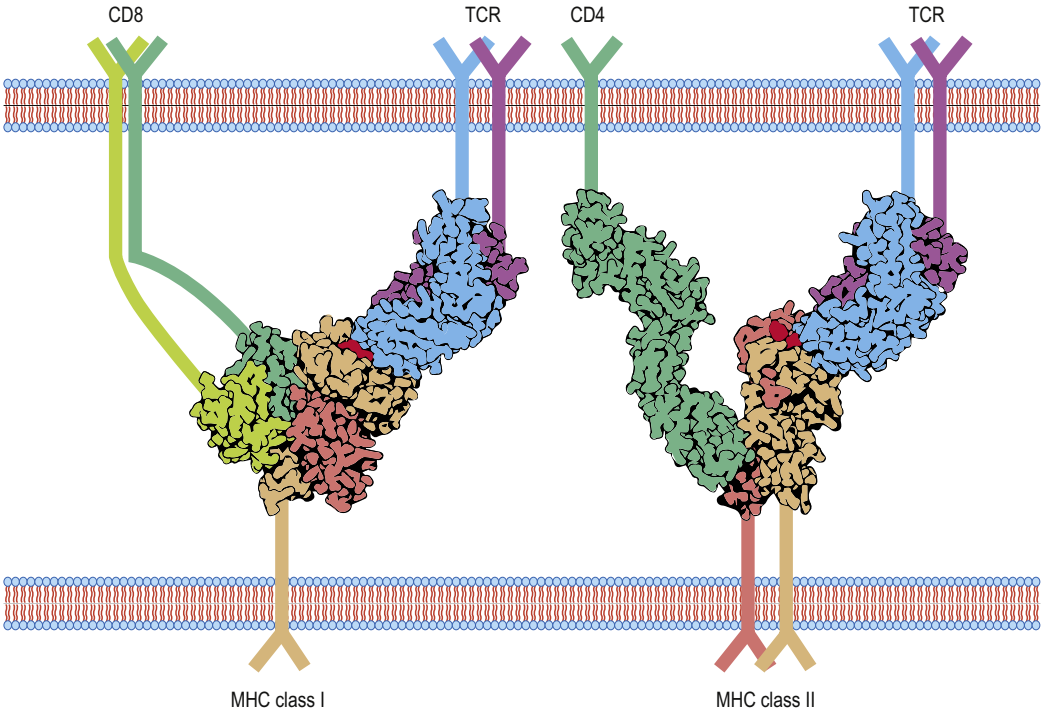
The proteasome is a general protein processing machine, which degrades exogenous and endogenous molecules to peptides and transfers them to other organelles for further processing in lysosomes and finally to amino acids (see Chapter 4, p. 170, Box 4.2). In antigen-presenting cells, the proteasome prepares peptides for presentation via MHC molecules to T cells but this process can be disrupted, for instance by infectious organisms such as cytomegalovirus (CMV), which produces proteins that prevent proper preparation of the peptide. This can occur at several stages, for instance by the viral proteins pp65 and US3-10. These act at various stages either prior to entry of the peptide into the endoplasmic reticulum or before entry of the composite MHC–peptide ligand into the secretory machinery of the cell for ultimate expression on the cell surface (eFig. 7.9). APCs, the immunoproteasome, are specialized to degrade antigens into a set of overlapping peptides. For this purpose, it contains two MHC-related proteins, LMP-1 and LMP-2. The peptide is bound specifically to a groove formed by the α_1 and α_2 helical domains of the MHC class I molecule and is anchored at specific sites to the β-pleated sheets that form the floor of the groove (eFigs 7.9 and 7.10).



eFig. 7.9 The proteasome (see Chapter 4 Box 4.2, p. 170) is a 'peptide-processing factory', which degrades proteins by ubiquitination and prepares the peptides for incorporation onto the major histocompatibility complex (*MHC*) molecules for ultimate presentation on the cell surface. CMV interferes with class I MHC antigen presentation by generating 'immunoevasins' to block CD8 T-cell cytotoxicity. The immunoevasins are the U-labelled proteins and pp65. *CRT*, Calreticulin; *PLC*, peptide-loading complex; *TAP*, transporter associated with antigen presentation; *TCR*, T-cell receptor.

Continued

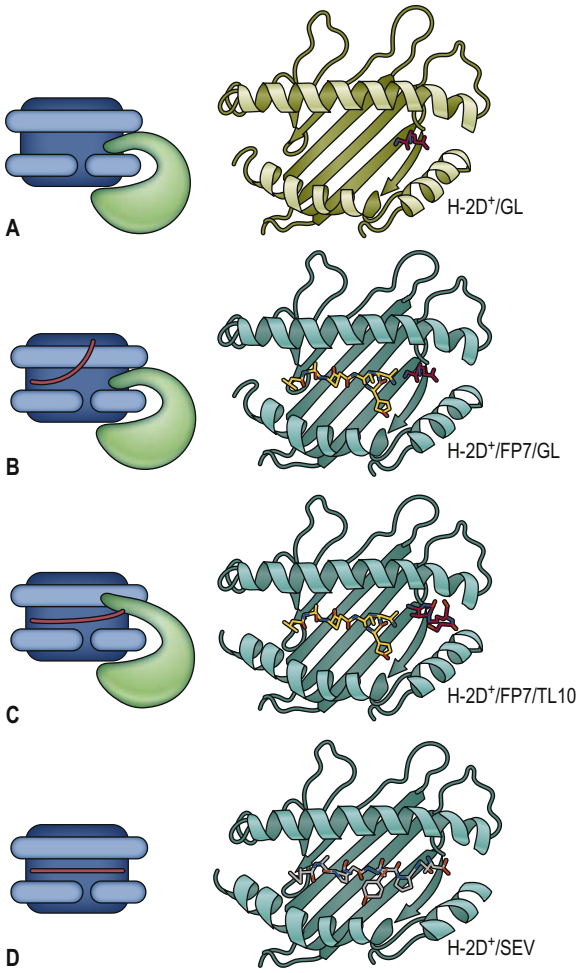
ANTIGEN PROCESSING—cont'd



eFig. 7.10 Models of antigen presentation. The antigen-presenting cell membrane (lower *grey band*) houses the major histocompatibility complex *(MHC)* molecules and presents peptide *(small red shapes)* embedded in the MHC 'groove' to the T cell: the composite MHC–peptide ligand is bound by the T cell receptor *(TCR)* and the entire structure is stabilized by CD8 molecules on cytotoxic CD8 T cells and CD4 molecules on CD4 helper T cells. (Image from http://www.rcsb.org/pdb/101/motm.do?momID=63 Courtesy the World Wide Protein Data bank.)

PEPTIDE EDITING DURING ANTIGEN PRESENTATION

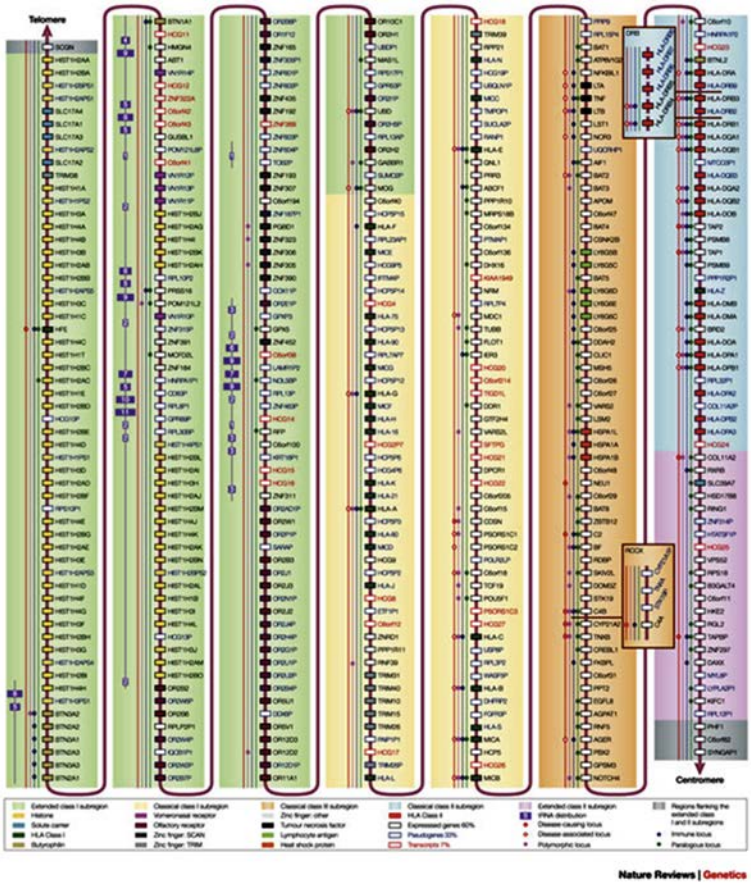
Each step in the antigen-processing pathway (see p. 428) is essential in ensuring correct presentation of normal non-immunogenic host peptides rather than peptides that may induce inflammatory disease. This applies particularly to genes that are involved at specific stages in the process. The peptide-loading complex comprises the transporter associated with antigen presentation in the ER (*TAP 1 and 2 genes*), tapasin, a chaperone protein involved in editing the peptide, a tapasin related protein (TAPBR) that does some preliminary editing, the oxidoreductase (ERp57) and the chaperone calreticulin. In addition, the ER aminopeptidases ERAP1 and 2 and the glucosyl transferase (UGGT) all collaborate to load the correct peptide into the MHC Class I groove (eFigs 7.9 and 7.11). Polymorphisms in ERAP1 and ERAP2 are linked to diseases such as ankylosing spondylitis and its associated eye disease (acute anterior uveitis).



eFig. 7.11 Hypothetical molecular model underlying tapasin-mediated MHC-I peptide selection and binding. (**A**) Tapasin is considered to have scoop loop *(dark green)*, which stabilizes the pMHC complex. (**B** and **C**) The candidate peptide binds to the MHC-I cleft first with its N terminus and can be released again if it displays insufficient affinity. If affinity is sufficient, the C-terminal of the candidate peptide is tested for its capacity to dislodge the scoop loop, which if successful (**D**) allows tapasin to release its grip from the cleft and a full-length complex is formed. (From Hafstrand, 2019.)

Continued

PEPTIDE EDITING DURING ANTIGEN PRESENTATION—cont'd



eFig. 7.12 A map of the extended MHC (xMHC) gene region in humans. The different regions are colour coded as in the key. Some specific regions are expanded in the inset boxes. (From Horton et al., 2004.)

PEPTIDE BINDING TO THE MHC ANTIGEN

The size of the peptide is much more restrictive for MHC class I than for MHC class II. Class I molecules will only bind peptides that are 9 or 10 amino acids long, but class II molecules will bind those of any size, but usually between 16 mer and 30 mer. Class I peptides are anchored to a deep pocket in the groove at the second residue. The C-terminal end of the peptide is also bound to a shallower pocket, normally at the ninth residue. This leaves the peptide essentially free in the middle, apart from some less strong side-pocket binding; as a result, the peptide is usually arched in the middle with its amino acid residues projecting outwards to the TCR (see eFig. 7.10). It is these exposed residues that determine the specificity of the reaction with the TCR.

In contrast, peptides in the class II groove overlie the sides of the molecule. Binding to the class II molecule, however, is restricted to the same number of residues as for class I (i.e. 9 or 10), except that they occupy the central portion of the peptide. Anchoring at the second residue is the strongest, frequently mediated by the non-polar residue proline. A further difference is that anchoring of the peptide to the class II molecule occurs at positions 2, 4, 6 and 9, with greater side-pocket interaction. This has the effect of straightening the peptide to take up a twisted, linear conformation.

These differences in class I and II peptide binding can be accounted for by specific amino acid sequences. In addition, the affinity of a peptide for a particular MHC allele is dependent on this sequence and is determined by how well each peptide fits into the pocket in the groove. Clearly, peptides composed of amino acids with large or highly charged side chains will have different requirements for binding than those with smaller amino acids, which might fit easily into the pocket.

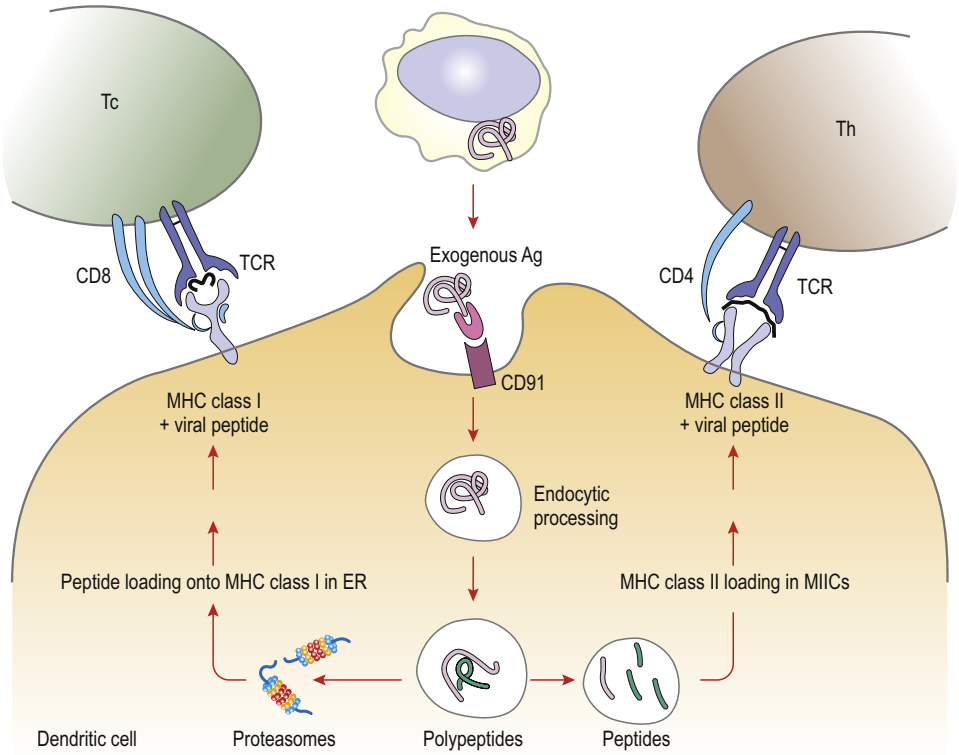


Fig. 7.26 Cross-presentation. In the *upper centre* of the figure an endogenously produced antigen *(Ag)* that eventually ends up in cellular debris has been captured by a scavenger receptor *(CD91)* on a dendritic cell. The antigen is internalized into the endocytic system and polypeptide fragments are produced. Some of these fragments remain in the exogenous processing pathway *(right)* and are degraded to peptides that are loaded on to MHC class II in MIICs. These peptides are presented to $CD4^+$ T cells. However, some of the polypeptide fragments are released from the endosomes into the cytosol *(left)*, where they are taken up by proteasomes and enter the endogenous processing pathway. These peptides are loaded on to MHC class I in the endoplasmic reticulum *(ER)* and are cross-presented to $CD8^+$ T cells. *MHC*, Major histocompatibility complex; *TCR*, T-cell receptor; *Th*, T helper cell. (From Mak and Saunders, 2006, with permission from Elsevier.)

brief interactions, which become increasingly prolonged as the TCR:MHC complex 'finds' a perfect fit.

In terms of the peptide, this is recognized as immunodominance. For any given antigen, only one in 2000 peptides is likely to have sufficient binding affinity to make that perfect fit and initiate an immune response. Moreover, it is unclear what represents optimal affinity of the peptide for the receptor: both too little and too much can prevent T-cell activation. This variable peptide affinity is reflected in the TCR, which can bind reciprocally many different peptides: thus, specificity is much less rigorous for the T cell than the B cell. These many factors, i.e. immunodominance of peptides, TCR degeneracy with regard to specificity and the need for repeated TCR activation as well as cross-linking more than one TCR in the immunological synapse before cell signalling can occur, all play a part in determining whether an immune response takes place and indeed what type of T cell is induced.

This has special relevance to immunotherapy and vaccine development for infectious disease and cancer. Methods based on protein crystals structure have been used for many years to predict the binding of any particular peptide-MHC complex to the TCR and thus the likelihood of T-cell activation. Computational and bioinformatics methods are now taking over in determining structure of MHC peptide complexes and the prediction of TCR binding affinity.

The major histocompatibility system

As indicated previously, most antigens initiate immune responses by being degraded inside the APC into small

peptides, which are then complexed with MHC molecules and presented on the surface of the APC to T cells. So, what are MHC molecules? MHC molecules were discovered by Peter Medawar (1956) during his studies on rejection of transplants between individuals of the same species but different genetic make-up and are now known to determine not only the susceptibility to graft rejection but also to infectious and autoimmune diseases generally.

What are MHC antigens and where are they found?

The MHC gene cluster on human chromosome 6 is a region of highly polymorphic genes whose products are expressed on a variety of cells. Class I genes are termed HLA-A, -B and -C, while those of class II are known as HLA-D. Class III genes lie between the centromeric class II genes and the telomeric class I genes. MHC genes differ from typical germline genes because they are polymorphic and responsible for some of the traits that distinguish one individual from another. Most germline genes are by definition non-polymorphic (i.e. identical) within a species. The MHC gene cluster has now been extended to include more than 500 genes (Box 7.8) and has been fully characterized.

Further information on the extended Human MHC map is available online at https://expertconsult.inkling.com

As indicated previously, MHC class I genes were discovered during studies in inbred mice of the genetic control of transplantation rejection phenomena. Many HLA class I genes are described, some with novel functions: for instance, HLA-G genes are expressed in the placenta and may have a role in protecting the fetus from attack by maternal NK cells. MHC class II genes were discovered later in analysis of mixed leucocyte reactions (MLR) in mice, a test that is the basis of tissue typing. The polymorphisms (genetic differences known as alleles) in MHC gene products that characterize each individual are the antigens responsible for inducing graft rejection, and in humans were originally discovered in pregnancy sera and blood transfusion samples by careful documentation of antibody responses for each allotype (allotype = different type of individual). Currently, direct gene sequencing techniques are used to identify specific alleles (allele = specific genetic difference belonging to one individual) and have led to subtype identification even within alleles (see p. 428).

Association of allotypes with disease has been investigated now for many years in genome-wide association studies (GWAS) (see Chapter 3, pp. 152–154) and includes the MHC genes. The long-standing association, for instance, of certain HLA-B27 haplotypes with ankylosing spondylitis and uveitis includes some of the genes involved in trimming the processed peptides for binding to the HLA-B27 peptide groove (*ERAP1* and *ERAP2*, see earlier). *ERAP1* shows haplotype susceptibility to the spondylitis component of the disease, which is increased when the uveitic component of the syndrome is present. Different MHC protein alleles differ in their ability to bind and present different peptides and this is probably the basis of each individual's susceptibility to disease. Studies of the association of disease susceptibility and specific HLA alleles have revealed many interesting linkages.

Further information on the extended human MHC map is available online at https://expertconsult.inkling.com

Most disease associations have, however, been detected in MHC class I genes. Perhaps the strongest known association of any human disease (>90%) is between HLA-A29 and a rare form of endogenous posterior uveitis, birdshot retinochoroidopathy. MHC mapping has greatly informed disease susceptibility among different population groups; and indeed, haplotype mapping has produced interesting information on the potential migration routes of *Homo sapiens* since the emergence of the species in Africa around 160,000 years ago (Box 7.10).

Organization of the MHC genes in the genome

The MHC genes are located on chromosome 6 (see Box 7.9), while the non-polymorphic component of the class I molecule β_2-microglobulin is on chromosome 15. The MHC region is very large, occupying about 4000 kilobases—as large as the entire genome of certain bacteria.

The class II genes, placed near the centromere, are commonly involved in crossing-over during meiosis. Unlike non-polymorphic genes, both alleles are expressed. Furthermore, there are two or three functioning β chain genes in class II, each of which can combine with the α chain. This allows some class II alleles to be expressed in more than one allelic form on the same cell. Thus, a heterozygous individual expresses six different class I alleles (two HLA-A, HLA-B and HLA-C from each parent) as separate molecules on each cell. In contrast, for class II, many more than six heterodimers can be inherited (commonly there are 10–20 different class II genes per cell) and this permits a large range of peptides to be bound on each cell. Furthermore, each MHC molecule can bind many different peptides with differing affinities, which alters their ability to activate T cells (see earlier).

As for other immune genes that involve recombination events such as the immunoglobulin and TCR genes, there is coordinate expression of the HLA genes, e.g. class I on chromosome 6 and β_2-microglobulin on chromosome 13 must be simultaneously activated, and equally so the class II genes during transcription of HLA-DP, -DQ and -DR. Regulation of MHC gene expression and function thus determines susceptibility to disease.

Further information on the regulation and transcription of MHC genes is available online at https://expertconsult.inkling.com

THE EXTENDED HUMAN MHC MAP

The gene map of the extended major histocompatibility complex (xMHC) is shown from telomere (left) to centromere (right) on the short arm of chromosome 6 I (eFig. 7.12). The five colour-coded sub-regions making up the xMHC span about 7.6 Mb and are defined as: the extended class I sub-region (green block; *HIST1H2AA* to *MOG*; 3.9 Mb), the classical class I sub-region (yellow block; *C6orf40* to *MICB*; 1.9 Mb), the classical class III sub-region (orange block; *PPIP9* to *NOTCH4*; 0.7 Mb), the classical class II sub-region (blue block; *C6orf10* to *HCG24*; 0.9 Mb) and the extended class II sub-region (pink block; *COL11A2* to *RPL12P1*; 0.2 Mb). Regions that flank the extended class I and II sub-regions are shown as grey blocks. Insets denote the hypervariable RCCX and DRB regions. Numbers and positions of tRNA genes are represented by indigo bars, the length of which is proportional to the gene number between other loci. Vertical lines connect the two main groupings of tRNA genes of 1.6 Mb and 0.5 Mb of the sequence (separated by 0.6 Mb). Circles to the left of each locus indicate disease status, polymorphism, immune status and paralogy as described in the text.

This gene map has been developed to determine the vast number of allelic variations which occur in HLA's and is goes under the name of the immune Polymorphism-ImmunoGenetics HLA database (IPD-IMGT/HLA database: web site http://www.ebi.ac.uk/ipd/imgt/hla).

This extensive information has now been used to identify HLA associated polymorphisms with disease. In particular, with second generation genomic sequencing now available, the vast amount of information has required the development of new methods to analyse the data such as the HLA mapper (Castelli, et al. 2018). Standardization of a vast amount of data relating to MHC-binding peptides is also a work in progress and is publicly available via the SysteMHC Atlas project.

REGULATION AND TRANSCRIPTION OF THE MHC GENES

Regulatory elements that lie 5′ upstream of the MHC class I β_2-microglobulin and MHC class II genes, plus other essential genes such as the invariant chain, tapasin genes, *ERAP1* and *ERAP2* genes and several others, coordinate the expression of MHC genes. These elements include the S, X, Y box and the complex is known as the class II transactivator (CIITA) gene, which regulates MHC class I and II. Mutations in these genes can cause immunodeficiency syndromes.

Cytokines modulate the rate of constitutive transcription and expression of MHC genes. Class I expression is increased by α, β and γ interferons and induction varies with cell type. IFN-γ also alters transcription of class II: in macrophages, endothelial cells and some parenchymal cells the change is upwards, but in B cells it is decreased. In contrast, IL-4 has the opposite effect on class II expression of B cells. Some cells, such as neuronal cells, do not respond at all to IFN-γ. The interactive responses of immune and tissue cells profoundly affect the final nature and character of the immune response and can convert a tissue-destructive process to a protective response or, depending on the tissue, to an allergic response (see later).

The genetic control of MHC expression and function is also controlled by other genes such as the *TAP1* and *TAP2* genes and the *TAPBR* gene (see p. 428), which regulate the entry of peptides into the endoplasmic reticulum and edit them for binding to the MHC, the invariant chain gene that regulates stability of the MHC complex, the *ERAP* genes that perform final trimming of the peptide in the endoplasmic reticulum and the calnexin/calreticulin genes that regulate the binding of the peptide to MHC class I.

Box 7.9 **Regulation of complement activation**

Complement proteins comprise around 15% of total serum proteins, much of which is synthesized in the liver, although some local production occurs in the tissues. Inhibitory activities exist for both fluid phase-activated complement and for cell surface-bound complement. Regulation occurs at all stages of activation.

Soluble inhibitors include: factor H, factor HL1 and properdin (alternative pathway); carboxypeptidase N, C4-binding protein, C1q and C1INH (classical pathway); CFHR1, clusterin and vitronectin (terminal complex formation). Surface-bound inhibitors include: CR1, CR2, CR3, CR4, CR-Ig, CD46, CD55 and CD59. Most of these are expressed on leucocytes, particularly myeloid cells, while some, such as CD55 and 59, are expressed on some tissue cells such as renal and retinal pigment epithelium. CD59 is particularly widely expressed.

Some receptors for complement effector proteins also modulate the overall response, such as C3aR and C5aR, C5L2, C1qR and SIGNR (the last on dendritic cells and microglia). Cell surface sialic acid preferentially binds factor H to factor B on material such as mucin, important on many mucosal surfaces including the conjunctiva.

T-cell activation

Any molecule, large or small, can act as an antigen but only macromolecules can activate lymphocytes and act as immunogens, i.e. produce an immune response. There is a minimum size of peptide for an effective response, which for MHC class I responses is precisely nine amino acids and between 12 and 30 for class II responses. Small molecules may sometimes activate lymphocytes if they are bound to a larger molecule; in this situation they are known as a hapten. The precise interaction between the T-cell receptor (TCR) and the MHC–peptide complex is under the control of genes regulating both sides of this interaction. The TCR genes have a greater role in determining the specificity of the response than the MHC genes.

The TCR and antigen binding

Processing of antigens to peptides in the APC potentially yields a large number of immunogenic peptides. However, only a small number of these peptides will activate lymphocytes and only one will be specific for a particular lymphocyte. This is determined partly by the nature of the antigen but more importantly by the TCR repertoire, i.e. the range of different TCRs that exist to deal with very large numbers of possible antigens. The immune system has developed a mechanism to deal with this problem, which involves the use of multiple germline genes, each of which undergoes somatic rearrangement on challenge with antigen. The same basic mechanism is used by B cells in their production of antigen-specific antibodies (see later). This is known as genetic recombination and is under the control of enzymes (recombinases), products of specific genes known as recombination activation genes (*RAG-1* and *RAG-2*), which are expressed only in lymphocytes.

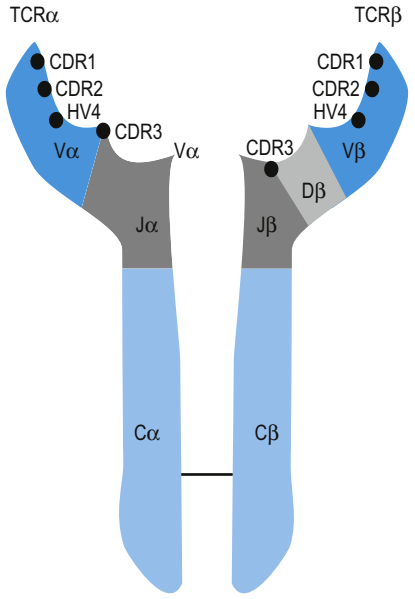


Fig. 7.27 Correspondence of TCR hypervariable regions to TCR gene segments. In this schematic example, the areas of the *TCRα* and *TCRβ* proteins derived from the indicated gene segments are shown in different colours. The *CDR1*, *CDR2* and *HV4* hypervariable regions are clustered in the variable domains of each chain, while the *CDR3* region encompasses the *VJ* joint in the TCRα chain and the *DJ* joint in the TCRβ chain. (From Mak and Saunders, 2006, with permission from Elsevier.)

TCRs are members of the immunoglobulin gene superfamily (Box 7.11) and are dimeric proteins composed of an α and a β chain that possess V (variable), D (diverse) and C (constant) regions as for immunoglobulin molecules (Figs 7.27 and 7.28). The V region contains the antigen-binding site and interacts with the peptide–MHC complex. At least 20 families of Vα chains are recognized and a similar number of Vβ families are also known. Several V genes (2–10) exist within each family. A single antigen binds to a single TCR and stimulates antigen-specific clonal expansion. However, as indicated above, several antigens have the potential to bind to the same TCR, each with different affinities, unlike antigen–antibody interactions. Precise

structural information on the conformation of a specific TCR *in vivo* has been difficult to obtain because of obscuration of the configuration by associated molecules involved in TCR activation. However, databases holding germline gene sequences as a reference resource have been set up, such as the International Immunogenetics information system (IMGT) linked through the *Ensembl* database.

Further information on germline gene databases is available online at https://expertconsult.inkling.com

At present there are many allelic variations in all four gene segments, C, D, J and V, for each of the α, β, γ and δ TCR chains, totalling over 200 genes. There are many thousands of possible combinations of the V, J and C genes, which increases the diversity of TCRαβ-binding capacity enormously. A single V gene encodes a sequence in the TCR, a complementarity determining region (CDR) (Fig. 7.27), which binds to that region on the antigenic protein corresponding to the peptide–MHC complex that binds to the TCR. This region of the antigenic protein is known as an epitope. Fig. 7.28 shows the regions of the peptide as it sits in the MHC groove that bind to the CDRs of the TCR. Each protein may contain several epitopes that interact with different TCRs and produce different responses *in vivo* (similar epitopes exist for CDRs on antibodies; see later). Some epitopes may even induce anergy (non-responsiveness) in a specific clone of T cells. TCR gene loci to a wide range of antigens from many species can be identified using the database programmes.

Further information on germline gene databases are available online at https://expertconsult.inkling.com

Immunization with a multideterminant antigen (i.e. an antigen that has several epitopes, which accounts for most protein antigens) will therefore lead to a polyclonal T-cell response in which several T-cell clones are variably activated by each determinant depending on its immunogenicity. Usually one or two epitopes on a molecule are immunodominant.

Epitope mapping of proteins is therefore possible. It has been shown that only about 30% of most proteins contain sequences that are recognized by lymphocytes. For certain autoantigens, such as retinal S antigen (implicated in the pathogenesis of uveitis, see later), various parts of the molecule have been mapped as regions that engage in physiological function such as binding activated rhodopsin (see Chapter 4, p. 265), regions, which bind antibodies, regions that are immunogenic (stimulates T cells *in vitro*) and regions that are pathogenic *in vivo* (for example, peptide sequences with the amino acids 280 to 364 induce autoimmune uveoretinitis in experimental animals) (Fig. 7.29). Cryptic epitopes, i.e. epitopes that are immunogenic but not normally detected by the immune system, are revealed only when the molecule is partly degraded.

Conversely, preferential usage of certain clones of T cells (with particular Vα and/or Vβ sequences) is recognized in certain diseases, such as rheumatoid arthritis and multiple sclerosis, and has also been described in non-infectious uveitis in humans, where the autoantigen is suspected but not known. Varicella zoster and *toxoplasma*-specific Vα and Vβ TCRs have been identified in ocular samples from patients with infectious uveitis. Thus, investigation of TCR sequences on these cells may provide clues to the nature of possible autoantigens.

Regulation of TCR gene expression is under the control of several gene enhancers such as activating transcription factor (ATF), cAMP-responsive element binding protein (CREB), T-cell factor (Tcf-1) and lymphoid enhancer factor (Lef-1) as well as specific transcription factors for each T-cell subset. For instance, GATA-3 regulates Th2 cells, while FoxP3 regulates Treg cells. Transcription factors specific for each T-cell subset are shown in Fig. 7.23.

The right time and the right place for T-cell activation

One of the long-standing conundrums in T-cell activation has been how to explain the logistics of the interaction between a T cell and an APC. As discussed previously, antigen traffics either as soluble antigen in the lymph or as cell-bound peptides in DCs to the secondary lymphoid tissues (SLOs), usually the peripheral draining lymph node. *In vivo* two-photon studies have shown that T cells entering the lymph node make repeated short interactions with several APCs and when a T cell decides that one particular antigenic peptide is the specific antigen for that receptor (based on its affinity for binding, i.e. how strongly it is bound), it engages in sustained contact (several hours to 2 days) until, when sufficiently activated (as determined by expression of activation markers such as CD69), it detaches from the APC. If this process takes place with the support of additional molecules that provide co-stimulation (co-stimulatory molecules), the T cell will undergo intense activation, proliferate (clonal expansion) and secrete cytokines. In the case of CD4 T cells, they become pathogenic effector cells secreting INFg (Th1) cells, IL17 (Th17 cells) or IL4 and IL13 (Th2 cells) under control of DC cytokines namely IL12, IL4 and TGFb/IL6 respectively (Fig. 7.22). If there is no co-stimulation, the T cell may proliferate to a lesser degree but does not become active for effector function but for tolerance and becomes a Treg secreting IL10 under control of TGFb alone. Once the clone of T cells has expanded sufficiently, they leave (egress) the lymph node essentially *en masse* under the control of a specific receptor ligand interaction (sphingosine-1 receptor) to enter the bloodstream. As more cell-associated antigen arrives from the tissues to activate further clones of T-cells, T-cell egress from the lymph node occurs in waves. As indicated previously (see p. 421), unactivated passenger naive T cells and possibly tolerized T cells pass through the lymph node via the efferent lymphatics to the next node in the chain

DATABASE FOR GENE BUILDING USING GERMLINE SEQUENCES

Information on specific TCR and immunoglobulin gene segment data can be obtained through the http://www.ensembl.org/index.html database which gives full instructions on how to use the protein 'gene build' programs. Gene building for TCRs and immunoglobulins is somewhat specialized due to the recombination events which take place between specific gene segments. This can be investigated using the link to the International Immunogenetics information system (IMGT) (http://www.imgt.org/), which can identify the locus on the genome showing the gene segment frequency for the specific Ig sequence of interest, e.g. the human IgH (immunoglobulin heavy chain) locus with specificity for herpes simplex virus type 1 (as of July 2013 there were 4 IGHV group, 4 IGHJ group and 4 IGHD group TCR gene segments on the IgH locus).

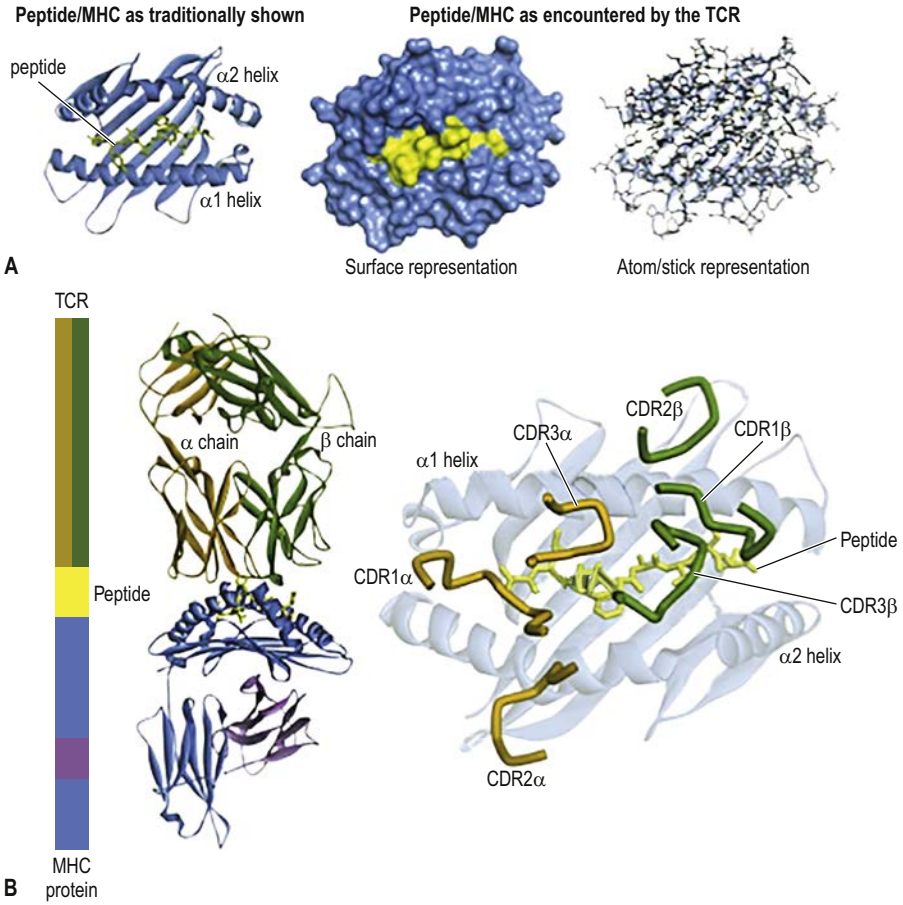


Fig. 7.28 Structural overview of how T-cell receptors *(TCRs)* engage peptide/major histocompatibility complexes *(MHC)*. (**A**) View of the composite/MHC ligand, illustrating a complex with a Class I MHC complex. The *left panel* shows a traditional image clearly demarcating the peptide (stick format) from the MHC (ribbon format). The *centre panel* shows a surface representation. The small exposed surface area of the peptide compared with that of the MHC is apparent, highlighting the close packing and proximity between a short peptide and the MHC. The *right panel* better illustrates what the TCR physically encounters: a collection of atoms in which the physical distinction between peptide and MHC is not so obvious. (**B**) How a TCR traditionally binds a peptide/MHC complex. The *left panel* shows a side view of a TCR-peptide/MHC complex, illustrating the positioning of the receptor over the ligand. The *right panel* shows the orientation of the CDR loops over the peptide/MHC. This image illustrates the DMF5 TCR binding the MART-1 antigen presented by HLA A-2. A wide range of positions of complementarity-determining region *(CDR)* loops over the peptide/MHC have been observed. (From Riley et al., 2018.)

and on to the thoracic duct thereby re-entering the circulation and thus able to circulate through the spleen. Similar induction of CD8 cytotoxic T-cell subsets takes place under control of IL2, IL15 and IL7.

Circulating activated effector T cells make the rounds many hundreds of times before entering the site of tissue inflammation since the blood vessels at these sites need to be 'softened up', i.e. start to express appropriate adhesion molecules and chemokines. Once the T cells enter the tissues, they are probably further activated by engaging with antigen or they may simply continue on their proliferative programme, secrete cytokines, kill infected cells via MHC class I mechanisms (Tc cells, Th killer cells) and recruit inflammatory monocytes/macrophages (M1-like) to clear infected tissues (macrophages), and later alternatively activated macrophages (M2-like) to restore tissue integrity (see https://expertconsult.inkling.com).

Macrophages are not all the same: Further information is available online at https://expertconsult.inkling.com

T cells entering the tissues during the later stages have been programmed as Tregs and probably function *in situ* to promote resolution of the inflammation, but it is

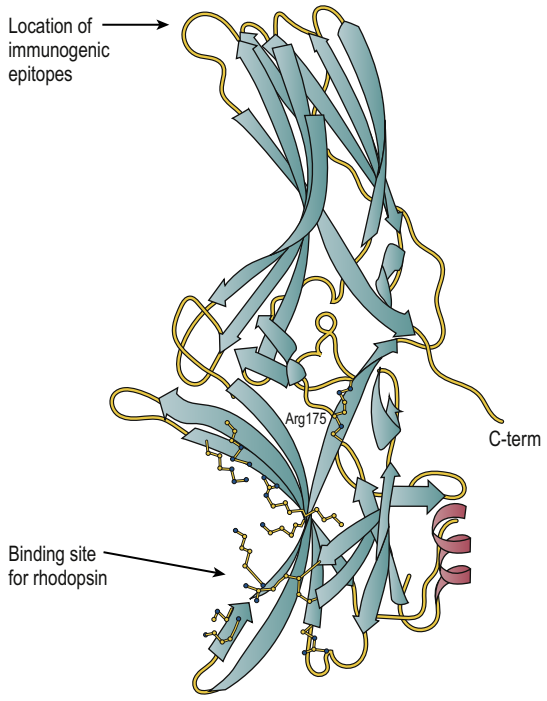


Fig. 7.29 Ribbon diagram of Retinal S antigen with b-strands in *blue*, a-helix and coil in *yellow*. *Top*: C-terminal domain. *Bottom*: N-terminal domain. The positively charged amino acids are drawn as ball and stick model (Lys 2, Lys 15, Arg 18, Arg 56, Arg 66, Arg 80, Lys 150, Lys 163, Lys 166, Arg 171, Arg 175). They are possibly in contact with the cytoplasmic phosphorylated loop of P-Rh*. The binding site for rhodopsin and the immunogenic epitopes on S-antigen are at opposite ends of the molecule. (From Granzin, 1988.)

also likely that Treg activity in the lymph node is equally important by preventing further activation/proliferation of T cells. This question has not yet been completely answered nor has the question of whether antigen needs to be completely cleared from the tissues before inflammation can subside and how this plays out probably determines whether the inflammation persists as chronic disease or whether the infection becomes latent (i.e. ignored by the immune system).

What are γδT cells and what is their role?

γδ T cells are T cells that express γ and δ IgG superfamily chains, almost identical to α/β chains of conventional T cells, but they have very limited diversity and recognize a broad range of antigens including self- and non-self-antigens such as heat-shock proteins and lipid antigens. They are, in essence, innate immune cells that become involved very early in the inflammatory process. Thus, they normally populate sites where the body is susceptible to pathogen entry such as the skin and the gut epithelium (first lines of defence) but they also circulate in the blood and respond to 'stress' induced by pathogens or endogenous stimuli (DAMPs). γδ T cells do not require classical MHC molecules to recognize antigen but they can respond to unprocessed peptide and other antigens. γδ T cells expand 2%–10% in response to alkylamines derived from microbes and edible plants and may represent a link between innate and acquired immunity. γδ T cells are therefore major players in surveillance during initial pathogen challenge and sense epithelial and immune cell-derived butyrophilin-like molecules during activation. γδ T cells recognize ligands such as CD1d bound to lipid and MICA (MHC Class I-related protein A) and are also important in tumour immunity.

γδ T cells are also implicated in autoimmunity; they appear to recognize heat-shock proteins (proteins expressed in 'stressed cells' and highly conserved across species) and it has been suggested that they play a role in autoimmune diseases such as rheumatoid arthritis and uveitis associated with Behçet's disease. γδ T cells have been reported to be essential in the induction of experimental autoimmune diabetes and are early participants in models of autoimmune uveitis. High levels of γδ T cells have also been cultured from the vitreous humour in a case of sympathetic ophthalmia. Interestingly, as part of the immune response to tumours, γδ T cells have been associated with better survival from choroidal malignant melanoma.

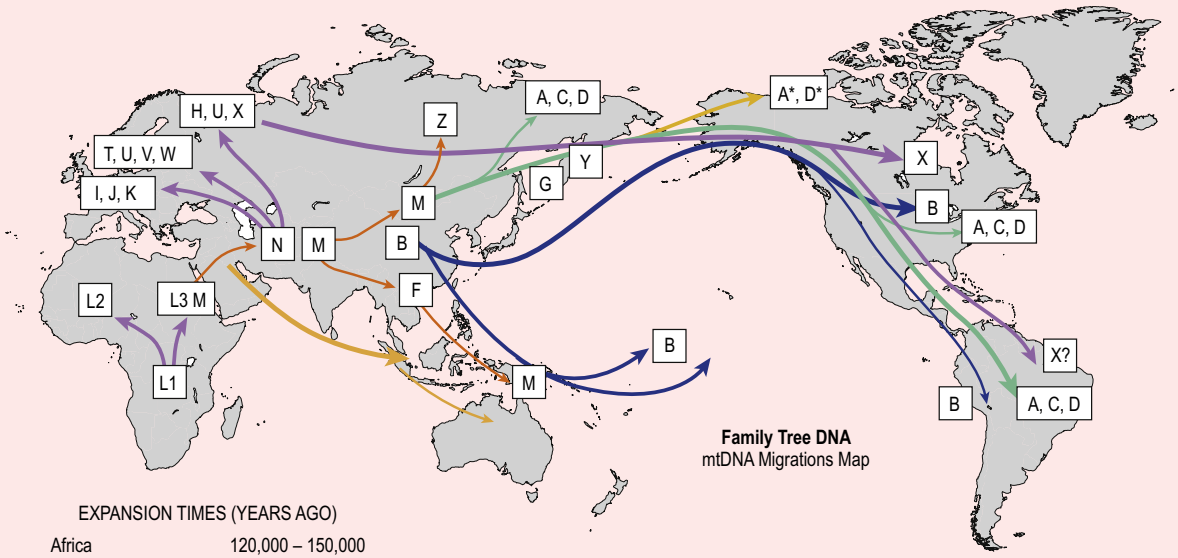
Superantigens

Some diseases induced by organisms are so rapid in their onset and catastrophic in their manifestations that it is difficult to explain their pathogenesis in terms of an adequate innate or acquired immune response. Examples include the toxic shock syndrome, meningococcal meningitis and leptospirosis. These disorders are caused by superantigens derived from bacteria (e.g. staphylococci, streptococci, mycobacteria, *Clostridia* and many other organisms that release superantigens), viruses (e.g. rabies and ebola viruses) and retroviruses (e.g. mouse mammary tumour virus and human immunodeficiency virus). Superantigens are generated by genetic elements known as 'pathogenicity islands', which are phage-related (a phage = a parasitic virus for bacteria, which uses the bacteria's metabolic machinery to replicate) genes (phage-related chromosomal islands, PCRIs), which they use for rapid transduction.

Staphylococcal enterotoxins (SE) are a good example of superantigens and are the commonest cause of food poisoning. There are more than 20 SEs, SEA–SEE. They have to bind to the MHC class II antigen to be presented to the T cell, but they can bind many polymorphic MHC class II molecules. Superantigens do not require processing by

Box 7.10 **Tracking map for mitochondrial DNA**

A haplotype is a set of allelic differences (i.e. genes) inherited from a single parent and has been used in anthropological studies to track the migration of humans over hundreds of thousands of years from the first migration of *Homo sapiens* out of Africa (see the tracking map for mitochondrial DNA; further information is available through the link http://mathildasanthropologyblog.wordpress.com/2008/06/16/mitochondrial-dna-haplotypes-for-dummies/).



EXPANSION TIMES (YEARS AGO)

Africa	120,000 – 150,000
Out of Africa	55,000 – 75,000
Asia	40,000 – 70,000
Australia/ PNG	40,000 – 60,000
Europe	35,000 – 50,000
Americas	15,000 – 35,000
Na-Dene/ Esk/ Aleuts	8,000 – 10,000

APCs because they activate the T cell by binding to the side of the TCR β chain. They are, therefore, quite promiscuous in that they can polyclonally activate several species of T cell. In superantigen infection, about 10%–20% of T cells are activated, whereas in conventional adaptive immune responses only 1 in 10^5 or fewer T cells are capable of recognizing the antigen. Indeed, few antigen-specific T cells can be detected either in the circulation or even in the lesion. The result is the very rapid production of cytokines from activated T cells as well as T-cell-induced macrophages, causing a 'cytokine storm'. Superantigens are the cause of toxic shock syndrome.

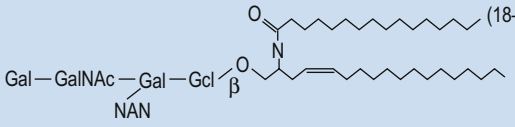
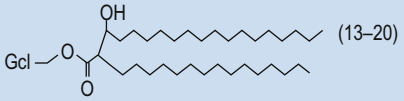
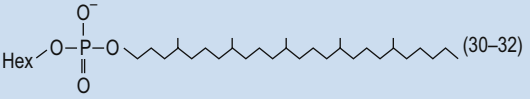
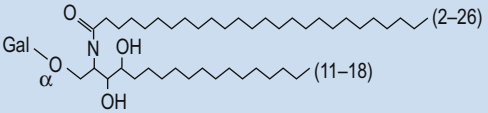
Polyclonal activation in superantigen infection may involve autoreactive T cells and thereby initiate an autoimmune disease, which may persist even if recovery from the infection occurs. Studies of preferential Vβ TCR gene usage by lymphocytes infiltrating the tissues may shed light on the nature of the autoantigens in conditions such as uveitis and rheumatoid arthritis and help to identify sequence similarities between autoantigens and superantigens.

Antigens can be presented by other MHC and MHC-like molecules

Antigens presented through the canonical MHC class I and II routes are exclusively small peptides. However, other molecules such as sugars and lipids can evoke T-cell and antibody responses. MHC class Ib molecules include human leucocyte antigen (HLA) -E, -F, -G (see Box 7.10) and in the mouse molecules such as Qa-1. Although these molecules have some diversity, the range of peptides they can present is restricted, e.g. Qa-1 presents only a small set of microbial peptides, while HLA-G is involved in maternofetal antigen presentation.

The MHC-like molecule CD1, which does not contain an α_2-microglobulin protein, has a highly hydrophobic

A

CD1	Lipid antigen	Source
CD1b	Ganglioside GM1 	Self
	Glucose monomycolates 	Mycobacteria
CD1c	Hexosyl-1-phosphoisoprenoids 	Mycobacteria
CD1d	α galactosyl ceramide 	Marine sponge

B

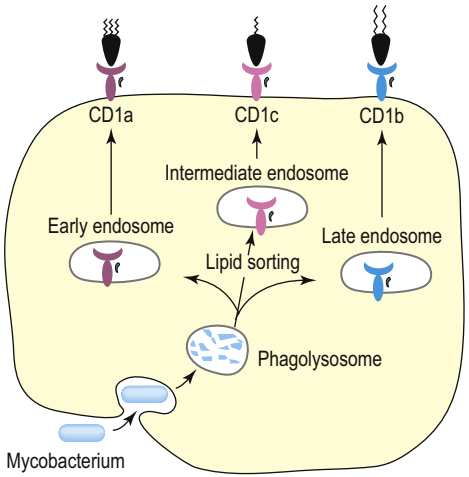


Fig. 7.30 Lipid antigen presentation by CD1 molecules. (**A**) Structures of lipid antigens presented by CD1 molecules. The precise structure of the CD1a antigen is unknown. (**B**) Sorting of mycobacterial lipids. At the *bottom left* of the figure, a mycobacterium has been phagocytosed and degraded. The mycobacterial lipids are sorted by structure into different endosomal compartments where they are loaded on to different *CD1* molecules for presentation to T cells. (From Mak and Saunders 2006, with permission from Elsevier.)

groove with two deep pockets that neatly binds the fatty acid chains of glycolipid antigens while the sugar moieties bind the TCR (Fig. 7.30). In relation to eye disease it has been suggested that both innate and adaptive immunity may be involved in AMD and that modification of self-antigens by products of lipid peroxidation such as carboxyethylpyrrole (CEP) may act as autoantigens and drive disease changes at the RPE-choroidal interface. Lipidated antigens may be presented via CD1 molecules; for instance, CD1d binds NK T cells via αGal-ceramide, while CD1c binds γδ T cells via an unknown antigen.

Co-stimulation: presentation of antigen requires 'help' from other molecules

MHC:peptide complexes, the ligands for binding to the TCR, will not activate effector T cells unless the cytokine

milieu is conducive (see Figs 7.2, 7.23 and 7.24). The cytokine milieu is itself dependent on whether the antigen is presented by an APC that has been activated by a DAMP (see p. 397). For instance, activation of a TLR or NALP3 with consequent induction of the inflammasome will release pro-inflammatory cytokines such as IL-1β, IL-12, IL-23 and TNF-α, all of which provide a suitable environment for activation of a Th1 or Th17 $CD4^+$ T cell; in contrast, if IL-10 or TGF-β are released, Tregs are more likely to be induced. In both cases, the TCR is triggered but different transcriptional pathways are induced.

In addition, activation of one TCR on a Th1 or Th17 cell is not enough: several TCRs need to be simultaneously engaged in a membrane structure known as the immunological synapse. This comprises the aggregated TCRs as well as several surrounding co-stimulatory molecules. In contrast, less is known about induction of Tregs or even engagement of naive T cells to induce anergy (i.e. non-responsiveness), but several 'co-inhibitory' molecules have now been discovered.

Co-stimulatory molecules

Activation of the T cell is achieved by the clustering of several ligand–receptor pairs between it and the APC. Clustering of these ligands at the point of contact between the cells helps to strengthen and to prolong the contact between the cells and thus facilitate the presentation of peptide to the TCR (Fig. 7.31). Mobilization of the various sets of molecules to the site of TCR–pMHC interaction occurs within localized membranous patches (lipid rafts, see Chapter 4, pp. 167–168) and the various ligand–receptor pairs participate in the formation of the immunological synapse (Fig. 7.31). Some of the molecules are involved in either T-cell co-stimulation or co-inhibition including adhesion molecules such as ICAM-1/LFA-1 (see Table 7.3), co-stimulatory molecular pairs such as CD40-CD40L, ICOS-ICSOL and CD80-CD28, as well as inhibitory molecular pairs such as PD1-PDL1, CD28-CTLA4, SIRP1a-CD47, which are now known as checkpoint inhibitors (see later) (reviewed by Chen and Shen, 2015).

Several of the molecules involved in T-cell activation are also shown in Fig. 7.31. However, two important ligand–receptor pairs that drive $CD4^+$ T-cell activation are not included in this diagram and are the B7:CD28 interaction and CD40:CD40 ligand (CD40L). Interaction between these molecules is a prerequisite for T-cell activation and, indeed, in the absence of this 'second signal' presentation of peptide to the TCR it is likely to lead to anergy rather than stimulation. B7 is a 45 kDa to 60 kDa cell surface glycoprotein expressed on activated B cells, dendritic cells and macrophages, which binds to CD28, a 44 kDa homodimeric molecule expressed constitutively on the T cell, and leads to the proliferation of, and IL-2 secretion by, T cells. B7 can also bind to a second receptor, cytotoxic T-lymphocyte antigen 4 (CTLA-4), which is slowly induced in T cells and is an inhibitory molecule with greater affinity for B7; as such this ligand pair is involved in downregulation of T-cell activation in the later stages of the response. CTLA-4 is also expressed at high level in Tregs, thus mediating part of their suppressive function. CTLA4 is a checkpoint inhibitor and when this is lifted, as in current therapy for several types of cancer, failing immune response to tumours is reactivated and the tumour is attacked.

CD40–CD40L interactions are essential to T-cell activation, and in reciprocal DC conditioning there is a two-way signalling interaction via this ligand pair. CD40 on APCs can be upregulated and is an approach to vaccination. In addition, it is present on many non-APCs such as endothelial cells. CD40L, however, is only present on activated T cells and even then transiently, so that the timing of CD40L expression in part regulates the overall duration of the immune response. CD40L and a further molecule, Lag3 (CD223), may be involved in APC 'licensing' (enabling) cytotoxic T cells.

Several other accessory molecules are involved in the T-cell/APC interaction leading to the immunological synapse to maximize the T-cell/APC contact. Activation of the T cell leads to the expression of integrin adhesion molecules on the cell surface, including LFA-1 and VLA-4 ($\alpha_4\beta_1$) (see Table 7.3). Reciprocal expression on cells promotes binding. VLA-4 also binds to extracellular matrix molecules, as does CD44—the hyaluronan receptor. Some molecules provide co-stimulation for specific T-cell activities, such as OX40 and ICOS-ICOSL and others such as PDL–PDL1 are inhibitors of T-cell activation.

The process of antigen presentation for the generation of an active immune response (immunity) therefore involves multiple molecular interactions and requires co-stimulatory activity for induction of positive adaptive immunity and co-inhibitory molecular interaction for downregulation and anergy induction. It is not clear, however, whether all of the above interactions are required to initiate T-cell clonal expansion or whether immunogenicity is a function not just of peptide/TCR specificity but also depends on which and how many of the accessory interactions are entrained.

Getting the message across: cell signalling through the immunological synapse

Cells activated through cell surface receptors alter cell function through second messenger systems in the cytoplasm. These are likewise linked to events in the nucleus that mediate protein transcription and ultimately to changes in cell function (see cell signalling, Chapter 4, p. 164). The immunological synapse (also known as the supramolecular activation complex, SMAC) provides this function

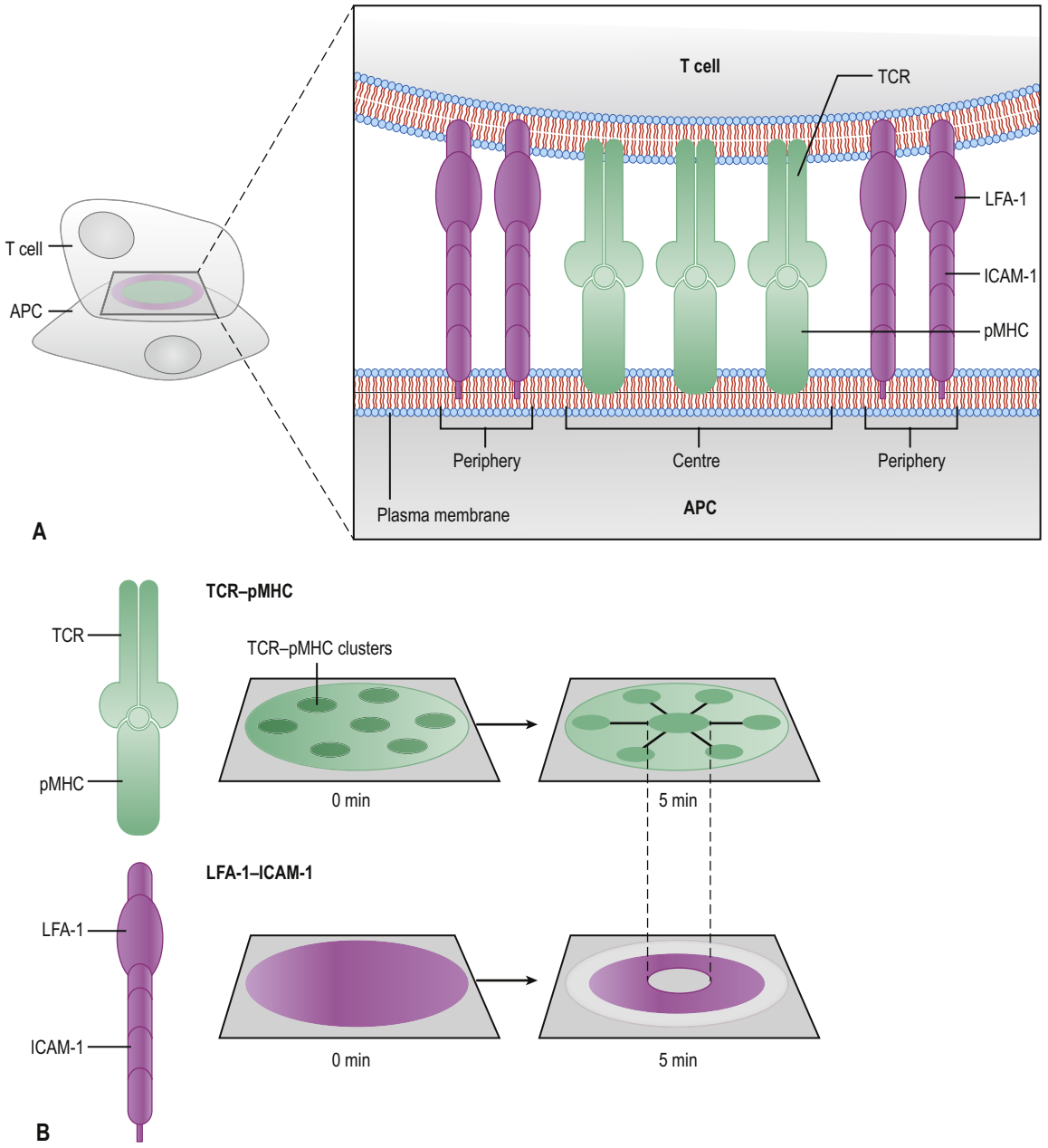


Fig. 7.31 The immunological synapse. This term has been used to describe the binding between the T cell and the antigen-presenting cell because of its structural similarity to the neurological synapse. There is a *centre* containing the ligand receptor pair (MHC peptide:TCR) surrounded by a ring of adhesion and other accessory molecules which are progressively recruited to the synapse as initial non-T-cell activating binding is converted to a firm prolonged contact necessary to activate the T cell. (**A**) A side and enhanced view of the synapse. (**B**) An *en face* view. Over a period of about 5 minutes, many TCRs are recruited to the centre of the synapse while further adhesion and accessory molecules firm up the outer ring. The synapse is generated in a membrane lipid raft (see text). *APC*, Antigen-presenting cell; *MHC*, major histocompatibility complex; *TCR*, T-cell receptor. (From Manz and Groves, 2010.)

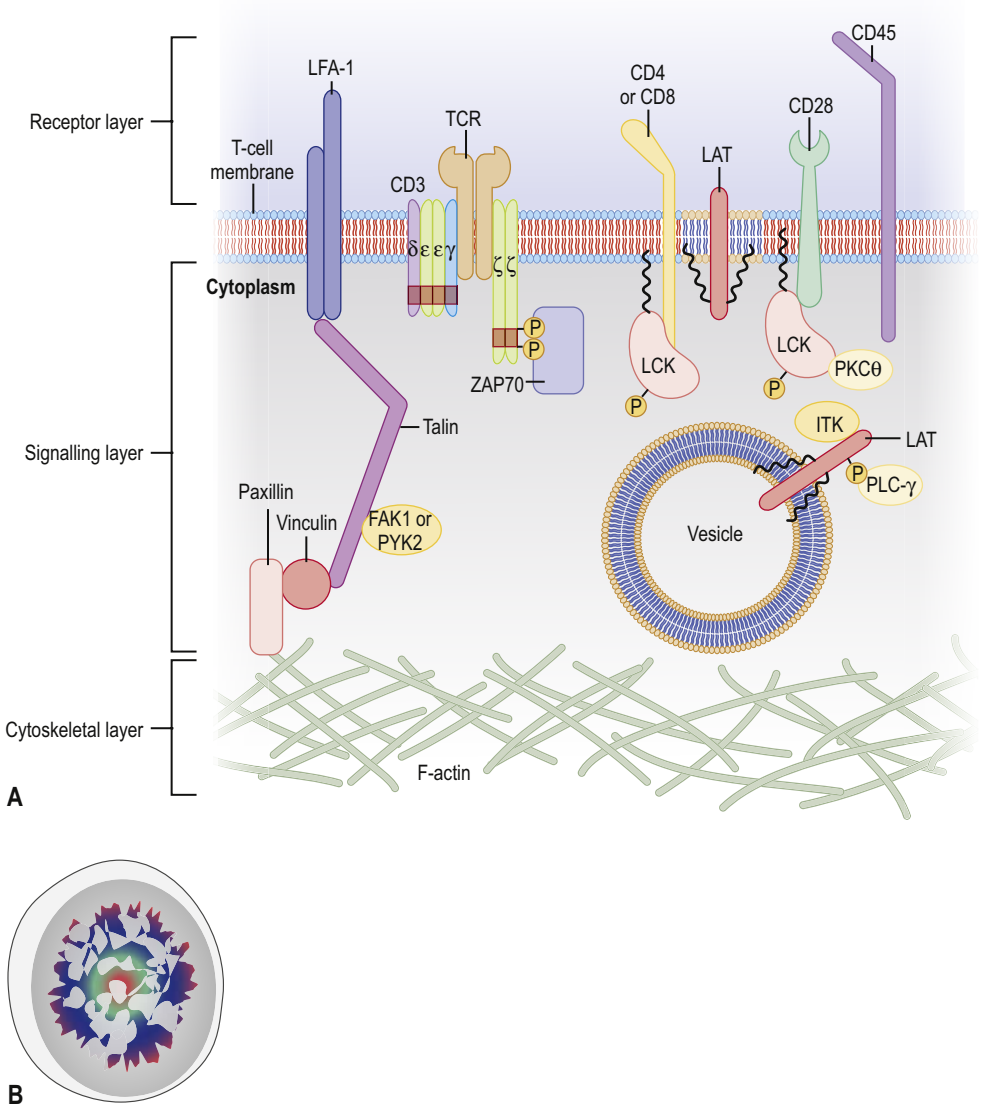


Fig. 7.32 Molecular details of the immunological synapses on the T-cell receptor side. The TCR linked to the adhesion molecule LFA-1, the CD4 T-cell-specific anchoring molecule CD4, the accessory activation molecules CD28 and CD45, activates the signalling machinery (VAV, ZAP-70, LCK, PKCθ, and LAT) as well as the T-cell cytoskeletal motor (vinculin, F-actin, etc.), which sets the T cell on a cytokine secretion programme, and migration towards chemotactic stimuli. The side view is shown in (**A**) and the *en face* view on the T cell in (**B**). (From Dustin and Depoil, 2011.)

for the T cell. The initial contact between the T cell and the APC mediated through reciprocal LFA-1 and ICAM-1 interactions and also probably by CD45, stabilized by the CD4 and CD8 molecules in the appropriate T cells, induces changes in the actin cytoskeleton that draw several molecules on both sides in lipid rafts into a stable adhesive complex (Fig. 7.32).

The effect is dependent on the signalling molecule VAV. In the centre of the synapse lies the CD3/CD4 (or CD8)/CD45/TCR–pMHC complex surrounded by an inner ring composed of CD2–LFA-3 paired molecules, itself surrounded by an outer ring of ICAM-1–LFA-1 molecules linked to talin and thus to the cytoskeleton (Fig. 7.32). The TCR is linked to its second messenger system through

a complex of transmembrane dimeric proteins, the CD3 complex, which links intracellularly to cytoplasmic vesicles via an adapter protein, linker for activation of T cells (LAT), and to a signal transduction complex, which belongs to a group of molecules known as immunoreceptor tyrosine-based activation motifs (ITAMs). The T-cell ITAM (known as $p59^{fyn}$) acts in conjunction with a second molecule ($p56^{lck}$) (LCK) as the major tyrosine phosphorylating mechanism in T cells.

The four TCR–CD3 complexes may each initiate discrete signalling events, the summation of which produces an effective stimulus for T-cell activation (similar to events in neural cells; see Chapter 5, p.315). A progressive accumulation of signal thus develops through the sustained presentation of antigen mediated by recruitment of co-stimulatory and accessory molecules into the membrane surrounding the immunological synapse. In contrast, short-lived contacts between APCs and T cells may fail to initiate synapse formation in the T cells and probably do not even involve the TCR. Downstream signalling from the Src family of tyrosine kinases is another set of protein tyrosine kinases with docking sites that bind SH-2 molecules. One of these is a protein known as ZAP-70 (associated with the ζ protein of ITAM, CD3) (Fig. 7.32), which plays a major role in sustained TCR signalling. In fact, the interactions between the various signalling molecules through the TCR are a good example of how signalling networks function.

ITAMS and ITIMS

As indicated previously, ITAMs such as CD3 and LCK, activate immune cells and ITIMs (immunoreceptor tyrosine inhibitory molecules) inhibit immune cells. ITIM pairs involving interactions between T cells include PDL-PDL-1, B7-CTLA4 and more while myeloid cells have numerous ITIMs (over 200). The characteristic feature of ITIMs is that they recruit phosphatases such as SHIP and SHP, 1,2, which remove the phosphate group from activating signalling molecules and promote inhibition of cellular function.

Action of drugs on cell activation

Certain immunosuppressive drugs, used clinically in transplantation and autoimmune diseases, are known to act at various stages in T-cell activation. These include steroids, ciclosporin A, tacrolimus, rapamycin and mycophenolate mofetil. Ciclosporin, tacrolimus and mycophenolate mofeteil are particularly relevant to ophthalmology because they are used in a variety of conditions such as sight-threatening uveitis, severe scleritis and corneal graft rejection.

Ciclosporin and tacrolimus specifically inhibit the transcription of the *IL-2* gene in CD4 T cells by binding to intracellular proteins (immunophilins) that subsequently bind a Ca^{2+}-regulated phosphatase, calcineurin. This enzyme is activated by Ca^{2+} influx during T-cell activation and is required for assembly of the nuclear transcription factor (NF-AT) involved in IL-2 secretion. Ciclosporin may also act via induction of TGF-β release, a cytokine involved both in Treg cell function and in induction of Th17 cells (see Fig. 7.2).

B-cell activation

Until now we have considered only cellular interactions with antigens. However, the first evidence for the existence of an adaptive immune system came through the discovery of antibodies, circulating proteins in the globulin fraction of the serum. Free antibodies are produced by plasma cells, the fully differentiated version of the B cell. B cells are derived from stem cell precursors in the bone marrow and are released into the circulation as $CD45^+$ $sIgM^+$ $CD19^+$ $CD20^+$ MHC class II^+ immature cells that migrate to the B-cell follicle regions of the SLOs. B cells comprise 5%–10% of the circulating lymphocyte pool. B cells are important not only as antibody producers but also as APCs in their own right.

Antigens, B cells, and T cells: what's the story?

Antigen recognition by B cells

The rules governing the size of peptide antigen required to induce an immune response apply to both T and B cells. B cells can also respond to carbohydrate and glycolipid antigens, producing T-independent B-cell responses. This occurs for instance with blood group antigens.

Early events in B-cell responses to peptide antigens require T-cell help, usually provided by Th cells, which require CD40:CD40L binding and are enhanced by cytokines such as IL2, 4, 5 and 6 produced by Th2 cells and other cells such as macrophages. Later events in B-cell maturation are promoted by cytokines such as IL17 and IL21 produced by T follicular helper cells (Tfh) under the control of the transcription factor Bach2, which targets the B-cell induced maturation protein-1 (Blimp-1) as well as SOCS, the chemokine CCR5 and the apoptotic protein Bcl-6. IL21 is implicated in a number of autoimmune disease involving antibodies such as SLE and rheumatoid arthritis. Tfh cells assist in the long-lived production of antibody by B cells.

A subset of Tfh cells also has a regulatory function (Tfr cells) and shares the transcription factor FoxP3 with Tregs. These cells have been implicated in the recently described set of diseases known as IgG4-related diseases, one of which is an orbital disease previously included as one form of pseudotumour of the orbit.

B cells acquire antigen from specialized follicular dendritic cells (fDCs), which present antigen as immune complexes bound to Fc and C3 receptors generating an activated germinal centre B cell. This cell undergoes isotype switching (see p. 407), somatic hypermutation and clonal expansion (proliferation) to produce memory B cells and long-lived plasma cells, each of which secretes a specific (monoclonal) antibody. B cells also present processed antigen on MHC class II molecules to Tfh, which reciprocally assist in the presentation of antigen to B cells by fDCs. This requires co-stimulation via CD40–CD40L interactions. This form of T-cell and B-cell recognition of the same peptide antigen is known as mature B-cell (plasma cell)-linked recognition. Other co-receptor molecules are involved in the formation of the contact site of the B–T-cell interface, e.g. CD30:CD30L and B-lymphocyte stimulator (Bly5) and its receptor TAC1.

B cells arriving from the bone marrow to the secondary lymphoid tissues traverse the T-cell area on their way to the B-cell follicles. However, if they meet an antigen-specific Tfh cell, they become trapped in the marginal T-cell zone and start proliferating and producing antibody (a primary lymphoid focus). Some of these B cells mature to antibody-producing plasma cells, which leave the T-cell area and return to the bloodstream and ultimately the bone marrow. In contrast, the majority enter the B-cell area where they form a B-cell follicle and ultimately form the B-cell germinal centre, an antibody-producing mini-factory, where extensive isotype switching takes place.

The B cells then migrate from the germinal centre to the medulla of the secondary lymphoid organ (lymph node, spleen or MALT), and out into the circulation where they home to the bone marrow and differentiate into mature antibody-secreting plasma cells. B cells may also populate sites of chronic inflammation where they develop into plasma cells. B-cell follicles are an important component of tertiary lymphoid organs (TLOs) (see p. 422) involved in local production of antibodies at sites of inflammation and indicate a worsening of disease in autoimmunity but are favourable prognostic entities in cancer. For instance, local antibody production in chronic inflammation in the eye (uveitis) is used as a biomarker of disease severity (the Goldmann-Witmer test, a measure of the ratio of intraocular vs plasma antibody levels) and is a useful diagnostic test of ocular infection in cases of uveitis, e.g. toxoplasmosis and herpes simplex uveitis.

Three checkpoints exist therefore for B-cell expansion: the first is at the stage of T-cell help, regulated by CD40L; the second is at the stage of selection in the germinal centre by the fDC delivering anti-apoptosis signals; and the third is at the stage of migration of the mature B cell (plasma cell) to bone marrow and other tissues. However, most B cells remain in the follicle/germinal centre and secrete antibody directly into the bloodstream via the efferent lymphatics and thoracic duct.

B cells as antigen-presenting cells

Endocytosis of antigen via its sIg receptor activates the B cell to process and present peptide fragments in association with MHC class II on its surface (see earlier). B cells are efficient APCs but can be activated to present antigen only via the sIg receptor. Therefore, in this regard they act as memory cells.

Although B cells are activated in an antigen-specific manner and their effector molecules (antibodies) are also antigen specific, they induce their effects in an antigen–non-specific way, predominantly via complement. B-cell responses are also greatly enhanced by activation of the B-cell co-receptor complex (CD19:CD21:CD81). CD21 is the receptor for complement component C3d, thus permitting antibody–antigen–complement interaction.

B-cell differentiation

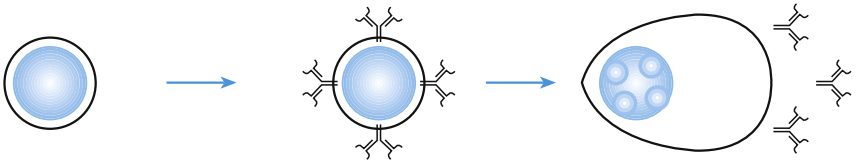
As indicated previously, B cells are derived from stem cell precursors in the bone marrow (Fig. 7.33). Immature B cells home to the SLOs. However, if they do not encounter specific antigen, they undergo apoptosis and are removed within 3–4 days. In contrast, when activated, they develop into lymphoblasts and ultimately plasma cells (see earlier). Some activated cells develop into long-lived memory cells.

Pre-B cells lack membrane-bound IgM and immature B cells may fail to respond to antigen despite the presence of sIgM, owing to lack of accessory molecules. Anergy or tolerance may instead be induced under these conditions. Immature and mature B cells express surface IgM and IgD, whereas activated B cells express IgG after heavy-chain isotype switching. At this stage most of the immunoglobulin is still membrane-bound. After some rounds of activation, the cell switches to high secretion of antibody and becomes a plasma cell.

Antibody generation during B-cell ontogeny

The immune system uses the same general mechanism to generate an almost infinite range of specific antibodies as it does for TCRs, via recombination of the heavy and light chains of the C (constant) regions and somatic mutations in the V (variable) regions of the respective antibody and TCR molecules under control of the *RAG* genes (see p. 431). The genes encoding antibody structure are located on chromosomes 14, 2 and 22 in humans (Fig. 7.34).

In memory B cells particularly, somatic mutations that occur during progression of the B cell through isotype switching to the plasma cells permit the antibody to be 'shaped' to fit the antigen better. This is known as 'affinity maturation' of the B cell and its immunoglobulin molecules, a process that describes the stronger affinity of antibodies for an antigen on repeated exposure (Fig. 7.35). Thus, antibodies produced after repeated exposure to the



STAGE OF MATURATION	Stem cell	Pre-B cell	Immature B cell	Mature B cell	Activated B cell	Antibody-secreting cell
PATTERN OF IMMUNOGLOBULIN PRODUCTION	None	Cytoplasmic μ only	Membrane IgM (κ or λ light chain)	Membrane IgM, IgD	Low rate Ig secretion; heavy chain isotype switching	High rate Ig secretion, reduced membrane Ig
IMMUNOGLOBULIN DNA, RNA	?	Rearranged H chain gene ($VDJ\text{-}C_\mu$); μ mRNA	Rearranged μ and κ or λ genes; μ and κ or λ mRNA	Alternate splicing of VDJ RNA (primary transcript) to form C_μ and δC mRNA	Alternate splicing of VDJ RNA to membrane or secreted Ig mRNA; H chain switching by deletion of C region exons or alternate splicing	Predominantly secretory form of H chain mRNA

Fig. 7.33 B-cell maturation. (Courtesy A. Abbas.)

antigen have a much higher binding capacity for the antigen and are more effective in dealing with its removal. This is the basis of various immunization protocols. Isotype switching is under the control of cytokines and varies for different immunoglobulins; thus IL-4 induces switching to IgG1 and IgE while IFN-γ induces IgG3 and IgG2a and TGF-β induces IgG2b and IgA switching.

Under normal bone marrow conditions of continual cell proliferation and maturation, with unique mutations being produced in each cell during cell division, a very large range of possible permutations of antigen specificity is met within a short time but initially with low affinity. When the correct match between a particular antigen and receptor occurs, antigen-specific clonal expansion ensues. However, this happens only rarely; most cells expressing a particular sIg do not meet the corresponding antigen and are eliminated.

Somatic mutations occur in the CDR of the V genes

Antigen-binding sites are located in the complementarity determining regions (CDRs) of the V segments of the H and L chains, particularly in the hypervariable region.

Details on the features and special characteristics of immunoglobulin isotypes are is available online at https://expertconsult.inkling.com

However, some overlapping binding to the conserved regions of these proteins may also occur. Mutations that increase the affinity of the antigen–antibody reaction occur in the hypervariable regions and become increasingly important with each exposure to the antigen (affinity maturation). Mutations occur in all three CDRs in both chains and may also occur in some intervening regions of the encoded sequence. Antibody diversity is therefore a result of several factors (Box 7.12).

Some microbial antigens (superantigens) have the ability to induce a polyclonal response in B cells in the absence of T cells, similar to T-cell superantigens (see p. 434). Superantigens bind to the majority of VH gene families, especially $VH3^+$ IgM. They also bind to conserved regions of the antibody outside the CDR at FR3 (framework region 3). Antibodies are very important to immunity and protection from microorganisms, especially for memory responses to viruses. In such situations the paracrystalline repeated structural epitopes on the virus surface may induce a B-cell response without T-cell help.

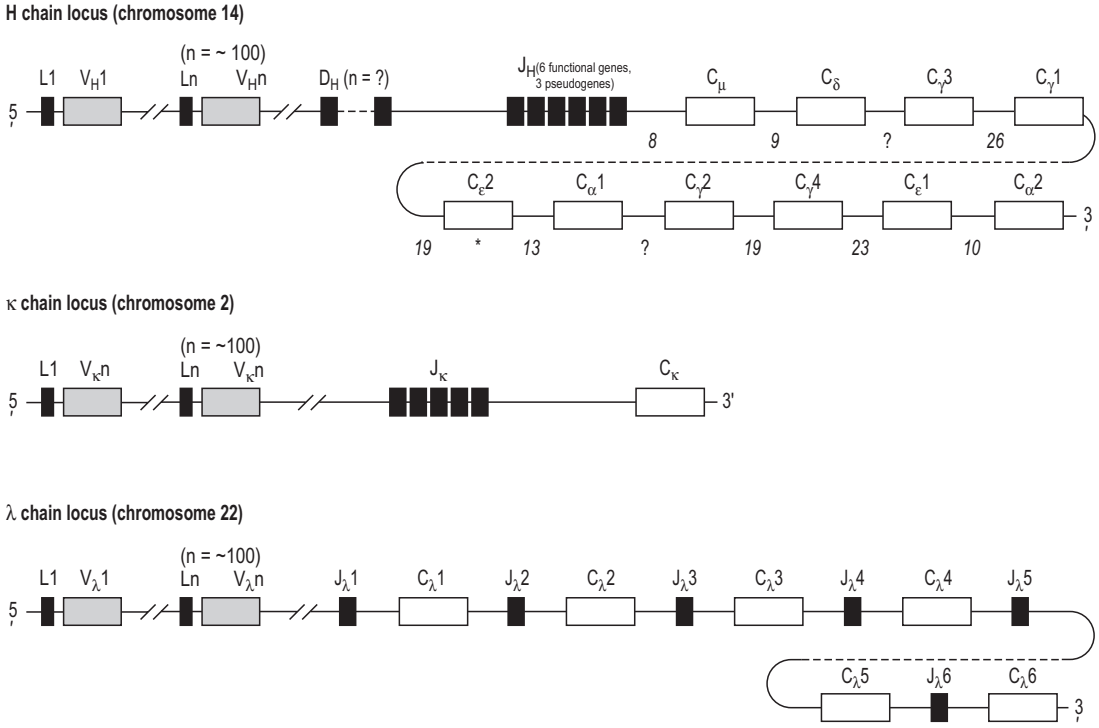


Fig. 7.34 The heavy (*H*) and light (L) chain genes are located in the sequence in which they are transcribed. There are over 100 V genes, each of around 300 base pairs in length and grouped in six or seven families on the basis of >80% homology. Each V gene is preceded by a leader sequence characteristic of secreted and transmembrane proteins. There are an unknown number of D genes and five or six J genes. The J and D genes code for the C-terminal region of the V gene including the third complementarity-determining region (CDR) (see later). Each C gene is composed of several exons represented by the single box. The D_H gene is the main genetic determinant of D–H–J diversity. The amino acid composition of the IgH CD3 region is the driving force for selection. Hydrophilic/aromatic amino acid groups are more likely to be selected than hydrophobic groups. Recombination of the gene segments to generate a mature immunoglobulin is achieved by excision of the intervening DNA and ligation of the immunoglobulin genes. This 'looping out' is facilitated by specific enzymes (recombinases) that act during the earliest stages of B-cell maturation. Segments of DNA, 3′ of each V gene and 5′ of each J gene in the non-coding intervening DNA of the light chain and similarly in the VDJ genes of the heavy chain, are excised and re-annealed. Recombinases themselves are under genetic control via recombination activation genes (*RAG* genes). Recombination occurs in a precise sequence. V_H genes join to the DJ gene, at which point the cell is committed to becoming a B cell. The VDJ sequence then binds to C_μ and a poly-AAA tail is added. The L chain follows the same sequence and the expressed L chain then assembles with the 'μ only' H chain in the endoplasmic reticulum. When this pre-B cell expresses the IgM on its cell surface, it becomes an immature B cell. The μ chain regulates somatic rearrangement by allelic exclusion (activation of genes on one set of chromosomes suppresses activation of the other chromosome) and by initiating L chain rearrangement (the κ chain is activated first; if this is non-productive, λ chain genes are activated). (Courtesy A. Abbas.)

Genetic control of antibody production

In addition to the normal TATA boxes of V region promoters, immunoglobulin genes contain regions modulated by trans-acting nuclear factors that regulate the promoters and enhancers. Trans-acting nuclear factors are DNA-binding proteins, some of which are specific to B cells. Others occur in a number of cells that have to respond rapidly and quantitatively, e.g. NF-κB (also involved in IL-2 transcription) and NF-AT (the target for the immunosuppressive drugs ciclosporin and FK506 used in the treatment of uveitis and other ocular inflammatory conditions).

Tolerance and autoimmunity

The immune system is organized not to respond unless it is absolutely necessary because unchecked responses

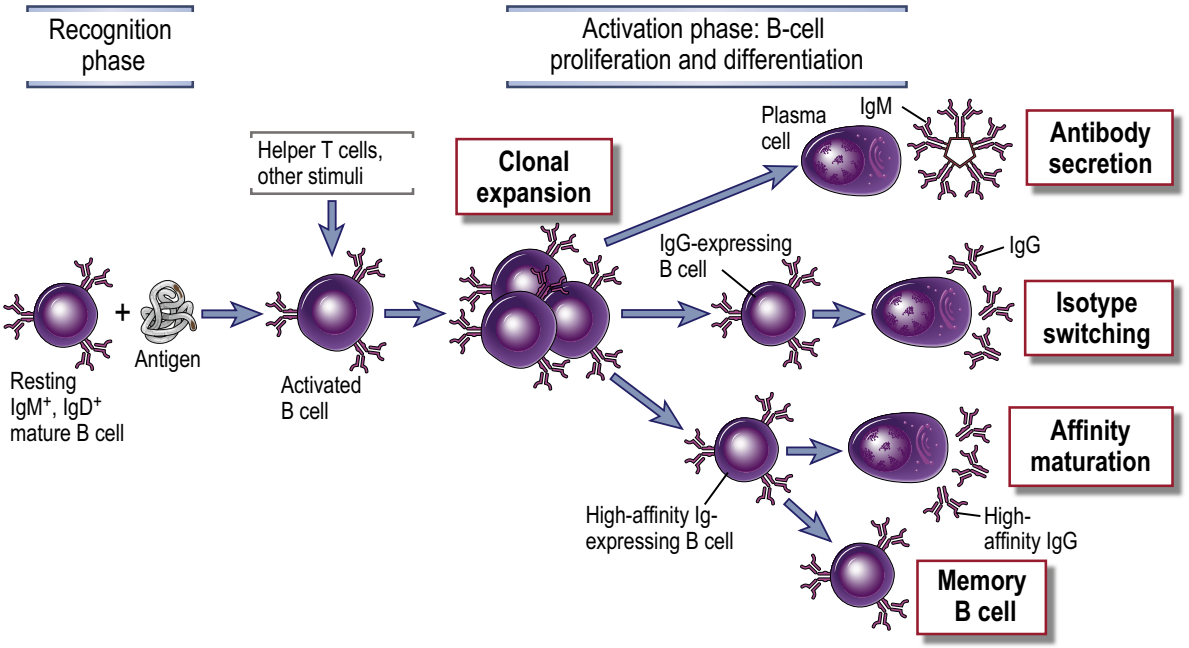


Fig. 7.35 Affinity maturation of the B cell and its immunoglobulin molecules. (Courtesy A. Abbas. *Cellular and Molecular Immunology*, Fig. 11.1)

to non-life threatening pathogens could compromise host function. Therefore, most of the immune system is geared not to respond, i.e. is under inhibitory regulation. This applies not only to infectious agents but also to any potential antigen including self-antigens. A prime example of appropriate non-response is found in the gut, skin and respiratory tract, which harbour trillions of bacteria and other microorganisms. This non-responsiveness of the immune system is known as *immune tolerance*. Tolerance requires recognition by the immune system (we all have autoantibodies and T cells that can be shown in laboratory tests to be specific to self-antigens) but the products are not pathogenic, i.e. they do not cause disease. Indeed, they are usually protective, e.g. natural antibodies and Tregs. Indeed, in the absence of Tregs, autoimmune diseases occur because effector anti-self T cells, which we also generate naturally, respond in an uncontrolled manner when they engage self-antigens .

What is tolerance?

Tolerance is a feature of the adaptive immune system. Innate immunity is constantly active and deals with all challenges. If the insult is small/minimal, innate immune mechanism can adequately deal with the problem. Tolerance may be defined as antigen-induced inhibition of the development, growth or differentiation of pathogenic antigen-specific lymphocytes, i.e. adaptive immunity. Tolerance has the following properties:

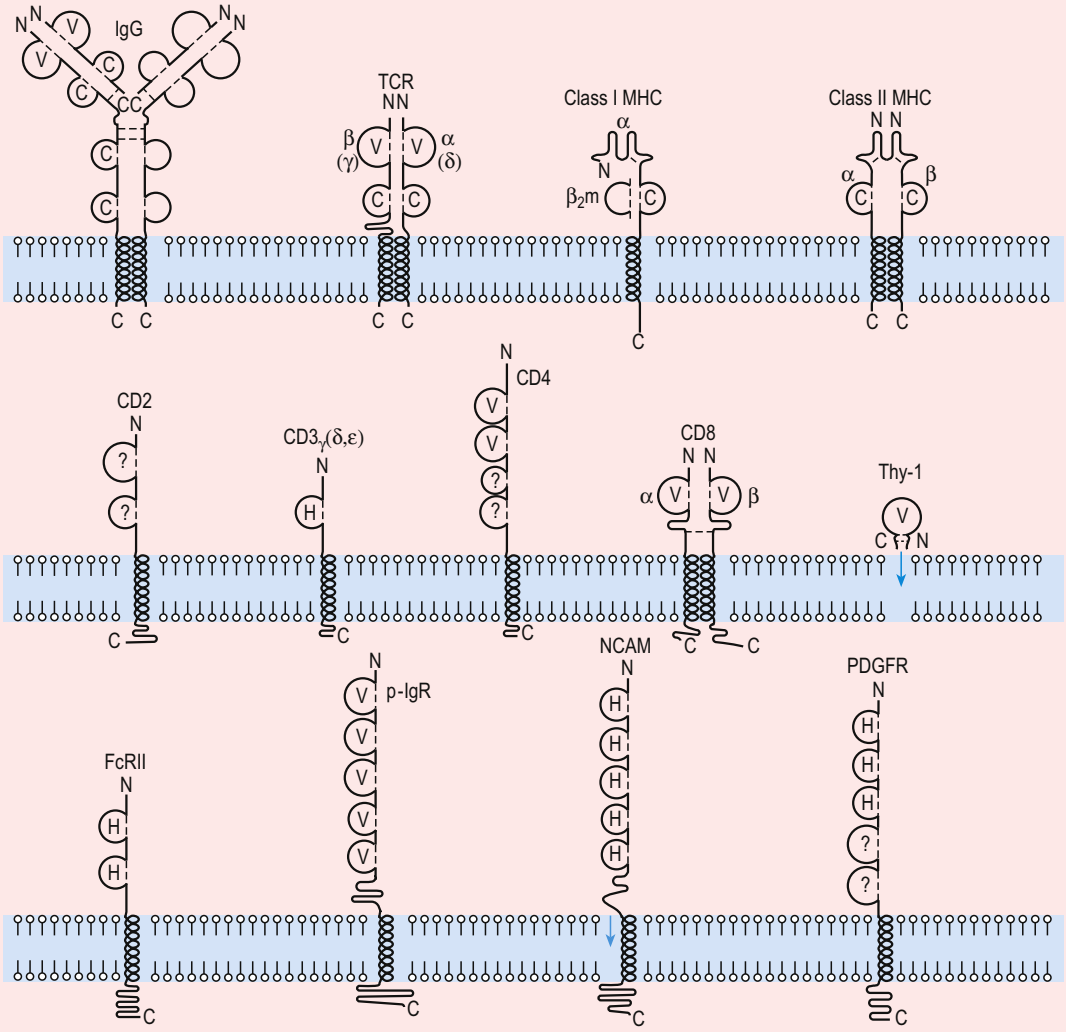
- It is antigen-specific: individuals who are tolerant to one antigen are not necessarily tolerant to all antigens or even a second related antigen
- Tolerance to self-antigens is acquired during development in the thymus and continues throughout life; however, immature lymphocytes develop tolerance more easily than adult ones.
- Tolerance develops both to foreign antigens (e.g. commensals) and self-antigens.
- Maintenance of antigen-specific tolerance requires persistence of the antigen throughout the life of the individual.

Sites of tolerance induction

Tolerance classically has been described to occur in two principle sites: centrally in the thymus (central tolerance) and peripherally in the secondary lymphoid organs (SLOs) such as the spleen and the lymph nodes. Deletional tolerance occurs as a result of death of the majority of T cells in the thymus, which are exposed to antigens of high affinity for TCRs such as self-antigens and for commensal antigens by what has been described as 'postrevision' of the TCR. Regulatory tolerance occurs through generation of antigen-specific regulatory cells in the thymus and the periphery, which control/prevent activation of pathogenic T cells. A third form of tolerance can

Box 7.11 **Immunoglobulin superfamily of proteins**

The immunoglobulin superfamily is characterized by the common structural motif, the immunoglobulin domain. This domain may occur singly in small molecules such as Thy-1 or multiply in both chains of dimers such as the T-cell receptor.



occur by failure of pathogenic T cells to continue responding to antigen and thus becoming anergic. Such cells still retain the potential to be activated under the right conditions.

Recently, Matzinger's concept that the organism does not truly distinguish between self- and non-self-antigens but does distinguish between immunogenic and non-immunogenic antigens (the danger hypothesis; see p. 381) and that self-antigens themselves can become dangerous under appropriate conditions has been extended to include a role for the tissues where the immune responses occurs. Thus, various non-lymphoid tissues can modify the local immune response and determine whether a pathogenic effect takes place (tissue-based tolerance).

Further information on tissue-based tolerance is available online at https://expertconsult.inkling.com

TISSUE-BASED TOLERANCE

Matzinger has suggested a third mechanism for immunological tolerance in addition to the two canonical mechanisms of deletional (thymic) and regulatory tolerance. She has suggested that the immune response is modulated by the tissue in which the reaction takes place. Thus, for instance, certain tissues constitutively secrete mediators that influence the nature (e.g. Th1 vs Th2), the strength (a cytokine storm or a chronic low-level inflammation) or even the outcome (complete healing of the tissue with resolution of the inflammation or permanent tissue damage with continued dysfunction). For instance, mediators such as immunosuppressive cytokines, TGF-β and vasoactive intestinal peptide (VIP) are common to both the eye and the gut. The ocular cells also secrete α melanocyte stimulating hormone (αMSH), while the gut produces thymic stromal lymphopoietin (TSLP), which ultimately promotes IgA production. IgA is the immunoglobulin generated by mucosal tissue to provide mucosal immunity. Matzinger and Kamala (2011) have suggested that the local tissues modulate or are in control of the immune response (see section on further reading).

Box 7.12 Factors controlling antibody diversity

Multiple germline genes
Somatic combinatorial diversity
Junctional diversity
H and L chains contributing to the antibody combining site
Somatic mutations

In the mouse it has been estimated that the potential antibody repertoire resulting from recombination mechanisms is 10^9–10^{11} (i.e. before antigen has initiated a response).

Thus, immune responses to the same antigen or microorganism may differ depending on the nature of the tissue type and location. This in fact is not a new concept but an elaboration of the notion of *immune privilege* (see later), which is particularly well expressed in the eye and the brain and is a means whereby ocular and brain tissues modify (downregulate/dampen) immune and inflammatory responses generally.

The same cellular mechanisms are utilized to induce central, peripheral and tissue-based tolerance.

Mechanisms of tolerance induction

Immune tolerance (better described as immunological non-responsiveness, although this is rather unwieldy) is considered under various forms, e.g. neonatal versus adult, central versus peripheral, innate versus acquired, etc. In fact, tolerance is always acquired; the differences arise when the acquisition of tolerance occurs in the thymus during development (deletional or central) or in the peripheral lymphoid tissues during adulthood (regulatory/peripheral).

Several mechanisms of tolerance have been suggested including clonal deletion, anergy and induction of regulatory cells.

Further information on general mechanisms of tolerance is available online at https://expertconsult.inkling.com

Failed tolerance

A failure of any of the three general mechanisms of tolerance that regulate the immune response to antigens (clonal deletion, anergy and regulatory cells) leads to the development of dysregulated immunity and in some instances, autoimmune disease. To some degree, tolerance is incomplete and natural autoimmunity to autoantigens is the norm in the form of IgM autoantibodies and non-pathogenic autoreactive T cells. This has been shown for most autoantigens, including several ocular/retinal antigens. Therefore, the detection of autoantibodies or autoreactive T cells is in itself insufficient to claim pathogenicity for these agents, making the interpretation of data difficult, for instance in studies of autoimmune retinopathy and autoimmunity in the context of chronic disease such as glaucoma and AMD.

Autoimmune disease is the dysfunction or damage of tissue caused by immune responses to autoantigens

Autoimmune diseases take many forms, from organ and even cell-specific antibody-mediated diseases such as myasthenia gravis (where the antigens are located at the neuromuscular junction) to widespread systemic diseases such as systemic lupus erythematosus, where the antigen(s) is (are) distributed in all tissues (DNA).

Tissue damage in autoimmune disease can be induced by any of the accepted forms of immunopathological mechanisms (types I–IV; Table 7.7). However, as for most immune mechanisms, autoimmune diseases are usually initiated by $CD4^+$ T cells. Several mechanisms have been proposed to account for this process:

- *Molecular mimicry* between foreign and autoantigen sequence homologies (as might be expected with the wide range of antigenic peptides occurring in infective organisms) occur with predictable frequency. Processing of foreign antigenic peptide might thus lead to activation of autoreactive T cells if the foreign antigen is 'mistaken' for self. This has been shown for several retinal antigens that have amino acid sequences similar to bacterial and viral antigens, including Gram-negative bacteria such as *Escherichia coli* and parasites such as *Onchocerca*, (which harbours Wolbachia), to cause endemic blindness in certain regions of Africa (river blindness).
- *Bystander activation:* the initiation of autoimmune disease may also occur by bystander activation, for instance during infection. Upregulation of co-stimulatory molecules on APCs in the lymphoid tissues in the vicinity of anergized but not deleted autoreactive T and/or B cells may lead to autoimmune disease. Anergy may require expression of the co-inhibitory molecule CTLA-4 (see pp. 437 and 440), which, if absent, may also allow self-autoreactivity.
- *Failure of Treg activity*: tolerance is mediated by Tregs (see earlier) that homeostatically inhibit autoreactive T cells, possibly by direct cell contact or release of inhibitory cytokines. Tregs constitute a major proportion of circulating $CD4^+$ T cells (~5%–10%). Reduced Treg activity is a feature of most autoimmune diseases.
- *Failure of deletion* (activation induced self-death): T-cell deletion and apoptosis is mediated by cell surface molecules such as Fas–FasL and TNF-receptor apoptosis inducing ligand (TRAIL)—death receptor (DR4/DR5).

HOW IS IMMUNE TOLERANCE DEVELOPED AND SUSTAINED

The immune system has several mechanisms for ensuring that it does not go into overdrive and turn on the host. When these mechanisms fail, the immune system adds to the damage caused by the inciting immunogen whether it is self or non-self.

Clonal deletion

This mechanism was originally proposed to explain the antigen-induced destruction of self-reactive lymphocytes that occurs in the thymus during development (negative selection) in relation to tissue-specific antigens expressed in the thymus under the control of genes such as *Aire*. This is a form of activation-induced cell death (AICD) (see p. 431) and can occur both centrally (in the thymus) and the periphery (lymph nodes) and is thus better described as deletional tolerance. It is dependent on high affinity of the antigen for the TCR.

It has been postulated that removal of all autoreactive cells occurs in the thymus, both for the large pool of thymic antigens and for tissue-specific antigens, which are transported to the thymus on circulating dendritic cells. This has been shown for certain brain (myelin basic protein) and retinal (interphotoreceptor binding protein and retinal S antigen) proteins under the control of *Aire* expressed in thymic medullary epithelial cells (Fig. 7.19).

However, this mechanism does not account for the induction of tolerance to all tissue-specific (non-thymic) antigens and peripheral tolerance mechanisms are also necessary, some of which also include T-cell deletion.

Anergy

Anergy describes T- and B-cell non-responsiveness to specific antigens in cells that respond to non-specific mitogenic stimuli, i.e. the cells are in a sort of 'suspended animation'. Anergy in lymphocytes is thought to arise via lack of co-stimulation required for antigen presentation to Th1/ Th2/ Th17 cells (the B7:CD28 and the CD40:CD40L interactions are considered particularly important; see pp. 437) and has been proposed to account for the non-responsiveness of lymphocytes to self-antigens. It is therefore considered to be a major mechanism of immune tolerance in the adult.

However, it is likely that clonal deletion and anergy involve the same processes. Thymic education of lymphocytes involves a discrete series of events in arming the T cell to develop towards a $CD4^+$ or a $CD8^+$ phenotype (see Fig. 7.19) and antigen specificity. Each of these steps requires more than one signal and, at any point, lack of a particular signal might induce anergy. If anergy persists, the cell is driven down a pathway towards apoptosis (i.e. deletion). Similar mechanisms exist in the bone marrow and in the germinal centres of lymph nodes and spleen. Apoptotic B cells are removed from germinal centres by 'tangible body' macrophages (tangible bodies contain condensed chromatin from apoptotic cells).

Thus, the same mechanisms for removal of autoreactive cells, involving anergy and cell death, probably occur in the peripheral and the central lymphoid tissues, although there is a fine balance between deletion and anergy in the induction of tolerance.

Active suppression by specific lymphocytes: T-regulatory cells

Transfer of lymphocytes from animals exposed to the CNS antigen myelin to mice developing autoimmune CNS inflammation can prevent autoimmune encephalomyelitis. The mice in this case are rendered tolerant to myelin and tolerance is mediated by cells that inhibit or regulate the immune response. Originally the cells were described as suppressor cells, probably acting at the level of antigen presentation. The concept of suppressor cells was initially considered dubious, mainly because of their lack of responsiveness to specific antigen *in vitro* (anergy). However, seminal work by Sakaguchi in (Sakaguchi et al., 1985) Japan showed that depleting mice of a subset of $CD4^+$ T cells expressing high levels of the IL-2 receptor (CD25) led to the development of spontaneous autoimmune disease with severe pathology particularly affecting the bowel (inflammatory bowel disease). These cells, rebadged as T-regulatory cells (Tregs, see pp. 391 and 427), are defined by the transcription factor FoxP3 (see Figs 7.2, 7.24 and 7.25) and are required for the prevention of autoimmune disease in humans as well as mice. Tregs are decreased in various forms of ocular inflammation such as Vogt–Kayanagi–Harada (VKH) disease. Other subsets of Tregs have also been discovered and they are generally characterized by the inhibitory cytokines they produce. Such cells include IL-10-producing Tr1 cells, the TGF-β-producing Th3 cells, $CD8^+$ suppressor T cells, which mediate their action through macrophages and NKT cells, B regulatory cells and γδ T cells. It is not clear how Tregs induce the suppression of T-effector cells, but the effect is believed to be via DC and not directly through Treg:T effector interactions (Fig. 7.36).

T-cell tolerance is different from B-cell tolerance

The Th cell is the critical control element for adaptive immune responses. Therefore, inhibition of Th responses should lead to inhibition of all self-reactivity. T cells can be rendered unresponsive (anergic) by very low doses of antigen and B cells usually require larger doses. T-cell tolerance induced by clonal anergy/deletion is antigen-specific and MHC restricted. In addition, natural/thymic (nTregs, tTregs, both terms used reciprocally) are generated in the thymus and released to the periphery. In the periphery, Tregs constitute the main regulator of autoimmunity (see earlier) and include both natural/thymic (n/t) Tregs and antigen-specific induced (iTregs) derived from naive T cells during the progression of all adaptive immune responses (i.e. to both foreign and auto-antigens). Anergic T cells and Tregs may also be induced by presentation of antigen by aberrantly expressed MHC class II antigen on parenchymal cells and endothelium in the absence of co-stimulation.

HOW IS IMMUNE TOLERANCE DEVELOPED AND SUSTAINED—CONT'D

B-cell tolerance occurs particularly to antigens that are 'T-independent'. This includes carbohydrate and glycolipid antigens such as the ABO blood groups and thus has direct clinical relevance. Tolerance is induced by anergy/deletion of antigen-specific B cells in the marrow by mechanisms similar to those described for T cells in the thymus. Studies using transgenic mice in which genes for neoantigens and the corresponding antibody were inserted into the genome have elegantly demonstrated that B-cell anergy in the marrow occurs by maturation arrest (see page 442) and failure of these autoreactive B cells to enter the peripheral lymphoid organs. After some time, these cells underwent apoptosis (i.e. were deleted). Interestingly, some of these B cells could render the autoantigen non-anergy-inducing by switching the κ/λ chain. B cells that had undergone 'receptor editing' of this nature appeared in the periphery (i.e. were not deleted).

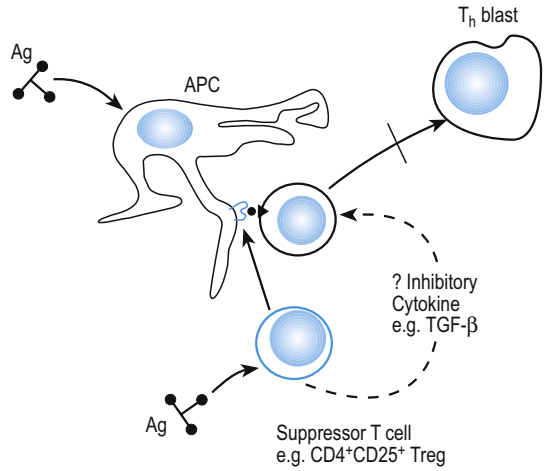


Fig. 7.36 Mode of action of regulatory cells. *Ag,* Antigen; *APC*, antigen-presenting cell.

Failure of the apoptotic machinery might lead to autoimmunity. TRAIL and Fas have both been shown to play a role in immune privilege in the eye and brain.

- *Polyclonal B-cell activation*: certain compounds such as endotoxin and bacterial glycolipids can activate B cells directly, either to produce cytotoxic antibody or to act as APCs and thus present autoantigen to responding T cells.
- *Superantigen*: simultaneous activation of several subsets of T cell by superantigens (see page 434), which do not require processing because they link the T cell and MHC antigen directly, may also activate autoreactive T cells. Preferential use of certain TCRs by superantigens might enhance the risk of autoimmune responses.

Auto-inflammatory disease

As indicated previously, although there are many theories concerning how autoimmune diseases might occur, definitive evidence has been difficult to establish. A new concept of autoinflammatory diseases has therefore taken hold. Evidence has been found for single gene defects in molecules involved in innate immune defence (pattern recognition receptors and their signalling pathways (see p. 409) and more recently, polygenic diseases such as Behçet's disease, diabetes mellitus and even AMD are being considered against a background of deficient innate immunity (Forrester et al., 2018).

Latent infections persist by giving false signals to the immune system and using immune privileged sites to hide

Most microorganisms either kill the host or are successfully cleared from the system. Some pathogens, however, persist in the tissues and have a preference for privileged sites (see p. 460) such as the eye and the brain where they are protected from immunological onslaught (Forrester et al., 2018). Viruses, fungi, mycobacteria and parasites all use this mechanism and many are associated with diseases that are of major medical and economic importance worldwide. Parasites are a good example: they invade the tissues and enter a symbiotic relationship, sometimes within the cytoplasm of the cell, particularly macrophages. Extracellular parasitic infections are even more frequent, such as those involving nematodes. Examples include filariasis, schistosomiasis and toxocariasis, the first and last of which can infect the eye. Ocular toxocariasis causes blindness in children, presenting as one cause of the white (cat's eye) pupillary reflex (see Chapter 9, p. 504).

Parasites present the immune system with enormous problems, partly because they go through a life cycle in which they present different antigens at different times; the host therefore has difficulty in generating an adequate response and may enter a chronic state of inflammation. The parasite achieves this by manipulating, evading or diverting the immune system. Although a Th1 response may be initially induced after infection by the parasite, the response is later diverted to become Th2-mediated with secretion of IL-2, IL-4 and IL-10. Several explanations have been proposed for this change, including defects in the presentation of parasite antigen, inappropriate production of regulatory cytokines such as IL-10 and IL-13 or the induction of a form of tolerance via prolonged binding of the antigen to the TCR. However, rather than being an anergic or impaired type of immune response, the switch from Th1 to Th2 may be quite pronounced. In some respects, it is similar to the response in allergic disease because the Th1/Th2 immune response system becomes similarly dysregulated and can lead to excessive mast cell infiltration and eosinophilia, particularly with helminth infestations.

As indicated previously, most parasitic infections are cleared during the initial immune response. For instance, toxoplasma infection (a protozoan parasite) involves induction of a strong INF-γ-mediated Th1 response through APC production of IL-12, IL-23 and IL-27, which eliminates most of the parasites. However, some parasites escape detection inside circulating monocytes, which then 'hide' in sites of immune privilege such as the eye, brain, muscle and liver. The parasite remains undetected until it changes its antigenic profile by dying or proliferating, thereby inducing a prominent immune response via the newly exposed antigens released into the tissues. Toxoplasmic retinochoroiditis is thought to occur by this mechanism. This relatively common cause of posterior uveitis is characterized by one or more large chorioretinal scars that undergo spontaneous reactivation, producing recurrent attacks of uveitis as the organisms change over time (see Chapter 9, p. 504)

Allergy and immediate hypersensitivity

The immune system's attempts to rid itself of foreign elements, most of which are harmless, can sometimes be so exaggerated or inappropriate that the host is damaged in the process. Such a situation arises in type I hypersensitivity (see Table 7.7) involving IgE responses to foreign antigens (allergens), thereby causing allergy. As will become clear, not all allergic responses are IgE mediated.

Atopy, asthma and allergic eye disease

Allergic reactions are usually effective, short-lived immune responses that clear the allergen. Allergic disease is where this fails.

Allergic reactions have the following features:

- allergen-specific Th2 cell activation and production of IgE by B cells
- binding of IgE to mast cells with mast cell degranulation
- very wide range of 'allergens' including animal dander, chemicals (haptens), house dust mite proteins, commensals including fungal proteins, plant and other environmental proteins
- symptoms of immediate hypersensitivity resulting from cytokines.

Allergic disease occurs when the allergic reaction (i.e. the immune response) is exaggerated causing overt symptoms. These can be very severe and acute, even life-threatening (as in acute asthmatic attacks). The allergen may be cleared in this process or it may persist and induce the development of chronic disease (e.g. eczema, atopy, chronic asthmatic wheeze).

The prototype allergic disease is asthma, in which reversible obstruction of the airways occurs. This is often the acute manifestation of a chronic inflammatory process in the mucous membranes and skin. Asthma frequently coincides with chronic dermatitis (eczema), rhinitis and sinusitis, and conjunctivitis. When there is a strong genetic predisposition to synthesize IgE specific for certain external antigens, the disease is known as atopy. However, patients with asthma are not necessarily atopic. Other forms of asthma in non-atopic individuals are recognized (intrinsic versus extrinsic asthma). Similarly, all patients with severe allergic conjunctivitis are not always atopic. These concepts are important clinically as seen by the selective efficacy of monoclonal antibody anti-IgE therapy in the treatment of refractory asthma.

Allergic conjunctivitis induced by specific antigens such as pollens or house dust mite may be chronic (perennial allergic conjunctivitis) or intermittent (seasonal allergic conjunctivitis) and usually produces mild symptoms. Chronic conjunctivitis in patients with atopy is more damaging, often with involvement of the cornea (atopic keratoconjunctivitis; see next section). A distinct clinical entity with massive follicular conjunctivitis and corneal opacification is known as vernal keratoconjunctivitis. In addition, contact lens wear is associated with a significant amount of allergic eye disease, producing giant papillary conjunctivitis.

In spite of the clinical differentiation between various forms of allergy, it is likely that the underlying mechanisms are similar. Activation of Th2 cells is mediated by migratory DCs, which capture allergen, activate Th cells in the LN to become Th2 cells, which secrete IL-4, IL-5, IL-13 and Tfh cells, which activate B cells in the B-cell follicle. Mast cells have long been recognized as effector cells in allergic disease. However, mast cells are dependent on IL-3 and IL-5 production by T cells, which is ultimately induced by antigen presentation to T cells (Fig. 7.37). In this regard, conjunctival DCs have been considered to be major contributors to allergen capture and presentation to T cells locally in the tissues. In severe chronic asthma and ocular surface (conjunctival) disease, there is massive thickening of the subepithelial mucosa with fibrosis and infiltration of other pro-inflammatory granulocytes such as eosinophils, which are important in tissue damage. It has been suggested that metabolite sensing G-protein coupled receptors (GPCRs) that line the intestinal tract dynamically respond to a wide range of dietary and nutrient factors and regulate microbiome-dependent immune and metabolic factors. Allergic diseases are considered to be a result of dysfunction in one or more of these GPCRs.

Mast cells and mast cell degranulation

Mast cells are derived from precursors in the marrow and mature into one of two phenotypes depending on the microenvironment in which they reside. Stem (c-kit) cell factor is an important mast cell growth factor. In humans, mucosal mast cells contain tryptase (M_T), histamine and heparin in their secretory granules, while connective tissue mast cells also contain chymase (M_{TC}). In the normal conjunctiva and choroid, M_{TC} cells predominate. However, in allergic conjunctivitis, M_T cells appear to increase in number in association with the expression of adhesion molecules such as E-selectin and ICAM-1. These changes are likely to be important in the pathogenesis of disease (see later). The significance of mast cell heterogeneity is not yet clear but it would appear that M_T are involved in the active stages of inflammation.

Mast cell degranulation is mediated both by IgE or by non-IgE mechanisms (see later). The role of

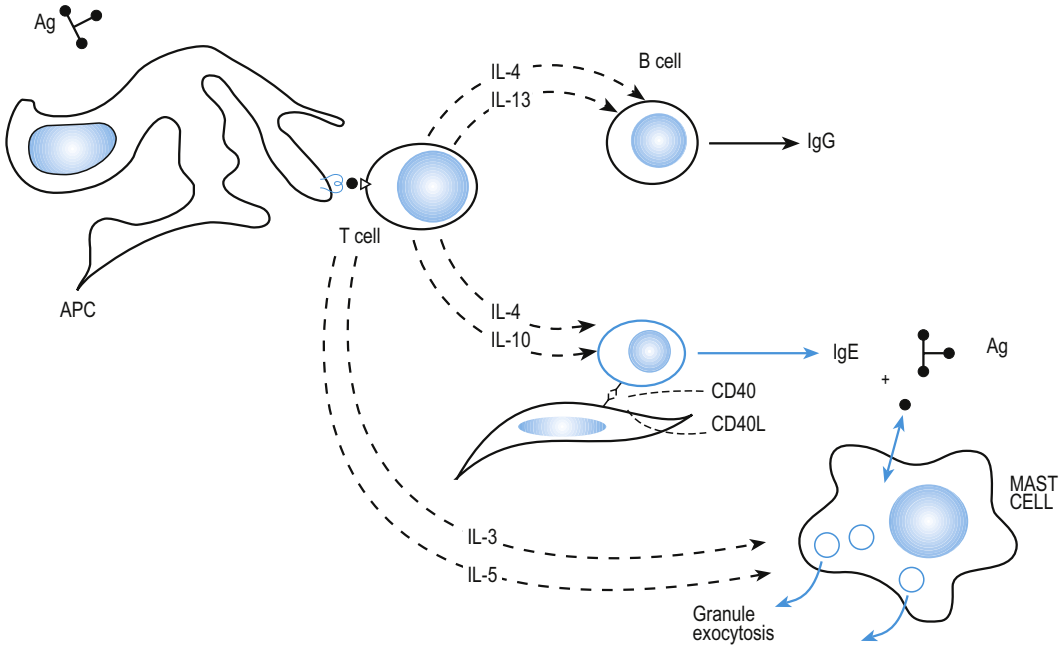


Fig. 7.37 Mast cell activation is dependent on antigen-specific activation of T cells *(Ag)*, which, in concert with concomitant engagement of the CD40 membrane protein by the CD40 ligand (CD40L), induces an isotype switch in the presence of Th2-type cytokines to direct the B cell to produce IgE. IgE–antigen complexes are then available to activate the mast cell. *APC*, Antigen-presenting cell.

antigen-specific IgE in this response can be demonstrated *in vivo* experimentally by the passive cutaneous anaphylaxis test. However, IgE is neither sufficient nor essential to induce mast cell-related allergic disease. Chemokines are also involved, mediating not only leucocyte recruitment but also mast cell activation through mediators such as the chemokines eotaxin 1 (CCL11) and MIP-1α (CCL3).

Mast cell degranulation leads to the release of preformed mediators including histamine and serotonin, which bind to a variety of receptors and induce second signalling events with release of a great variety of secondary agents depending on the cell type. This includes NO•, prostacyclin, smooth muscle relaxants and many others. Several receptors exist for histamine (H_1, H_2, H_3, etc.), which can be distinguished by pharmacological agents (Fig. 7.38). Mast cells also synthesize and release newly formed materials after stimulation. These include prostaglandin D_2 (vasodilator and bronchoconstrictor), leukotrienes (LTC_4, LTD_4, LTE_4, previously known as slow-releasing substance A, cause prolonged bronchoconstriction) and platelet-activating factor (PAF).

Cytokine release by mast cells is the major inflammatory pathway. TNF-α, IL-1, IL-4, IL-5, IL-6, IL-13 and several colony-stimulating factors are released and help to recruit cells to the site of antigen exposure on the mucous membrane.

Immunoglobulin and helper T cells

There are two types of IgE receptor: FcεRI, which is present in mast cells, basophils and activated eosinophils, and FcεRII (CD23), a low-affinity receptor present on many cell types including B, T and dendritic cells, monocytes and some thymic epithelial cells. FcεRII is important for initial antigen capture by dendritic cells.

IgE-mediated degranulation of mast cells occurs via high-affinity receptors (FcεRI). Signalling via the IgE receptor is inducible only by IgE–antigen complexes and requires local production of IgE by infiltrating B cells.

Isotype switching in B cells from IgG to IgE requires binding of the cell surface antigen CD40 on the B cell to a ligand on the T cell (gp39). This can also be induced by mast cells that express the CD40 ligand and promote local production of IgE in the tissue under cytokine stimulation (Fig. 7.39). Thus, the disease can be perpetuated locally and enter a phase of chronicity, as occurs in asthma.

Not all allergy/asthma is mediated via IgERI or II binding and IgG-linked mechanisms have also been implicated.

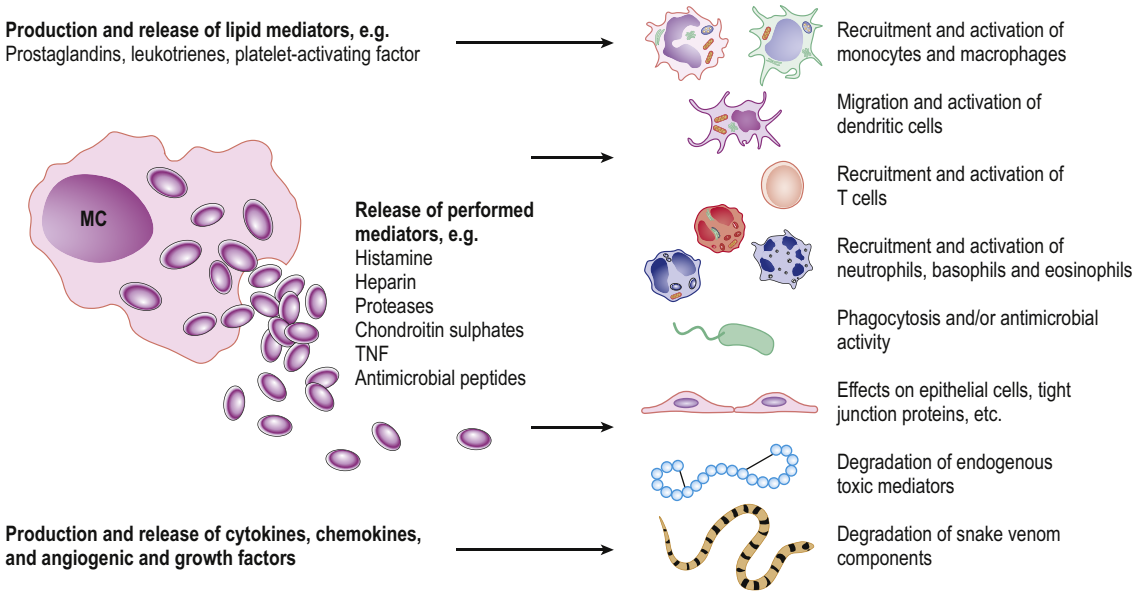


Fig. 7.38 Mast cells have preformed granules ready to be released instantly on activation. They can also produce massive amounts of mediators *de novo* and are thus very potent cells with diverse effects on a wide range of cells as indicated. (From Metz and Maurer, 2007.)

Mechanism of disease production in allergic disease

Exposure of atopic individuals to allergen is accompanied by an immediate response mediated by mast cell degranulation and a later response occurring after 4–6 hours. The later response is induced by cytokine release from mast cells, which causes recruitment of M_T cells and basophils (see pp. 387–389) to the tissue and the expression of adhesion molecules, which promote eosinophil accumulation. In addition, release of the eotaxin (CCL11) by airway epithelial cells is probably the main stimulus for eosinophil recruitment. T cells are not involved in this stage of the disease.

Eosinophils are strictly regulated because of their potential toxicity. In the absence of infection, very few eosinophils are produced. Even in the presence of an eosinophil, entry to the tissues is tightly controlled by epithelial eotaxins (via CCR3) and by other chemokines such as MCP-3 (CCL7), MCP-4 (CCL13) and RANTES (CCL5) .

Release of granule content by eosinophils, particularly erythropoietin (EPO), eosinophil major basic protein (MBP) and eosinophil cationic protein (ECP), causes much of the tissue damage. Eosinophils are attracted to the site of inflammation by specific adhesion molecules, particularly VCAM-1, on the endothelial cells. Induction of VCAM expression on endothelial cells is mediated by cytokine release from other inflammatory cells such as macrophages and T cells. Th2 cells induced by activated DC and ILC2 cells activated by alarmins such as IL33 released form the epithelium produce IL5, IL4 and IL13, the two latter cytokines promoting IgE production, via B-cell activating factor (BAFF). Th9 cells are also involved in maintaining the mast cell population in asthma. Th2 cells are intricately involved in destructive tissue damage during the progression of allergy from a reaction to a disease, but IL5 is involved mostly in exacerbations of disease. TH17 cells are involved in recruitment of neutrophils in severe asthma, which are attracted to the site as a result of upregulated adhesion molecule expression, particularly of E-selectin and ICAM-1 in the acute stage. Neutrophils are themselves responsible for considerable tissue injury in certain forms of allergic disease and characterize a subtype of refractory asthma. The risk of superadded infection in these cases is unclear. IL13 is also important in generation of viscous mucous plugs that occlude the bronchi and obstruct breathing, leading to death in some cases.

In the later stages of the disease, increased numbers of T cells may be recruited as part of the continuing inflammatory response because of persistent adhesion molecule expression on endothelial cells, especially VCAM. T cells are then available to interact with APCs, particularly activated B cells, causing further release of IL-3, IL-5, IL-13 and GM-CSF, which perpetuates the disease.

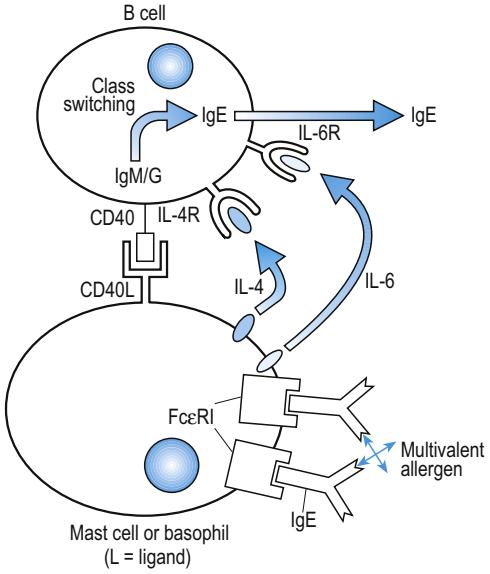


Fig. 7.39 Mast cells can themselves initiate isotype switching via CD40 because they express the CD40L antigen when they combine with allergen–IgE complexes. In this way the allergic response is amplified and perpetuated.

Mast cells also have a protective role

Mast cells may also have a role in the initiation of the immune response as effectors of innate immunity mediating adaptive immunity, particularly to organisms inducing Th2 responses such as helminths and ticks (Fig. 7.40). Mast cells express MHC class I and class II antigens and presentation of exogenous helminth antigen to T cells leads to activation reciprocally of the mast cell with release of chemokines and cytokines. These direct the overall response towards a Th2-type response.

Organ and tissue transplantation

Transplants or grafts are described as syngeneic when they are between genetically identical individuals, allogeneic when they are between individuals of the same species and xenogeneic when they are between individuals of different species. Autologous grafts refer to the transplantation of tissue within the same individual, for instance when the cornea from one eye is transplanted to the fellow eye. Allogeneic and xenogeneic grafts are normally not accepted unless the immune system is suppressed (usually by immunosuppressive drugs) and rejection is described as acute, chronic or accelerated.

The discovery of HLA antigens (see pp. 428–430) came about as a result of studies that attempted to explain how organ and tissue grafts, especially skin grafts, between members of the same species were rejected. Only in the case of genetically identical individuals were grafts accepted. For instance, in skin grafts, not only was there rejection of the initial graft between two non-identical individuals but also subsequent attempts to graft donor skin between the same two individuals led to an even more rapid rate of rejection, indicating that there was some sort of memory response to the antigens in the original graft.

Although potentially any antigen in the donor graft could act as a target for rejection (a 'danger' signal similar to an altered autoantigen or an antigen from a pathogen such as unsuspected herpes simplex or prion infection), the strongest candidate antigens are those where there is the greatest frequency of polymorphisms (differences between individuals or alloantigens) and top of the list are MHC antigens because they are highly polymorphic. Cells with the strongest expression of MHC antigens are leucocytes and rejection of grafts is therefore considered to be the result of the presence of alloantigens on 'passenger leucocytes' in the donor graft. For solid organs and bone marrow grafts, tissue matching on the basis of both blood group and HLA antigens is therefore routinely performed and the closest match attempted in order to ensure graft survival. However, even in these circumstances immunosuppressive drugs are often necessary and usually for the lifetime of the patient.

Further information on direct, indirect and semi-direct allorecognition is available online at https://expertconsult.inkling.com

Polymorphisms (genetic differences) are at their maximum in HLA antigens. However, they also occur in many other antigens to variable degrees and these are known as minor antigens. Minor antigens can also generate a host immune response and induce graft rejection, but the intensity of the response is considerably less. Mismatching at the minor alloantigen loci is a cause of chronic rejection, while accelerated rejection occurs predominantly between xenogeneic grafts (where differences exist in non-protein antigens such as carbohydrate antigens) or in individuals who have been sensitized to the donor, e.g. via previous blood transfusion. Disparities in MHC genetic loci have classically been determined by the mixed leucocyte response (MLR), an *in vitro* test in which leucocytes derived from one individual (donor) are mixed with those from another individual (host) and an immune response determined by the proliferation of the host leucocytes (the donor leucocytes are irradiated to prevent them from proliferating so that any response is the result of antigen presentation to host T cells). If the number of MHC disparities between the host and the donor is large, there is a correspondingly high proliferative response. However, the great advances in whole

DIRECT, INDIRECT AND SEMI-DIRECT ALLORECOGNITION

Since, in any one individual, leucocytes express a large number of MHC alloantigens (see p. 428), many allopeptides will be presented to the host, thus leading to polyclonal expansion of both CD4 and CD8 T-cells. Damage to the graft is the result of alloantigen-specific $CD8^+$ cytotoxic T cells recognizing MHC antigens on the donor tissue. The number of mismatches on the MHC antigens between the donor and host determines how many different clones of CD8 T cells will be expanded and thus how quickly and acutely the graft will be rejected.

When the *donor* APCs directly present the allopeptide derived from allo-MHC-peptide complexes to the recipient, it is known as the *direct pathway* of graft rejection (direct allorecognition). In contrast, when *recipient* APCs present donor allopeptide (derived from allo-MHC-peptide complexes, which they would obtain after phagocytosis of apoptotic donor cells), it is the *indirect pathway* of graft rejection cells (indirect allorecognition). A third newer concept has been proposed, i.e. that *recipient* APCs can present unprocessed but conformationally re-arranged *donor* MHC-allopeptide expressed on the recipient APC surface (again obtained after phagocytosis of donor cells), and is known as the semi-direct pathway. Grafts are therefore more likely to be accepted if the MHC antigens between donor and host are as closely matched as possible and this is clearly the case in identical twins.

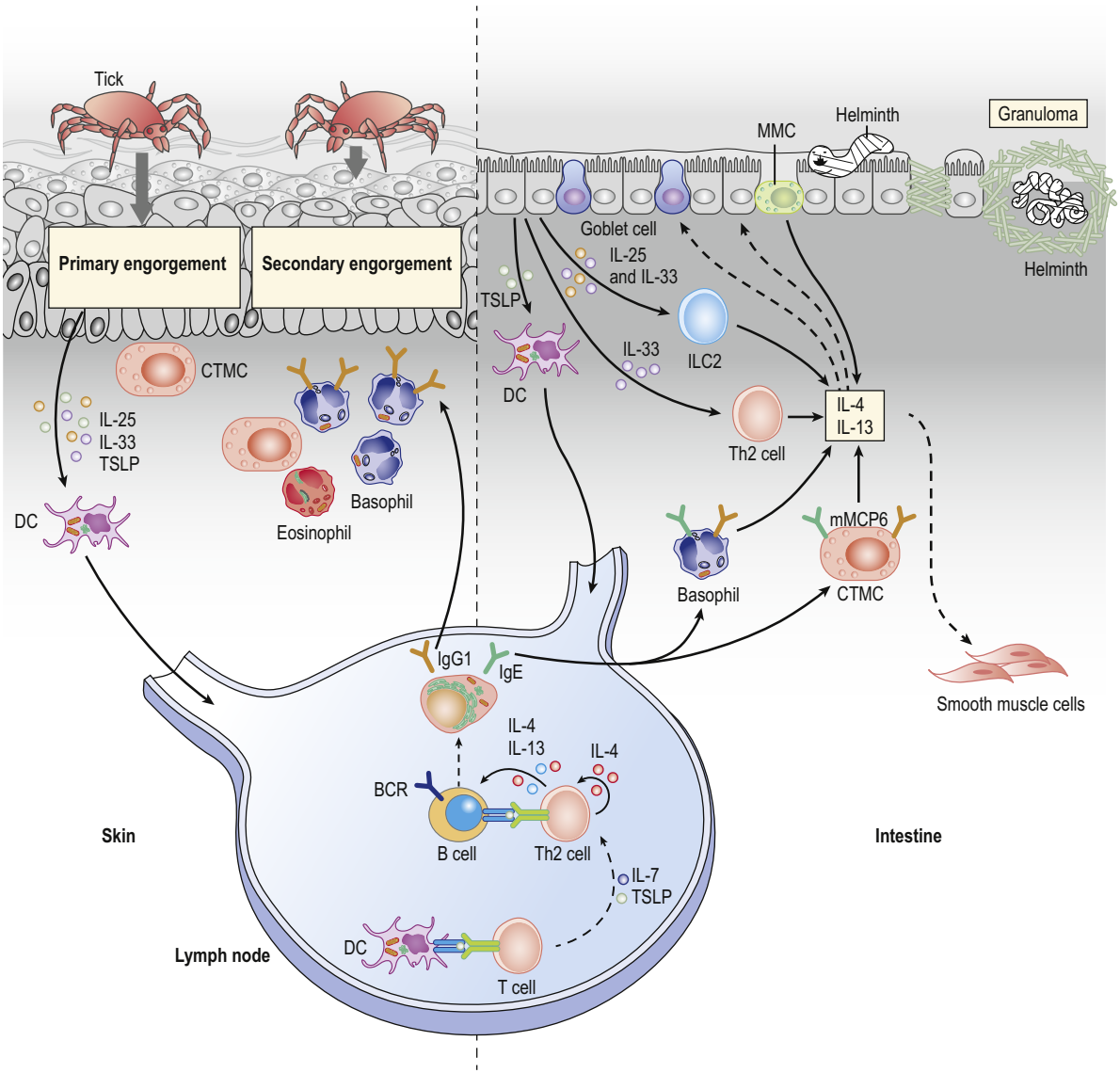


Fig. 7.40 Parasites, like any antigen, are taken up by phagocytes such as dendritic cells (DCs), which, selectively, through particular toll-like receptors, activate signals that polarize T-helper cells towards Th2 through activation of STAT5. IL-4, IL-5 and IL-13 produced by the DCs themselves or by neighbouring tissue and other cells (e.g. mast cells) help to drive the response down a Th2 route. The process is amplified by continued production of Th2 cytokine, the formation of a parasite-enclosed granuloma and other cytokines such as IL-33, released from damaged and dying cells. (From Voehringer, 2013.)

and partial genome sequencing allows molecular typing of genetic loci to be much more sophisticated and allows closer matching than ever before. In fact, the problem now is determining the cutoff for mismatched pairing.

Most donor solid organ, vascularized and bone marrow transplants induce a strong early direct alloimmune response, which requires long-term therapeutic immunosuppression to be controlled if the graft is to survive. Graft rejection is mediated by $CD8^+$ T cytotoxic T cells in this circumstance. Simultaneously, as donor leucocytes die, an indirect alloimmune response develops, which is slower but more insidious and is mediated by host $CD4^+$ T cells activating innate

immune cells such as macrophages and other cells. Many types of chronic rejection, for instance of kidney allografts, are the result of indirect allorecognition and host $CD4^+$ T-cell activity.

Corneal transplantation differs from other forms of skin and solid organ transplantation in the number of donor antigens present in the graft, since there are no leucocyte-rich blood vessels in the central cornea, and even leucocytes that are resident in the donor corneal stroma are generally not equipped to activate the host T cells directly (i.e. in the quiescent cornea they do not express high levels of MHC antigen).

Further information on the immune response to corneal allografts is available online at https://expertconsult.inkling.com

Rejection of corneal grafts is thus mainly through indirect allorecognition. However, if rejection does occur, regrafting with a second donor who expresses some MHC antigens that are similar to the first donor will induce a strong $CD4^+$ T-cell memory response and cause rapid rejection of the second graft. For this reason alone, there is a strong case for matching corneal grafts, although this practice is not currently routine because of the assumed high success rate of unmatched corneal grafts.

The role of other cells in the immune response to allografts is increasingly recognized. Th cells, both Th1 and Th17 cells, almost certainly have a role via CD40–CD40L interactions and may perhaps involve cross-presentation of antigen. In addition, 'danger signals' that activate cells of the innate immune system such as macrophages play a significant role and might develop a level of innate immune memory termed trained immunity if prolonged inflammation and continuous macrophage recruitment occurs. Surgical technique influences the acceptance or failure of grafts, both for vascularized grafts, such as kidneys, and even more so for corneal grafts. Donor-specific anti-HLA antibodies have the potential to activate complement, for instance in living-donor kidney allografts. In apparently accepted clear corneal grafts, rejection can be triggered by an unrelated event such as herpes simplex virus (HSV) infection in the graft or even by a simple procedure such as corneal suture removal. Indeed, it has been shown that failure of grafts in HSV infection is due to an exaggerated memory T cell response to HSV rather than alloantigen (Kuffova, et al. 2016). Presumably, activation of innate immune cells such as macrophages and dendritic cells via the inflammasome and antigen presentation with co-stimulation is sufficient to lead to activation of previously primed or resident memory T cells leading to rejection.

Tregs also play an important role in the overall immune response to allograft. In the initial inflammation associated with the surgical trauma, both naive T cells and Tregs extravasate into the tissues. In addition, APCs migrating to the draining lymph node activate both Tregs and non-Tregs, which home to the graft with different kinetics. The balance between T-effector cells (Th1/Th17) and Tregs determines the survival of the graft.

Finally, rejection of xenografts (the first ever corneal graft performed was a "goat to human" xenograft) is mediated by preformed antibody and complement and generates hyperacute rejection, induced through failure of complement regulatory protein to act across species barriers.

Tumours induce immune responses but not very effectively

Clinicopathological studies have long indicated that tumours induce some form of immune response in the host. This is based mainly on observations of T-cell, macrophage and NK-cell infiltration of tumours independent of whether there is any inflammatory response or tissue necrosis (tumour infiltrating lymphocytes, TIL). Experimental studies of chemically induced tumours have also shown that transplantation of resected tumours to the original host, or to a host previously sensitized to tumour antigens, leads to rapid rejection of the tumour, but not when the tumour is transplanted to a syngeneic naive host. Rejection is tumour-specific. Thus, tumours possess specific antigens and these induce a MHC class I-restricted T-cell cytotoxic response.

The concept that cancer cell mutations occur very frequently but are held in check from developing into full-blown tumours is long-standing and indeed NK cells are specially equipped to kill and delete malignantly transformed cells as well as infected cells (see pp. 386 and 387). Antigen-specific cytotoxic T cells (mostly $CD8^+$ but also $CD4^+$ killer cells) may also play this role. In addition, ILCs (see p. 387) that reside particularly at barrier sites and interact with the microbiome also indirectly shape immunity to tumours.

Tumour antigens may be tumour-specific, i.e. only expressed on that tumour, or they may be tumour-associated, in which case they are also expressed on normal cells. Most tumour antigens are identified in transplantation-type experiments, such as that described previously, in which cloned cytotoxic T cells are generated. This allows identification of the tumour-derived peptides and the genes that regulate their expression in the tumour. Tumour antigens are believed to be recognizable by cytotoxic T cells in this manner, owing to antigenic mutation that converts the normal self-protein, which would induce tolerance, to a non-self-protein, which is recognized as foreign and induces immunity. Thus, some tumour antigens are products of normal cells such as tyrosinase in melanoma cells.

IMMUNE RESPONSE TO CORNEAL ALLOGRAFTS

Corneal grafting is unusual. Successful corneal allografts between unrelated donors and recipients have been performed since the early part of the 20th century, long before HLA antigens were discovered. This was attributed to immune privilege of the eye, and specifically to the lack of MHC^+ 'passenger' leucocytes in the normal cornea. However, it is now accepted that the apparent success rate of corneal graft applies only to 'low-risk' recipients, namely patients with opaque corneas without significant chronic inflammation or infiltration of the cornea with abnormal blood vessels. Work from Australia has re-examined the question of the privileged status of the cornea as an explanation for protection against rejection when used as an unmatched allograft and without concomitant immunosuppression. In fact, the 5-year survival of unmatched corneas in high-risk recipients, i.e. those patients who have ongoing low-grade inflammation or blood vessel infiltration as in herpes simplex keratitis (see p. XXX), is less than for renal allografts, where steady improvements in survival of allografts has continued (Williams and Coster, 2007). Recent experimental studies of high-risk graft rejection have shown that sensitization, even to a single corneal minor (i.e. non-MHC) antigen, in high-risk grafting (particularly in regraft procedures where the first graft failed) can be highly immunogenic and leave a strong immunological memory that drives rejection. To reduce the risk of rejection caused by memory T cells sensitized to the antigens present in the first graft, HLA matching of all grafts should be performed so that a second graft has the fewest HLA matches to the first donor and thus maximizes the chances of acceptance of the regraft (Vitova et al., 2013).

However, many tumour antigens are produced by oncogenes, genes that are implicated in the cell cycle and in cell differentiation. These include *p21 ras, HER-2/neu*, the very important *p53* gene and others not normally expressed on cells, such as the MAGE series of proteins expressed by melanoma cells including cells of the choroid. Several viral genes may also be expressed in tumours, including those for the SV40 T antigen, the human papillomavirus E6 gene and the EBV antigen. Certain B-cell tumours are indirectly linked to EBV genes, particularly in the presence of immunodeficiency, and appear to involve the translocation of the *myc* oncogene to the immunoglobulin locus. Retroviruses also have the potential to induce malignant transformation of normal cells including the *src*, the *myc* and the *k-ras* gene. The human T-cell lymphotropic virus-I (HTLV-1) gene is also implicated in certain aggressive T-cell tumours and interestingly is also involved in some forms of intraocular inflammation (uveitis).

Certain tumour antigens are recognized by antibodies as well as by cytotoxic T-cell lines. These include the oncofetal antigens, carcinoembryonic antigen and α-fetoprotein. Other tumour-associated antigens are linked to tumours such as the surface glycoprotein MUC-1 in breast cancer, S-100 in neural crest cell tumours including malignant melanoma, and cytokeratins in epithelial cell tumours.

Tumour infiltrating lymphocytes (TIL) are conventionally considered to be effector cells against the tumour at least in some instances. Indeed, the process of immune surveillance in which T cells and antigen-presenting cells migrate through the tissues detecting altered self-antigens applies particularly to tumours and is necessary for the initiation of the anti-tumour response. TIL include both $CD4^+$ and $CD8^+$ cytotoxic cells and the former probably supply essential cytokines to the latter to promote tumour-killing ability. Some tumours aberrantly express MHC class II antigens plus co-stimulatory molecules; they may directly present tumour antigens to $CD4^+$ T cells and initiate an immune response *in situ*.

NK cells may also be important in killing tumours (as well as removing individual transformed cells as they arise), particularly those that have been induced by viruses. IL-2-activated NK (lymphokine-activated killer, LAK) cells have a markedly enhanced ability to lyse tumour cells and they are in trials as immunotherapeutic agents. Macrophages are also important in tumour killing, usually via release of TNF-α, a cytokine that was first identified by its ability to induce necrosis in tumours. Tumour cells appear to be unable to synthesize superoxide dismutase, which is required to protect cells from the TNF-α-induced release of cytotoxic superoxide free radical.

Despite the variety of mechanisms for tumour killing, many tumours evade death. Tumours utilize a variety of strategies: e.g. they downregulate MHC class I, thus inhibiting cytotoxic T-cell killing; although they express foreign antigens, they may not express co-stimulatory molecules necessary for T-cell activation; they may secrete immunosuppressive cytokines such as TGF-β; tolerance to tumour cells may occur if the tumour antigen is expressed during the neonatal period; antigenic modulation of the tumour may occur if the antigen binds non-complement binding antibody; or the tumour antigen may be prevented from gaining access to the immune system by a dense glycocalyx on the cell surface. The immune response to tumours is more likely to be modulated by any or some of these mechanisms, thus explaining why tumour immunity is imperfect and cancer remains a major cause of death.

Special mention must be made of tumour evasion of the immune system by promoting tolerance rather than immunity. In particular, Tregs comprise a significant component of TILs and prevent development of an effective antigen-specific anti-tumour T effector cell response. Tregs express immunoinhibitory/co-inhibitory molecules such as PD-1 and CTLA-4, which are known as 'checkpoint' molecules. New treatments using checkpoint inhibitors, i.e. monoclonal antibodies or fusion proteins that cancel the inhibitory effect of these molecules, are changing the outcomes of some patients by lifting the Treg control and allowing natural anti-tumour immune responses to clear the tumours.

There are also other cells that suppress the immune response. Myeloid cells expressing either granulocyte ($Gr1^+$) or monocyte ($Lys6C^+$) and termed myeloid-derived suppressor cells (MDSC) are TIL recognized to have a significant immunosuppressive effect, not only against tumours but also against any immune response. MDSCs are thus being proposed as therapy for autoimmune and other immune-mediated disease. In contrast, other cell-based therapies are in the vanguard of cell vaccination therapies for cancer treatments: these are based on the use of activated, pro-inflammatory DCs loaded with tumour-specific antigens to enhance strong anti-tumour responses and have been introduced for certain forms of prostate cancer. The main difficulty with this approach is identifying peptides specific for the tumour that are sufficiently immunogenic.

The eye and the immune system

The eye participates in all aspects of immune responses like any other tissue, but the immune response is modulated by the cells and tissues of the eye. In this respect, the eye (and the brain) are regarded as 'immunologically privileged', a concept that has now been broadened to be included as an additional mechanism of tolerance induction. Both the innate and acquired immune systems function in ocular defence mechanisms.

The innate immune system and the eye

Reference has already been made to the several physical and chemical protective barriers to ocular infection included in the blink reflex and the lids and the components in the tears such as lysozyme, lactoferrin and complement (see also Chapter 4, p. 207). Lysozyme is effective against Gram-negative bacteria and certain fungi but is ineffective against Gram-positive organisms such as *Staphylococcus aureus*. Lactoferrin and transferrin, however, are more effective in defence against Gram-positive bacteria because they bind iron, an essential cofactor for eukaryotic as well as prokaryotic cell growth. In addition, tears have specific anti-adhesive properties for bacteria and therefore inhibit bacterial attachment and invasion of the ocular surface and are incidentally important for prevention of contact lens contamination.

Tears also contain polymorphonuclear leucocytes (PMNs), which increase in number when the lids are closed for prolonged periods, e.g. during sleep. The anti-adhesive properties of tears extend to these cells, ensuring that they pass through the lacrimal passages and do not penetrate the corneal surface. Many of these leucocytes pass directly from the conjunctival vessels through the epithelial layers into the tear film. PMNs contain numerous antibacterial protein enzymes, including proteinase 3, myeloperoxidase (which generates free radicals), calprotectin, β-lysin and the cathepsins.

Tear lipid also has an antibacterial effect. This applies to both short- and long-chain fatty acids, the former affecting surface properties of the bacterial cell membrane and the latter having a direct effect on metabolism. In addition, cationic peptides, such as the defensins, as well as surfactant protein-D are part of the conjunctiva-associated lymphoid system (CALT, see p. 422).

The cells of the ocular surface also express, or can be induced to express, many of the innate immunoreceptors such as TLRs and are involved in many of the cornea and conjunctival innate immune responses to infectious agents. All three types of T cells (αβTCR, γδT cells and ILCs) are present in the conjunctiva some of which secrete IL22. More importantly, the conjunctiva contains numerous stromal and intraepithelial DCs (similar to Langerhans' cells in the skin), which act as APCs in the eye-draining lymph nodes in the afferent limb of the immune response. It is thus possible to become sensitized to environmental antigens and allergens via the conjunctiva (Fig. 7.41).

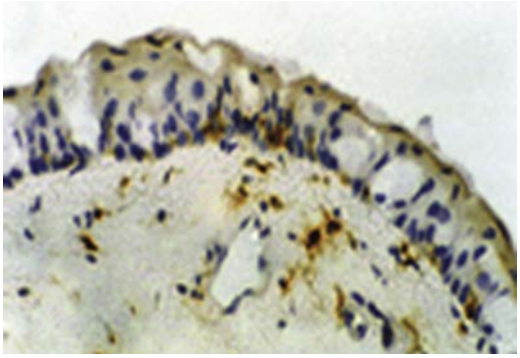


Fig. 7.41 Conjunctival dendritic cells (DCs) in the rat. Both intraepithelial and stromal DCs are shown expressing major histocompatibility complex class II (immunoperoxidase, *dark brown* stain).

The adaptive immune system and the first line of defence in the eye

Tears contain immunoglobulins such as IgA (see p. 406) and occasional specific immune cells such as lymphocytes. IgA is produced by B cells in the lacrimal gland and secreted as sIgA in the tears. Remarkably, the level of IgA in the tears is influenced by the gut microbiota in a gut-eye axis indicating a systemic effect on ocular IgA production. In addition, the lacrimal gland produces other immunosuppressive cytokines such as TGF-β. TGF-β-deficient and thrombospondin (TSP) deficient mice exhibit extensive ocular surface. TSP-activated TGFb via the peptide KRFK and topical treatment with this peptide prevent ocular surface pathology including infiltration of Th1 and Th17 cells, thus indicating the importance of tear TGFb in maintaining ocular surface homeostasis (Soriano-Romani et al., 2018).

Conjunctival sensitization to environmental antigens is best demonstrated by the fact that these do not elicit responses in normal individuals but do so in allergic (Th2) type responses. In these people, common antigens including pollens, house dust mite and animal dander (particularly cat dander) elicit a characteristic response. Although the effect is produced locally, the initial sensitization is a systemic one via activation of Th2 cells in the draining lymph nodes (see pp. 442–443). This results in local conjunctival mast cells becoming loaded with antigen-specific IgE, which renders these cells acutely sensitive to re-exposure to antigen. The chemokine CCL2 appears to be partly involved in this process of mast cell degranulation in giant papillary conjunctivitis (GPC) and the receptor CCR7 is active in directing conjunctival DC migration during this process.

If the sensitization is long-lived and recurrently activated with frequent mast cell degranulation episodes, it can lead to chronic inflammation. Thus, seasonal allergic conjunctivitis can progress to perennial conjunctivitis and, in atopic individuals, may manifest as severe atopic conjunctivitis and/or vernal keratoconjunctivitis (VKC). Much of the damage in VKC is thought to be induced

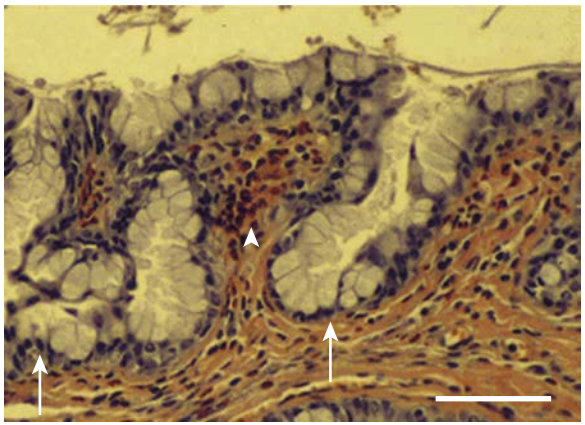


Fig. 7.42 Histopathology in allergic conjunctivitis in an experimental model. Note the pseudotubular structure formations in conjunctival epithelium (*arrows*) and infiltration of eosinophilic cells in the subepithelial stroma *(arrowhead)*. Bar = 100 μm. (From Hara et al., 2012.)

by eosinophils, because eosinophil cationic protein levels in tears are greatly increased in both atopic and VKC (Fig. 7.42). Giant papillary conjunctivitis is a feature of VKC as well as contact lens-induced allergic disease and the papillae have been shown to contain B-cell follicles as well as extensive T-cell infiltration. M cells, characteristic cells involved in mucosal inflammation in the gut, are also prominent (Zhong et al., 2007). In addition, different types of allergic conjunctivitis are associated with different patterns of cytokine production and Th-cell patterns. Th2-like profiles are linked to VKC, while Th1 is associated with atopic keratoconjunctivitis. There may also be a defect in histaminase function in allergic eye disease causing prolonged histamine effects after mast cell degranulation. However, it is likely that T, especially Th2, cells play the most prominent initiating role in most forms of allergic conjunctivitis.

The corneal surface also contains populations of intraepithelial and stromal leucocytes, some of which have the characteristics of dendritic APCs (see p. 427). The role of these cells in protective immune responses is not clear but there is evidence that, like the conjunctival cells, they can capture antigen and transport it to the draining lymph node. In addition, they are likely to respond to cytokines produced by surrounding epithelial cells and resident tissue macrophages in response to TLR engagement by common invading microorganisms such as herpes viruses and bacterial pathogens. This may be a direct effect of the pathogen; for instance, the obligate intracellular parasite *Chlamydia*, which is the cause of the worldwide blinding disease trachoma, signals through TLR2 inside epithelial cells. Indeed, the possibility of regulating host–pathogen interactions has been tested in an experimental model of *Pseudomonas* keratitis in which 'silencing' of TLR9 expression was achieved using small interfering RNA for this receptor.

Progressive ocular surface disease

Certain other ocular surface diseases occur that do not appear to be allergic in nature but are considered to be immune-mediated if not autoimmune in their pathogenesis. These include cicatrizing conjunctival disorders, various forms of 'melting' corneal disorders and a number of scleral and orbital inflammations.

Cicatrizing disease of the conjunctiva

Subconjunctival fibrosis occurs in certain rare disorders such as benign mucous membrane pemphigoid and the Stevens–Johnson syndrome. Both are considered to be autoimmune in nature by virtue of the detection of antibodies to basement membrane components such as integrins. Pemphigoid is characterized by progressive cicatrization of the conjunctiva stroma leading to severe shallowing of the fornices (see Chapter 9, p. 506). In this condition, specific autoantibodies against conjunctival epithelial β_4 integrin, a component of the hemidesmosome (see p. Chapter 4, pp. 184–186), have been identified. This is in contrast to bullous pemphigus, a skin disorder that does not affect the conjunctiva but is characterized by widespread areas of epithelial detachment and in which antibodies against other proteins such as desmoglein and plectin may be detected. Dysregulation of TGF-β in conjunctival cells has been reported in ocular pemphigoid. The inflammation-associated subconjunctival fibrosis is associated with upregulation of the aldehyde dehydrogenase enzyme, ALDH1A3 involved in the production of retinoids and inhibitors of this this enzyme, such as disulfiram, prevented conjunctival fibrosis experimentally (Ahadome et al., 2016). This condition would appear to be mediated through a Type 2 immune response, involving M2 macrophages.

The Stevens–Johnson syndrome has similar appearances but is much more acute, although self-limiting. This condition is normally associated with drug administration in which the drug is considered to act as a hapten (see p. 405). The severity of the condition is limited in its effects by the degree and duration of exposure to the drug.

Keratitis and 'melting' corneal ulcers

Many forms of keratitis are considered to be immune-mediated, including postherpetic disciform keratitis (see Chapter 8, p. 474) in which residual herpes simplex virus antigen may play a role. Fungal infections such as aspergillus and fusarium induce a strong innate immune response

involving IL6/23 and IL17 with marked neutrophil activation. Experimentally, thymic stromal lymphopoietin TSLP and IL-7-like protein activated corneal DC to induce a Th2-like anti-fungal response in Apsergillus infection (Sun et al., 2018).

In addition, certain debilitating corneal diseases characterized by peripheral corneal thinning and ulceration, sometimes leading to perforation, may be autoimmune or at least immune-mediated, mainly because they represent vasculitic complications of classic autoimmune diseases such as rheumatoid arthritis and also because of the absence of overt infection. In this condition, reductions in Tregs cells have been reported as well as increases in B cells and anti-corneal epithelial antibodies. Some forms of peripheral ulcer such as Mooren's ulcer have been linked to a cornea-associated antigen, calgranulin, also found in peripheral blood neutrophils and filarial nematodes. These disorders occur just distal to the source of corneal epithelial stem cells at the limbus, which suggests a defect at this level. Immunohistology shows predominantly T cells and macrophages infiltrating to peripheral cornea and conjunctiva (Shinomiya et al., 2013).

Inflammatory disorders of the orbit and sclera

While many inflammatory conditions of the sclera (scleritis) are considered to be autoimmune, or at least immune-mediated, it is important to consider infectious aetiologies such as syphilis and other bacterial causes. Spontaneous (autoimmune) inflammatory disorders of the episcleral tissue and the sclera (episcleritis and scleritis) represent a type IV immunopathological disease with close association with rheumatoid arthritis and a similar pathogenesis. Scleritis occurs in rheumatoid arthritis patients who have vasculitis. Rare histological studies have shown activated $CD4^+$ T cells in the lesion in a perivascular location, in addition to macrophages. $CD8^+$ T cells have also been demonstrated, but in fewer numbers. There is a prominent vasculitic component to the disease with extensive necrosis of the scleral layers. Typical granulomatous lesions are a feature of this condition (nodular scleritis). However, not all of these lesions are T-cell-dominated, because B-cell 'follicles' have been identified in some cases, suggesting the development of tertiary lymphoid structures (see pp. 422–424). In addition, increased matrix metalloproteinase activity has been detected in these lesions (e.g. collagenase and stromelysin) (Fig. 7.43). Although the antigen(s) for scleral inflammation has not been identified, it is presumed to be a component of the extracellular matrix such as dermatan sulphate proteoglycan or type 1 collagen. Scleritis most commonly affects the anterior sclera but if it affects the posterior sclera, it is more difficult to diagnose. In addition, it may be mistaken for a less well-defined group of orbital inflammatory disorders known as pseudotumour of the orbit, for which the aetiology remains obscure but which responds to systemic steroid therapy. Recently, orbital

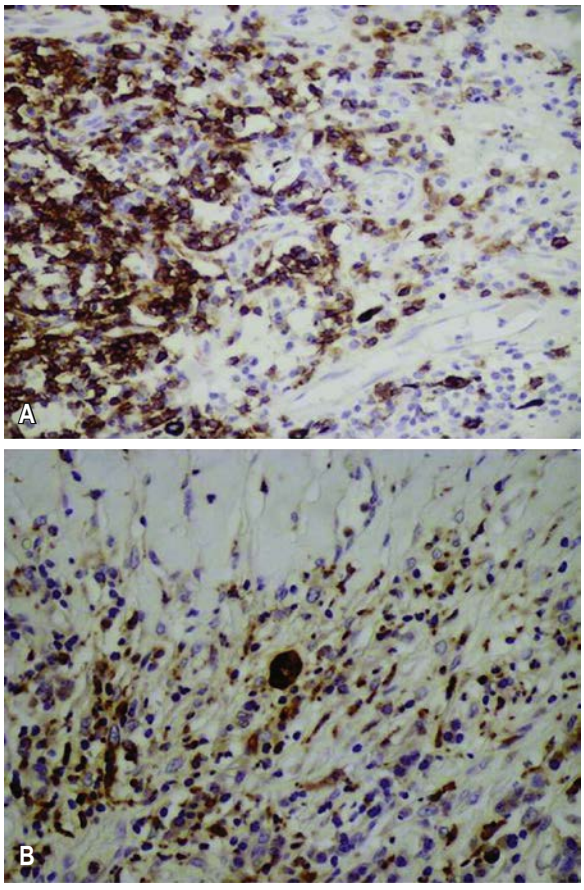


Fig. 7.43 Immunopathology of scleritis. (**A**) Autoimmune scleritis. Infiltration of CD20-positive B cells in the autoimmune group (original magnification ×100). (**B**) Necrotizing scleritis. Large numbers of CD68-positive macrophages in the idiopathic necrotizing scleritis groups (original magnification ×80). (From Usui et al., 2008.)

inflammatory lesions are reported to occur in an inflammatory disease characterized by extensive infiltration of IgG4-secrteing plasma cells and B cells, which affects several organs (Fig. 7.44). The disease is characterized by an extensive lymphadenopathy and 50% of cases are EBV positive.

A specific form of pseudotumour known as orbital myositis, in which an acute inflammatory swelling of a single ocular muscle occurs, is particularly responsive to steroid therapy. The autoantigen in this disorder is assumed to be a component of the ocular muscle.

Swelling of the orbital muscles also occurs in dysthyroid eye disease, causing proptosis and exophthalmos, but in this condition all four muscles are variably involved. This disorder is closely linked to Graves' disease of the thyroid in which thyroid autoantigens such

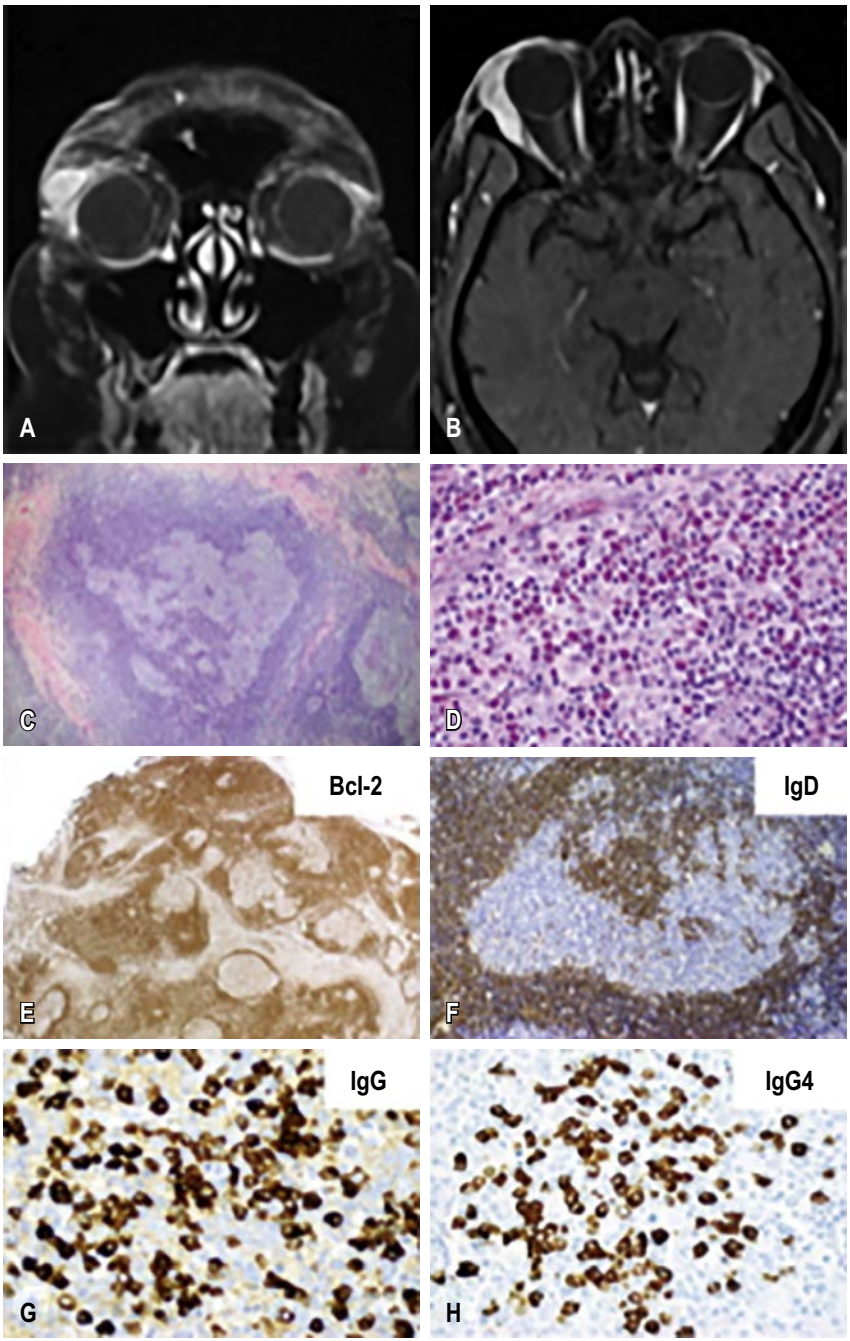


Fig. 7.44 Radiologic and histologic features of PTGC IgG4-ROD (Case 1). Post-contrast MRI scans in coronal (**A**) and axial (**B**) sections show diffuse, right lacrimal gland enlargement with homogeneous contrast enhancement. Microscopically, the lacrimal gland shows fibrosis and enlarged, irregular lymphoid follicles resembling tertiary lymphoid organs (see page 424) with prominent mantle zones invading germinal centres (**C**). Numerous eosinophils and histiocytes are also present in interfollicular areas (**D**). Bcl-2 and IgD immunohistochemical stains highlight mantle zone lymphocytes within germinal centres (**E, F**). IgG4:IgG ratio is over 70 % and >100 IgG4+ plasma cells per HPF are noted (**G, H**). (From Vizcaino et al., 2018.)

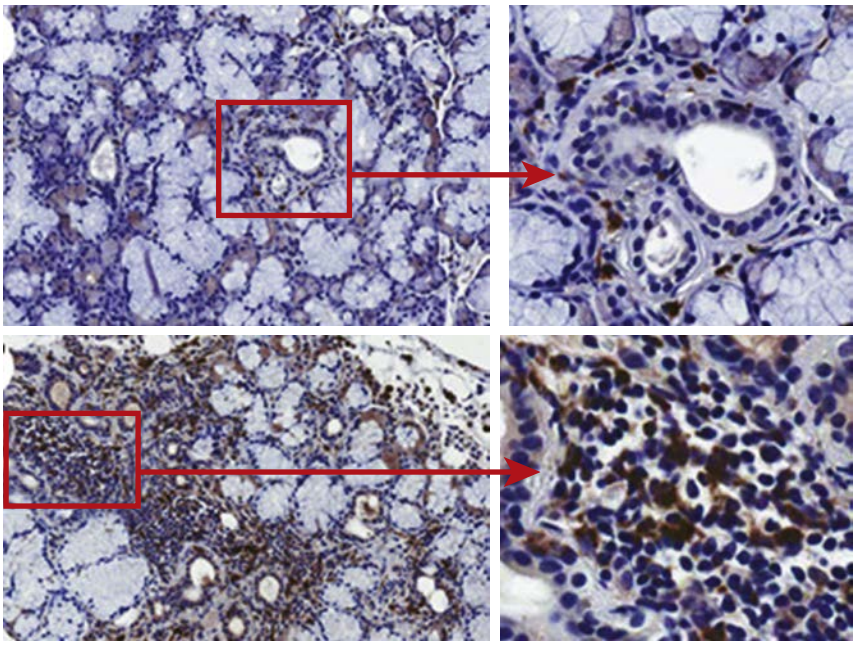


Fig. 7.45 *Upper panels:* normal mouse salivary gland has only a few resident macrophages (immunoperoxidase staining for CD68) (original magnification ×20, ×40 *right panel*). *Lower panels:* diseased salivary gland is heavily infiltrated with CD68+ macrophages (magnification ×20, ×40). (From Greenwell-Wild et al., 2011.)

as thyroglobulin and the thyroid-stimulating hormone receptor are implicated. Patients with dysthyroid eye disease have circulating lymphocytes that react with ocular muscle cell membrane antigens. However, thyroglobulin does not appear to be the important antigen for ocular muscle damage. Instead autoantibodies to thyroid-stimulating hormone receptor (TSHr) are considered pathogenic although both TSH stimulatory and inhibitory antibodies have been described. Whether TSH functional effects and disease in the orbit are related is not clear. Models of thyroid ophthalmopathy using T cells sensitized to the thyroid-stimulating hormone receptor have been reported. Immunological studies have shown that the T-cell infiltrate is almost exclusively Th1 in type, with secretion of IL-2 and IFN-γ.

Inflammatory disease of the lacrimal gland may be primary (autoimmune), as in Sjögren's syndrome, or secondary, as in sarcoidosis, in which the aetiology is unknown. In both disorders there is a deficiency of tear secretion that produces a secondary keratoconjunctivitis (keratoconjunctivitis sicca or the dry eye syndrome), common in the elderly. Primary Sjögren's syndrome involves other secretory glands such as the salivary glands and is characterized by specific autoantibodies against ribonucleoproteins (antiRo and antiLa) whose role in the pathogenesis of the condition is not clear. Biopsies of salivary gland tissue have shown a predominant T-cell infiltrate, but with little evidence of T-cell activation (as evidenced by the lack of IL-2 receptor (CD25) expression). In contrast, conjunctival biopsies from patients with Sjögren's syndrome have shown significant T-cell infiltrates with activation markers. In addition, extensive adhesion molecule expression has been detected in the lacrimal gland, not only on endothelial cells but also on the acinar epithelial cells. Both VCAM-1 and E-selectin are upregulated on the endothelium, indicating that this chronic disease is in a state of persistent activation. Eventually these glands undergo involutionary atrophy. Recent extensive microarray studies have shown that mammalian chitinases, innate immune products of macrophages with specificity against chitin-containing pathogens, are highly active in severe acinar gland destruction of Sjögren's syndrome (Fig. 7.45). In a Sjögren's-like syndrome in mice infected with cytomegalovirus, NK+ cells were protective in removing activated T cells, a process that was dependent on the TNF-related apoptosis inducing ligand (TRAIL).

The eye as a privileged site: what does this mean for the survival of corneal grafts?

The eye has been considered immunologically 'different' since the first corneal graft in a human was performed more than 100 years ago and shown not to be rejected. However, the eye as an immunologically privileged site was

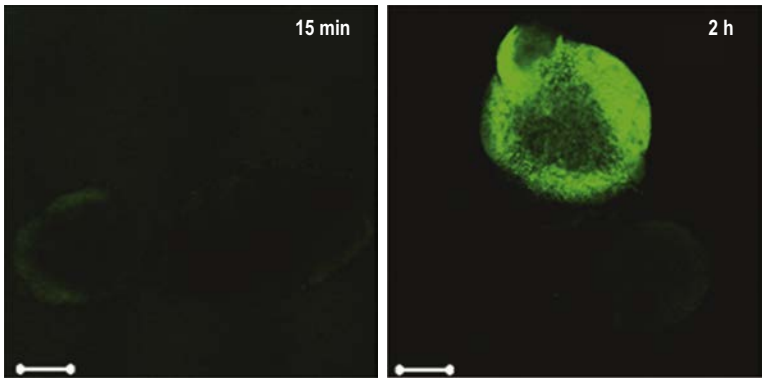


Fig. 7.46 Immunofluorescence of submandibular draining lymph node from a mouse after application of a *green fluorescent-labelled protein* to the surface of the cornea. Two hours after application, the protein was readily detectable in the lymph node. (From Dang et al., 2013.)

not formally recognized until 1945 when Peter Medawar in his seminal studies on alloreactivity in rejection of skin grafts showed that skins grafts placed in the anterior chamber of the eye or the brain were not rejected: i.e. these tissues demonstrated *tolerance* (see p. 443) to foreign antigens. The basis for this phenomenon was attributed to the avascularity of the cornea in the case of corneal grafts and to the lack of lymphatic drainage for intraocular structures. However, alloreactive T cells specific for corneal antigens can be detected in the circulation and appear to be generated in the secondary lymphoid tissues (SLO), indicating that despite the apparent lack of ocular lymphatics, foreign antigens in the cornea (and indeed in other ocular structures such as the retina) are clearly recognized in the SLO. Direct testing of antigen trafficking to the submandibular draining lymph node using green fluorescent-labelled proteins applied to the abraded corneal surface has shown that ocular antigen does reach an eye-specific draining lymph node via lymphatics (Fig. 7.46).

The concept of immune privilege as applied to the eye (and brain) has been modified in the context that all tissues have the potential to modulate the immune response (tissue-based tolerance, Matzinger and Kamala, 2011, see p. 381) There is common consensus that tissues behind by the blood–retina and blood–brain barriers (the retina and the brain parenchyma) and that are not traversed by migratory bone marrow-derive innate immune cells are privileged in the sense that immune cells are excluded from these tissues in the normal state. However, tissues such as the uveal tract and the meninges are highly vascular and support robust immune responses similar to other vascular tissues not privileged in this way. Similarly, the mesenchymal tissues such as the cornea, sclera and ocular muscles are not different in their ability to respond to immune challenge from other similar tissues such as tendons, muscles, acinar tissues, liver, etc. However, the level of immune reactivity at these sites is reduced compared with immune responses at external barrier sites such as the skin and the gut or lung mucosal tissues, which respond most robustly to challenge in part because of their extensive complement of ILCs. The highly privileged status of the retina and brain, however, renders them susceptible to latent infection, which, when it reactivates, generates uncontrolled damaging tissue destruction (Forrester et al., 2018).

Further information on immune privilege in the central nervous system is available at https://expertconsult.inkling.com

Tissues that have a moderate ability to mount immune responses to challenge achieve this level of response through tissue expression of various inhibitory ligand-receptor pairings such as Fas–FasL, PD1–PDL1 and TRAIL–D5 (see p. 437), which are immune checkpoint molecules. Importantly, Tregs (both $CD4^+$ and $CD8^+$ Tregs) both of which are generated in the SLOs, particularly the spleen, contribute to this immune regulation exerted by tissues.

Despite these risky properties of the ocular tissues, bacterial infection during intraocular surgery is remarkably infrequent. This has been attributed in part to the direct bacteriostatic properties of the aqueous humour, which have been shown to inhibit bacterial growth *in vitro*. The nature of this activity is not known but several antibacterial proteins are present in the aqueous humour, including complement, immunoglobulin, defensins and β-lysin. In addition, it is dependent on the size of the bacterial inoculum and the virulence of the organism.

Intraocular inflammation

A common cause of visual loss is intraocular inflammation. This is a blanket term for the many forms of uveitis and conditions such as retinitis and even optic neuritis. It may also include inflammation of the ocular mesenchymal tissues such as the cornea and sclera because they are frequently accompanied by inflammation in the uveal tract.

IMMUNE PRIVILEGE IN THE CENTRAL NERVOUS SYSTEM (CNS)

As indicated previously, CNS immune privilege comes with risks. While it is important for the preservation of sight and life that immune responses within the CNS do not get out of control when inflammation occurs, the modified immune defences that the retina and brain have may not be enough to prevent an exaggerated response. Thus, certain organisms can proliferate or reactivate with less of a trigger than if they were activated in other sites. For instance, toxoplasmosis is a common cause of ocular inflammation but is still relatively infrequent given that almost a third of the population have been infected, mainly through eating infected meat. This indicates that for most of the population toxoplasmosis is controlled and indeed, during the initial infection in the gut, the organism is rapidly cleared almost completely. Some organisms (tachyzoites) escape into the bloodstream carried within infected monocytes and find their way to sites of immune privilege where they can lie dormant and never cause a reaction. However, if the immune defences are diminished, as occurs in HIV-positive patients or patients taking immunosuppressive drugs, *Toxoplasma* organisms can escape immune control and reactivate, causing a severe tissue destructive inflammation including in the brain and retina. Similar pathogenesis occurs with other organisms such as cytomegalovirus, herpes simplex, *Mycobacterium tuberculosis* and many others (Forrester et al., 2018).

Intraocular inflammation may be exogenous, in which the organism is clearly evident, as in bacterial endophthalmitis (see Chapter 9, p. 500) or herpes zoster retinitis, or there may be no obvious infectious agent. Most such cases are described as idiopathic or non-infectious uveitis and may affect the anterior segment (iridocyclitis, inflammation of the iris and ciliary body) or posterior segment (posterior uveitis). Posterior uveitis may take many clinical forms such as multifocal choroiditis and retinal vasculitis and formal definition and classification of the numerous entities is work in progress (Standardization of Uveitis Nomenclature (SUN) committee).

Further information on classification of intraocular inflammation is available online at https://expertconsult.inkling.com

Non-infectious intraocular inflammation is thought to be autoimmune or at least immune-mediated. As described earlier (see page 430), around 50% of cases of acute anterior uveitis have a strong association with HLA-B27 antigen as well as ERAP1 and are linked to ankylosing spondylitis. They are also associated with low-grade Gram-negative enteric infection with organisms such as *Yersinia*, *E. coli* and *Klebsiella*. Posterior uveitis is much less closely linked to MHC class I antigens, except for certain well-defined syndromes such as birdshot retinochoroidopathy (HLA-A29) and Behçet's retinal vasculitis (HLA-B51 in oriental and Middle Eastern people). Vogt–Koyanagi–Harada disease has been closely linked to HLA-DR4 (subtype DRB1 0405; see Box 7.11) and there is a similar association with sympathetic ophthalmia.

Posterior uveitis, however, has close similarity to certain experimental $CD4^+$ T-cell-mediated uveoretinal inflammations (experimental autoimmune uveoretinitis, EAU) in which the autoantigens have been well defined. Most of these are derived from the outer retinal layers and include the visual protein rhodopsin. The relationship between EAU and non-infectious posterior uveitis is, however, tantalizingly tenuous because patients with these diseases do not have significantly raised levels of antibodies to retinal antigens, although some patients manifest T-cell responses to these antigens.

As a result of the difficulty in meeting the criteria for autoimmunity, many forms of uveitis are now being considered as part of auto-inflammatory diseases, originally defined as episodes of inflammation in the absence of B- or T-cell activation. Uveitis is part of many such syndromic disease. The possibility that most, if not all, non-infectious cases of uveitis have an infectious agent as the initiating cause has also been raised, with the concept that an out-of-control persistent immune response sustains recurrences of inflammation after the infection is cleared or becomes latent (Forrester et al., 2018).

The role of the immune system in ageing and degeneration in the eye

The recent understanding of the innate immune system, its inducers and receptors and its close link with the adaptive immune system, has revealed the role of immune processes in several ocular and non-ocular diseases that were previously considered genetic, developmental, metabolic and age-related/degenerative and even in cancer. This includes many conditions such as diabetes and its complications, obesity, neurodegenerations such as Alzheimer's and classic genetic diseases such as retinitis pigmentosa. Inflammation induced via PAMPs, DAMPS and PRRs (see p. 381) with activation of inflammasomes and secretion of IL-1 and its many family members is now known to play a significant part in the pathogenesis of many of these diseases, initiated by the original abnormality, be it developmental, viral induced, genetic or other. For instance, diabetic retinopathy is now recognized to be the result of leucocyte activation, in part via upregulation of the CCR5 receptor in response to high glucose levels. Succinate, generated in excess via the glycolytic pathway, is a central ligand for inducing inflammation in diabetes, ageing, cancer and obesity.

In the eye the most prominent example of the newly appreciated involvement of innate immune mechanisms and inflammation is in AMD. The risk of developing AMD, the commonest cause of blindness in developed nations, is directly linked to mutations in a number of innate immune genes, particularly complement factor H (see p. 411) and the risk increases with accumulation of mutations.

Further information on AMD and the immune system is available online at https://expertconsult.inkling.com

Macrophages and DCs are directly implicated in the pathology of AMD—classic inflammatory macrophages in the early stages and alternative (angiogenic) macrophages when active choroidal neovascularization occurs in the wet form of the disease. In the atrophic form of AMD, inflammasome products IL-1 and IL-18 appear to be important in causing progressive atrophy of the RPE, while in the wet form of the disease IL-18 has been proposed to be anti-angiogenic. However, there are many different aspects to AMD pathogenesis that involve more than complement and induction of the inflammasome and AMD truly reflects a multifactorial disease whose principle risk factor remains age (Fig. 7.47).

Much is still to be learned about the role of the innate immune system and possibly the adaptive immune system in AMD pathogenesis, in addition to other diseases such as

CLASSIFICATION OF INTRAOCULAR INFLAMMATION

Classification of the many forms of uveitis/intraocular inflammation as diagnostic entities has lagged behind many other disease systems, mainly because of difficulties in reaching agreement about what constitutes each entity. For instance, diseases such as presumed ocular histoplasmosis syndrome are accepted entities in the USA but the same clinical disease in Europe is termed punctate inner choroidopathy. In addition, many of the conditions present with creamy-white lesions on ophthalmoscopic examination, but these 'white dot' syndromes comprise many different entities, both infectious and non-infectious.

Accordingly, a large international consortium of experts has gathered together many of the different clinical examples and has been cataloguing and characterizing the different uveitis syndromes, both in terms of general levels of activity/severity and risk of threat to sight as well as the phenotypic characteristics that define each disease. This effort is named the Standardization of Uveitis Nomenclature (SUN) project and the group is continuing to publish its findings. This work will be of great value to clinicians generally, as well as to the conduct of clinical trials of new drugs (Trusko et al., 2013).

AMD AND THE IMMUNE SYSTEM

The identification of associations of genetic mutations in complement proteins with AMD has opened a rich area of research into the possible role of inflammation in both wet and dry AMD. Wet AMD is characterized by the growth of abnormal vessels from the choroid through Bruch's membrane into the subretinal space where bleeding and vessel leakage cause an acute loss of vision if the lesion is at the fovea. In dry AMD, there is progressive atrophy of the retinal pigment epithelium, with loss of photoreceptors and progressive choroidal avascularity. In both conditions, macrophages have been implicated. In wet AMD, activated macrophages release VEGF, which promotes new vessel formation, while in dry AMD, macrophages are involved in progressive cell death of the RPE layer. Macrophages are switched on by many mediators, in particular complement components (see p. 412), and complement inhibitors are in place to prevent excessive complement activation. Mutations in complement proteins, such as complement factor H (CFH), CFH-related proteins 1 and 3, factor B/C2, C3 and factor I, are linked to increased risk of AMD and for the inhibitory proteins. Evidence has been produced to show that proteins such as complement factor H are less effective in controlling C activation if they contain one or more of the mutations. The pathogenetic role of complement factors in AMD is, however, complex. Partial activation of complement can produce factors that promote removal of tissue debris and thus prevent accumulation of material that is presumed to lead to overt AMD such as drusen. However, defective inhibitors clearly promote overactivation of complement. Although most of the complement components are produced in the liver, several can be generated locally within the retina and other parts of the eye and therefore locally generated inflammatory responses are possible given the correct triggers. Furthermore, factor B deficient mice are protected against infiltration of macrophages in a light-induced model of retinal degeneration. The possibility that adaptive immunity may be involved has also been explored: IL-1b produced during innate immune responses inhibits the T-regulatory activity of retinal pigment epithelial cells and thus potentially allows secondary T-cell activation. Additionally, AMD patients appear to have elevated levels of NKT cells in the circulation.

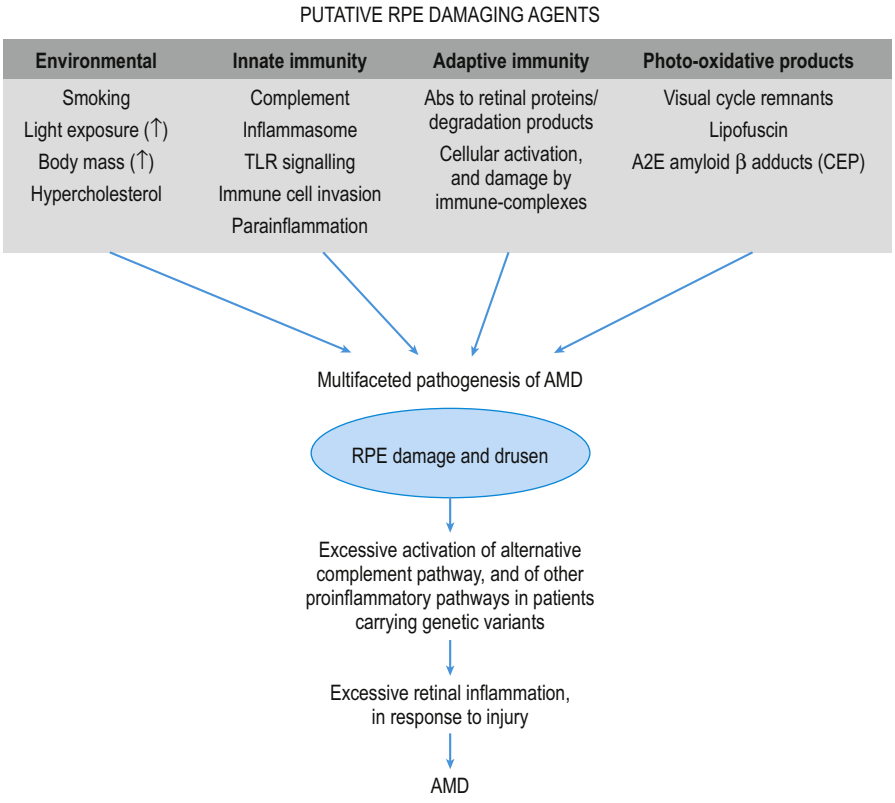


Fig. 7.47 Flow diagram outlining the main checkpoints in the development of age-related macular degeneration *(AMD)*. *RPE*, Retinal pigment epithelium. (From Whitcup et al., 2013.)

glaucoma and intraocular tumours, not to mention more obvious inflammatory diseases such as uveitis. However, a rich harvest of potential therapeutic targets using immune modulators has yet to be reaped.

Conclusion

The eye and its several tissues might be involved in any of the immune responses described in this chapter, either as a primary target of attack (e.g. in disciform herpetic keratitis or toxoplasmic choroiditis) or as part of a generalized immune disorder such as in Wegener's granulomatosis or sarcoidosis. The pathological processes and the mechanisms of initiation of immune responses are fundamentally similar from tissue to tissue. However, as stated in the introduction, each tissue has its unique microenvironment and this undoubtedly plays a part in the final expression of the immune response. Importantly, the eye (and the brain) has been revealed as a preferred site for latent infection by the increased incidence of immunosuppression in the community, not only as part of the AIDS epidemic and the increased use of immunosuppressive drugs but also by the natural history of diseases such as Ebola where survivors of the disease harbour latent virus, as well as the recognition that conditions such as recurrent anterior uveitis can be caused by latent cytomegalovirus infection.

A list of further reading and references is available online at https://expertconsult.inkling.com

References

Ahadome, S. D. et al., 2016. Aldehyde dehydrogenase inhibition blocks mucosal fibrosis in human and mouse ocular scarring. JCI Insight 1, e87001. https://doi:10.1172/jci.insight.87001.

Bach, J.F., 2018. The hygiene hypothesis in autoimmunity: the role of pathogens and commensals. Nat. Rev. Immunol. 18, 105–120. https://doi.org/10.1038/nri.2017.111. 2018.

Braza, M.S., et al., 2018. Inhibiting inflammation with myeloid cell-specific nanobiologics promotes organ transplant acceptance. Immunity 49, 819–828. https://doi.org/10.1016/j.immuni.2018.09.008. e816.

Castelli, E.C., Paz, M.A., Souza, A.S., Ramalho, J., Mendes-Junior, C.T., 2018. Hla-mapper: an application to optimize the mapping of HLA sequences produced by massively parallel sequencing procedures. Hum. Immunol. 79, 678–684. https://doi.org/10.1016/j.humimm.2018.06.010.

Ebbo, M., Crinier, A., Vely, F., Vivier, E., 2017. Innate lymphoid cells: major players in inflammatory diseases. Nat. Rev. Immunol. 17, 665–678. https://doi.org/10.1038/nri.2017.86.

Flajnik, M.F., 2018. A cold-blooded view of adaptive immunity. Nat. Rev. Immunol. 18, 438–453. https://doi.org/10.1038/s41577-018-0003-9. 2018.

Forrester, J.V., McMenamin, P.G., Dando, S.J., 2018a. CNS infection and immune privilege. Nat. Rev. Neurosci. 19, 655–671. https://doi.org/10.1038/s41583-018-0070-8.

Forrester, J.V., Kuffova, L., Dick, A.D., 2018b. Autoimmunity, autoinflammation, and infection in uveitis. Am. J. Ophthalmol. 189, 77–85. https://doi.org/10.1016/j.ajo.2018.02.019.

Gaetan, D., Vincent, R., Philippe, L., 2019. The n-3 docosapentaenoic acid (DPA): a new player in the n-3 long chain polyunsaturated fatty acid family. Biochimie. https://doi.org/10.1016/j.biochi.2019.01.022.

Guilliams, M., et al., 2014. Dendritic cells, monocytes and macrophages: a unified nomenclature based on ontogeny. Nat. Rev. Immunol. 14, 571–578. https://doi.org/10.1038/nri3712.

Hafstrand, I., et al., 2019. Successive crystal structure snapshots suggest the basis for MHC class I peptide loading and editing by tapasin. Proc. Natl. Acad. Sci. U. S. A. 116, 5055–5060. https://doi.org/10.1073/pnas.1807656116 .

Kaplan, M.H., Hufford, M.M., Olson, M.R., 2015. The development and in vivo function of T helper 9 cells. Nat. Rev. Immunol. 15, 295–307. https://doi.org/10.1038/nri3824.

Koeken, V.A., Verrall, A.J., Netea, M.G., Hill, P.C., van Crevel, R., 2019. Trained innate immunity and resistance to Mycobacterium tuberculosis infection. Clin. Microbiol. Infect. https://doi.org/10.1016/j.cmi.2019.02.015.

Koelwyn, G.J., Moore, K.J., 2019. Defining macrophages in the heart one cell at a time. Trends Immunol. https://doi.org/10.1016/j.it.2019.01.010.

Kuffova, L. et al., 2016. High-Risk Corneal Graft Rejection in the Setting of Previous Corneal Herpes Simplex Virus (HSV)-1 Infection. Invest Ophthalmol Vis. Sci. 57, 1578–1587. https://doi:10.1167/iovs.15-17894.

Li, Z., et al., 2018. Adult connective tissue-resident mast cells originate from late erythro-myeloid progenitors. Immunity, 49, 640–653. e645. https://doi.org/10.1016/j.immuni.2018.09.023.

Lin, C., et al., 2019. Fever promotes T lymphocyte trafficking via a thermal sensory pathway involving heat shock protein 90 and alpha4 integrins. Immunity 50, 137–151, e136. https://doi.org/10.1016/j.immuni.2018.11.013.

Mantovani, A. et al. (2004). The chemokine system in diverse forms of macrophage activation and polarization. Trends Immunol. 25, 677–686. https://doi:10.1016/j.it.2004.09.015.

Marakalala, M.J., Martinez, F.O., Pluddemann, A., Gordon, S., 2018. Macrophage heterogeneity in the Immunopathogenesis of tuberculosis. Front. Microbiol. 9, 1028. https://doi.org/10.3389/fmicb.2018.01028.

Medawar, P.B., 1956. The immunology of transplantation. Harvey Lect, 144–176.

Menny, A., et al., 2018. CryoEM reveals how the complement membrane attack complex ruptures lipid bilayers. Nat. Commun. 9, 5316. https://doi.org/10.1038/s41467-018-07653-5 .

Riley, T.P., Baker, B.M., 2018. The intersection of affinity and specificity in the development and optimization of T cell receptor based therapeutics. Semin. Cell Dev. Biol. 84, 30–41. https://doi.org/10.1016/j.semcdb.2017.10.017.

Sakaguchi, S., Fukuma, K., Kuribayashi, K., Masuda, T., 1985. Organ-specific autoimmune diseases induced in mice by elimination of T cell subset. I. Evidence for the active participation of T cells in natural self-tolerance; deficit of a T cell subset as a possible cause of autoimmune disease. J. Exp. Med. 161, 72–87. https://doi.org/10.1084/jem.161.1.72.

Song, P., et al., 2017. Downregulation of interferon-beta and inhibition of TLR3 expression are associated with fatal outcome of severe fever with thrombocytopenia syndrome. Sci. Rep. 7, 6532. https://doi.org/10.1038/s41598-017-06921-6 .

Starling, S., 2018. Innate lymphoid cells: lipid surveillance by skin ILCs. Nat. Rev. Immunol. 18, 78–79. https://doi.org/10.1038/nri.2018.1.

Sun, L., Chen, C., Wu, J., Dai, C., Wu, X., 2018. TSLP-activated dendritic cells induce T helper type 2 inflammation in Aspergillus fumigatus keratitis. Exp. Eye Res. 171, 120–130. https://doi.org/10.1016/j.exer.2018.03.014.

Vizcaino, M.A., Joseph, S.S., Eberhart, C.G., 2018. Orbital progressive transformation of germinal centers as part of the spectrum of IgG4-related ophthalmic disease: clinicopathologic features of three cases. Saudi. Saudi. J. Ophthalmol. 32, 56–61. https://doi.org/10.1016/j.sjopt.2018.02.017.

Yousefi, S., et al., 2019. Untangling "NETosis" from NETs. Eur. J. Immunol. 49, 221–227. https://doi.org/10.1002/eji.201747053 .

Zhong, X., Liu, H., Pu, A., Xia, X., Zhou, X., 2007. M cells are involved in pathogenesis of human contact lens-associated giant papillary conjunctivitis. Arch. Immunol. Ther. Exp. 55, 173–177. https://doi.org/10.1007/s00005-007-0022-x.

Further reading

Anderson, G., Beischlag, T.V., Vinciguerra, M., Mazzoccoli, G., 2013. The circadian clock circuitry and the AHR signaling pathway in physiology and pathology. Biochem. Pharmacol. 85, 1405–1416.

Bach, J.F., 2018. The hygiene hypothesis in autoimmunity: the role of pathogens and commensals. Nat. Rev. Immunol. 18, 105–120. https://doi.org/10.1038/nri.2017.111. 2018.

Bajénoff, M., 2012. Stromal cells control soluble material and cellular transport in lymph nodes. Front. Immunol. 3, 304.

Chen, Y.S., Shen, C.R., 2015. Immune checkpoint blockade therapy: the 2014 tang prize in Biopharmaceutical science. Biomed. J. 38, 5–8. https://doi.org/10.4103/2319-4170.151150.

Cho, I., Blaser, M.J., 2012. The human microbiome: at the interface of health and disease. Nat. Rev. Genet. 13, 260–270.

Chong, A.S., Alegre, M.L., 2012. The impact of infection and tissue damage in solid-organ transplantation. Nat. Rev. Immunol. 12, 459–471.

Costello, E.K., Stagaman, K., Dethlefsen, L., Bohannan, B.J., Relman, D.A., 2012. The application of ecological theory toward an understanding of the human microbiome. Science 336, 1255–1262. https://doi.org/10.1126/science.1224203.

Craft, J.E., 2012. Follicular helper T cells in immunity and systemic autoimmunity. Nat. Rev. Rheumatol. 8, 337–347.

Croft, M., Duan, W., Choi, H., Eun, S.Y., Madireddi, S., Mehta, A., 2012. TNF superfamily in inflammatory disease: translating basic insights. Trends Immunol. 33, 144–152.

Croker, B.A., Lawson, B.R., Rutschmann, S., Berger, M., Eidenschenk, C., Blasius, A.L., et al., 2008. Inflammation and autoimmunity caused by a SHP1 mutation depend on IL-1, MyD88, and a microbial trigger. Proc. Natl. Acad. Sci. U.S.A. 105, 15028–15033.

Dang, Z., Kuffova, L., Liu, L., Forrester, J.V., 2013. Soluble antigen traffics rapidly and selectively from the corneal surface to the eye draining lymph node and activates T cells when codelivered with CpG oligonucleotides. J. Leukoc. Biol. 95, 431–440.

Dustin, M.L., Depoil, D., 2011. New insights into the T cell synapse from single molecule techniques. Nat. Rev. Immunol. 11, 672–684.

Ebbo, M., Crinier, A., Vely, F., Vivier, E., 2017. Innate lymphoid cells: major players in inflammatory diseases. Nat. Rev. Immunol. 17, 665–678. https://doi.org/10.1038/nri.2017.86.

Engelhardt, B., Ransohoff, R.M., 2012. Capture, crawl, cross: the T cell code to breach the blood-brain barriers. Trends Immunol. 33, 579–589.

Fallarino, F., Grohmann, U., Puccetti, P., 2012. Indoleamine 2,3-dioxygenase: from catalyst to signaling function. Eur. J. Immunol. 42, 1932–1937.

Flajnik, M.F., 2018. A cold-blooded view of adaptive immunity. Nat. Rev. Immunol. 18, 438–453. https://doi.org/10.1038/s41577-018-0003-9.

Forrester, J.V., Xu, H., 2012. Good news-bad news: the yin and yang of immune privilege in the eye. Front. Immunol. 3, 338.

Forrester, J.V., McMenamin, P.G., Dando, S.J., 2018a. CNS infection and immune privilege. Nat. Rev. Neurosci. 19, 655–671. https://doi.org/10.1038/s41583-018-0070-8.

Forrester, J.V., Kuffova, L., Dick, A.D., 2018b. Autoimmunity, autoinflammation, and infection in uveitis. Am. J. Ophthalmol. 189, 77–85. https://doi.org/10.1016/j.ajo.2018.02.019.

Gaetan, D., Vincent, R., Philippe, L., 2019. The n-3 docosapentaenoic acid (DPA): a new player in the n-3 long chain polyunsaturated fatty acid family. Biochimie. https://doi.org/10.1016/j.biochi.2019.01.022.

Geering, B., Stoeckle, C., Conus, S., Simon, H.U., 2013. Living and dying for inflammation: neutrophils, eosinophils, basophils. Trends Immunol. 34, 398–409.

Ginhoux, F., Merad, M., 2011. Microglia arise from extra-embryonic yolk sac primitive progenitors. Med. Sci. 27, 719–724.

Granzin, J., Wilden, U., Choe, H.-W., Labahn, J., Krafft, B., Buldt, G., 1998. X-ray crystal structure of arrestin from bovine rod outer segments. Nature 391, 918–921. https://doi:10.1038/36147.

Greenwell-Wild, T., Moutsopoulos, N.M., Gliozzi, M., Kapsogeorgou, E., Rangel, Z., Munson, P.J., et al., 2011. Chitinases in the salivary glands and circulation of patients with Sjogren's syndrome: macrophage harbingers of disease severity. Arthritis Rheum. 63, 3103–3115.

Greter, M., Lelios, I., Pelczar, P., Hoeffel, G., Price, J., Leboeuf, M., et al., 2012. Stroma-derived interleukin-34 controls the development and maintenance of Langerhans cells and the maintenance of microglia. Immunity 37, 1050–1060.

Guilliams, M., et al., 2014. Dendritic cells, monocytes and macrophages: a unified nomenclature based on ontogeny. Nat. Rev. Immunol. 14, 571–578. https://doi.org/10.1038/nri3712.

Hafstrand, I., Sayitoglu, E. C., Apavaloaei, A., Josey, B. J., Sun, R., Han, X., et al., 2019. Successive crystal structure snapshots suggest the basis for MHC class I peptide loading and editing by tapasin. Proc. Natl. Acad. Sci. U. S. A. 116, 5055–5060. https://doi.org/10.1073/pnas.1807656116.

Hara, Y., Shoji, J., Hori, S., Ishimori, A., Kato, H., Inada, N., et al., 2012. Evaluation of eosinophilic inflammation in a novel murine atopic keratoconjunctivitis model induced by crude Dermatophagoides farinae antigen. Allergol. Int. 61, 331–338.

Horton, R., Wilming, L., Rand, V., Lovering, R.C., Bruford, E.A., Khodiyar, V.K., et al., 2004. Gene map of the extended human MHC. Nat. Rev. Genet. 5, 889–899.

Hunter, C.A., Kastelein, R., 2012. Interleukin-27: balancing protective and pathological immunity. Immunity 37, 960–969.

Joffre, O.P., Segura, E., Savina, A., Amigorena, S., 2012. Cross-presentation by dendritic cells. Nat. Rev. Immunol. 12, 557–569.

Kieser, K.J., Kagan, J.C., 2017. Multi-receptor detection of individual bacterial products by the innate immune system. Nat. Rev. Immunol. 17, 376–390. https://doi.org/10.1038/nri.2017.25.

Kaplan, M.H., Hufford, M.M., Olson, M.R., 2015. The development and in vivo function of T helper 9 cells. Nat. Rev. Immunol. 15, 295–307. https://doi.org/10.1038/nri3824.

Kim, M., Johnston, M.G., Gupta, N., Moore, S., Yucel, Y.H., 2011. A model to measure lymphatic drainage from the eye. Exp. Eye Res. 93, 586–591.

Kiss, E.A., Diefenbach, A., 2012. Role of the aryl hydrocarbon receptor in controlling maintenance and functional programs of RORgammat(+) innate lymphoid cells and intraepithelial lymphocytes. Front. Immunol. 3, 124.
Klein, T., Bischoff, R., 2011. Physiology and pathophysiology of matrix metalloproteases. Amino Acids. 41, 271–290.
Koeken, V.A., Verrall, A.J., Netea, M.G., Hill, P.C., van Crevel, R., 2019. Trained innate immunity and resistance to Mycobacterium tuberculosis infection. Clin. Microbiol. Infect. https://doi.org/10.1016/j.cmi.2019.02.015.
Koelwyn, G.J., Moore, K.J., 2019. Defining macrophages in the Heart one cell at a time. Trends Immunol. https://doi.org/10.1016/j.it.2019.01.010.
Kolaczkowska, E., Kubes, P., 2013. Neutrophil recruitment and function in health and inflammation. Nat. Rev. Immunol. 13, 159–175.
Kumar, P., Rajasekaran, K., Palmer, J.M., Thakar, M.S., Malarkannan, S., 2013. IL-22: an evolutionary missing-link authenticating the role of the immune system in tissue regeneration. J. Cancer 4, 57–65.
Lalor, P.F., Faint, J., Aarbodem, Y., Hubscher, S.G., Adams, D.H., 2007. The role of cytokines and chemokines in the development of steatohepatitis. Semin. Liver Dis. 27, 173–193.
Lazennec, G., Richmond, A., 2010. Chemokines and chemokine receptors: new insights into cancer-related inflammation. Trends Mol. Med. 16, 133–144.
Li, Z., Liu, S., Xu, J., Zhang, X., Han, D., Liu, J., et al., 2018. Adult connective tissue-resident mast cells originate from late erythro-myeloid progenitors. Immunity. 49,640–653 e645. https://doi.org/10.1016/j.immuni.2018.09.023.
Lin, C., et al., 2019. Fever promotes T lymphocyte trafficking via a thermal sensory pathway involving heat shock protein 90 and alpha4 integrins. Immunity 50, 137–151. e136. https://doi.org/10.1016/j.immuni.2018.11.013.
Liston, A., Masters, S.L., 2017. Homeostasis-altering molecular processes as mechanisms of inflammasome activation. Nat. Rev. Immunol. 17, 208–214. https://doi.org/10.1038/nri.2016.151.
Manz, B.N., Groves, J.T., 2010. Spatial organization and signal transduction at intercellular junctions. Nat. Rev. Mol. Cell Biol. 11, 342–352.
Mantovani, A., et al., 2004. The chemokine system in diverse forms of macrophage activation and polarization. Trends Immunol. 25, 677–686. https://doi.org/10.1016/j.it.2004.09.015.
Marakalala, M.J., Martinez, F.O., Pluddemann, A., Gordon, S., 2018. Macrophage heterogeneity in the Immunopathogenesis of tuberculosis. Front. Microbiol. 9, 1028. https://doi.org/10.3389/fmicb.2018.01028.
Matzinger, P., Kamala, T., 2011. Tissue-based class control: the other side of tolerance. Nat. Rev. Immunol. 11, 221–230.
Menny, A., et al., 2018. CryoEM reveals how the complement membrane attack complex ruptures lipid bilayers. Nat. Commun. 9, 5316. https://doi.org/10.1038/s41467-018-07653-5.
Merad, M., Ginhoux, F., Collin, M., 2008. Origin, homeostasis and function of Langerhans cells and other langerin-expressing dendritic cells. Nat. Rev. Immunol. 8, 935–947.
Metz, M., Maurer, M., 2007. Mast cells – key effector cells in immune responses. Trends Immunol. 28, 234–241.
Mills, K.H., 2011. TLR-dependent T cell activation in autoimmunity. Nat. Rev. Immunol. 11, 807–822.
Minton, K., 2011. Granulocytes: a weighty role for eosinophils. Nat. Rev. Immunol. 11, 299.
Mjosberg, J., Bernink, J., Peters, C., Spits, H., 2012. Transcriptional control of innate lymphoid cells. Eur. J. Immunol. 42, 1916–1923.
Mosser, D.M., Edwards, J.P., 2008. Exploring the full spectrum of macrophage activation. Nat. Rev. Immunol. 8, 958–969.
Murray, P.J., Wynn, T.A., 2011. Protective and pathogenic functions of macrophage subsets. Nat. Rev. Immunol. 11, 723–737.
Neyt, K., Perros, F., GeurtsvanKessel, C.H., Hammad, H., Lambrecht, B.N., 2012. Tertiary lymphoid organs in infection and autoimmunity. Trends Immunol. 33, 297–305.
Ng, D., Gommerman, J.L., 2013. The regulation of immune responses by DC derived type I IFN. Front. Immunol. 4, 94.
Quintana, F.J., 2013. The aryl hydrocarbon receptor: a molecular pathway for the environmental control of the immune response. Immunology 138, 183–189.
Rask-Madsen, C., Kahn, C.R., 2012. Tissue-specific insulin signaling, metabolic syndrome, and cardiovascular disease. Arterioscler. Thromb. Vasc. Biol. 32, 2052–2059.
Riley, T.P., Baker, B.M., 2018. The intersection of affinity and specificity in the development and optimization of T cell receptor based therapeutics. Semin. Cell Dev. Biol. 84, 30–41. https://doi.org/10.1016/j.semcdb.2017.10.017.
Satpathy, A.T., Wu, X., Albring, J.C., Murphy, K.I.M., 2012. Re(de) fining the dendritic cell lineage. Nat. Immunol. 13, 1145–1154.
Shinomiya, K., et al., 2013. Immunohistochemical analysis of inflammatory limbal conjunctiva adjacent to Mooren's ulcer. Br. J. Ophthalmol. 97, 362–366. https://doi.org/10.1136/bjophthalmol-2012-302631. 2013.
Song, P., Zheng, N., Zhang, L., Liu, Y., Chen, T., Bao, C., et al., 2017. Downregulation of interferon-beta and inhibition of TLR3 expression are associated with fatal outcome of severe fever with Thrombocytopenia syndrome. Sci. Rep. 7, 6532. https://doi.org/10.1038/s41598-017-06921-6.
Soriano-Romani, L., Contreras-Ruiz, L., Lopez-Garcia, A., Diebold, Y., Masli, S., 2018. Topical application of TGF-beta-activating peptide, KRFK, prevents inflammatory manifestations in the TSP-1-deficient mouse model of chronic ocular inflammation. Int. J. Mol. Sci. 20. https://doi.org/10.3390/ijms20010009.
St John, A.L., Abraham, S.N., 2013. Innate immunity and its regulation by mast cells. J. Immunol. 190, 4458–4463.
Starling, S., 2018. Innate lymphoid cells: lipid surveillance by skin ILCs. Nat. Rev. Immunol. 18, 78–79. https://doi.org/10.1038/nri.2018.1.
Sun, L., Chen, C., Wu, J., Dai, C., Wu, X., 2018. TSLP-activated dendritic cells induce T helper type 2 inflammation in Aspergillus fumigatus keratitis. Exp. Eye Res. 171, 120–130. https://doi.org/10.1016/j.exer.2018.03.014.
Swain, S.L., McKinstry, K.K., Strutt, T.M., 2012. Expanding roles for CD4(+) T cells in immunity to viruses. Nat. Rev. Immunol. 12, 136–148.
Timar, C.I., Lorincz, A.M., Ligeti, E., 2013. Changing world of neutrophils. Pflugers. Arch. 465, 1521–1533.
Trusko, B., Thorne, J., Jabs, D., Belfort, R., Dick, A., Gangaputra, S., et al., 2013. The Standardization of Uveitis Nomenclature (SUN) project. Development of a clinical evidence base utilizing informatics tools and techniques. Methods Info. Med. 52, 259–265.
Usui, Y., Parikh, J., Goto, H., Rao, N.A., 2008. Immunopathology of necrotising scleritis. Br. J. Ophthalmol. 92, 417–419.

van de Veerdonk, F.L., Netea, M.G., 2013. New insights in the immunobiology of IL-1 family members. Front. Immunol. 4, 167.
Vantourout, P., Hayday, A., 2013. Six-of-the-best: unique contributions of gammadelta T cells to immunology. Nat. Rev. Immunol. 13, 88–100.
Veldhoen, M., Brucklacher-Waldert, V., 2012. Dietary influences on intestinal immunity. Nat. Rev. Immunol. 12, 696–708.
Vitova, A., Kuffová, L., Klaska, I.P., Holan, V., Cornall, R.J., Forrester, J.V., 2013. The high-risk corneal regraft model: a justification for tissue matching in humans. Transpl. Int. 26, 453–461.
Vizcaino, M.A., Joseph, S.S., Eberhart, C.G., 2018. Orbital progressive transformation of germinal centers as part of the spectrum of IgG4-related ophthalmic disease: clinicopathologic features of three cases. Saudi J. Ophthalmol. 32, 56–61. https://doi.org/10.1016/j.sjopt.2018.02.017.
Voehringer, D., 2013. Protective and pathological roles of mast cells and basophils. Nat. Rev. Immunol. 13, 362–375.
Whitcup, S.M., Sodhi, A., Atkinson, J.P., Holers, V.M., Sinha, D., Rohrer, B., et al., 2013. The role of the immune response in age-related macular degeneration. Int. J. Inflam. 348092.
Williams, K.A., Coster, D.J., 2007. The immunobiology of corneal transplantation. Transplantation 84, 806–813.
Yousefi, S., Stojkov, D., Germic, N., Simon, D., Wang, X., Benarafa, C., et al., 2019. Untangling "NETosis" from NETs. Eur. J. Immunol. 49, 221–227. https://doi.org/10.1002/eji.201747053.
Yucel, Y.H., Johnston, M.G., Ly, T., Patel, M., Drake, B., Gümüş, E., et al., 2009. Identification of lymphatics in the ciliary body of the human eye: a novel 'uveolymphatic' outflow pathway. Exp. Eye Res. 89, 810–819.
Zhong, X., Liu, H., Pu, A., Xia, X., Zhou, X., 2007. M cells are involved in pathogenesis of human contact lens-associated giant papillary conjunctivitis. Arch. Immunol. Ther. Exp. 55, 173–177. https://doi.org/10.1007/s00005-007-0022-x.
Ziegler, A., Kentenich, H., Uchanska-Ziegler, B., 2005. Female choice and the MHC. Trends Immunol. 26, 496–502.
Zipfel, P.F., Skerka, C., 2009. Complement regulators and inhibitory proteins. Nat. Rev. Immunol. 9, 729–740.

Chapter 8

Microbial infections of the eye

CONTENTS

Introduction
Source of microbes that cause ocular infections
- Environmental
- Latent viruses
- Insect-borne ocular infections

Host defences at the ocular surface
- Physical barriers
- Epithelium
- Resident macrophages and dendritic cells in the cornea
- Commensal bacteria
- Pathogen recognition receptors and recruitment of neutrophils
- Toll-like receptors (TLRs)
- NOD-like receptors
- Contact lenses are a major risk factor for microbial keratitis

Viral infections of the eye
- DNA viruses
- RNA viruses
- Vector-borne viruses
- Viral retinitis

Bacterial infections of the eye
- The ocular surface and bacterial infections

Fungal infections of the eye
- Protozoan causes of ocular disease
- Helminths can cause ocular infection
- Onchocerciasis (river blindness)
- Loa loa (loiasis, eye worm)

Conclusions

Introduction

The outcome of microbial infections depends on the balance between virulence of the infecting organism and the strength of the host response which will determine the severity of disease and the longevity of infection. In most infections, tissue damage is a direct result of cytotoxic activity of microbial products such as exotoxins: however, it is the host inflammatory response that causes collateral tissue damage associated with anti-microbial activity. Microbial virulence may then be defined as the ability of the organisms to survive in the presence of the host immune response. This chapter will focus on infections of the anterior eye based on infections at the ocular surface, but it will also discuss infections of the posterior eye including endophthalmitis and ocular toxoplasmosis.

Source of microbes that cause ocular infections

Environmental

Many of the organisms that cause severe ocular infections are listed in Table 8.1 and are either ubiquitous in the environment, or are part of our normal body flora and do not cause infections or disease in immunocompetent, healthy individuals. These microbes are opportunistic pathogens that require a breach of the physical barriers of the ocular surface to cause corneal infections. For example, *Pseudomonas aeruginosa* and *Acanthamoeba* are normal freshwater organisms in ponds, lakes and in household water supplies, including showerheads. Similarly, *Aspergillus* and *Fusarium* moulds, which are plant saprophytes and pathogens, are ubiquitous in the air we breathe. Spore counts are higher in hot and humid areas of the world, especially in agricultural regions, and during harvest season. *Aspergillus fumigatus* is among the most ubiquitous organisms worldwide, and except in extremely cold environments, most individuals inhale *Aspergillus* spores on a daily basis. *Aspergillus fumigatus* is the cause of the green mould found on food in our kitchens and airborne *Aspergillus* and *Fusarium* spores and water-borne *Pseudomonas* and *Serratia* are normal inhabitants of our bathrooms. The number of organisms in

Table 8.1 **Microbial pathogens of the eye**[a]

Viruses	Genus, species	Major site of infection
DNA Viruses	Adenovirus	Conjunctiva
	Herpes simplex virus	Cornea
	Herpes zoster virus	Cornea
	Cytomegalovirus	Retina
RNA viruses	Zika virus	Conjunctiva, retina
	Dengue virus	Retina
Bacteria		
Obligate intracellular	*Chlamydia trachomatis*	Conjunctiva (trachoma)
Gram-negative bacteria	*Pseudomonas aeruginosa*	Cornea
	Serratia marcescens	Cornea
Gram-positive bacteria	*Streptococcus pneumoniae*	Conjunctiva, cornea, vitreous/retina
	Staphylococcus aureus	Cornea
	Bacillus cereus	Vitreous/retina
	Chlamydia trachomatis	Conjunctiva
Fungi		
Yeast	*Candida albicans*	Cornea
Mould	*Fusarium oxysporum, F. solani*	Cornea
	Aspergillus fumigatus, A. flavus	Cornea
Protozoa	*Toxoplasma gondii*	*Retina*
	Acanthamoeba castellani	Cornea
	Microsoporidia	Cornea
Helminths		
Nematodes	*Toxocara canis*	Retina
	Onchocerca volvulus (river blindness)	Cornea, retina
	Loa loa (eye-worm)	Conjunctiva

[a]Although a search of case reports will produce a much longer list of organisms, the table includes only the most common causes of ocular infection.
CMV, Cytomegalovirus; *PA*, *Pseudomonas aeruginosa; Sm*, Serratia marcescens; *SP*, *Streptococcus pneumoniae*; *Sa, Staphylococcus aureus*.

the environment is generally higher in warmer climates and during the summer in temperate climates.

In addition to the external environment, multiple organisms are present in the human conjunctiva as part of our normal body flora (Adenovirus, *Campylobacter*), the skin (*Staphylococci*) and the nasopharynx (*Streptococcus pneumoniae, Candida albicans*).

Toxoplasma gondii is ubiquitous in the environment, although the sexual stage of the organism is specific to the family of *Felidae*, which includes domestic cats. Humans are infected orally following inadvertent ingestion of food that is contaminated with cat faecal material, including litter boxes.

Latent viruses

Neurotropic DNA viruses persist as latent viruses causing ocular infections. Almost all individuals are exposed to herpes simplex virus (HSV-1) and herpes zoster virus (HZV) early in life, which remain in a latent form in sensory ganglia, including the trigeminal ganglia; however, once reactivated, HSV and HZV undergo retrograde migration along the nerve fibres of the trigeminal ganglia where they cause lesions in the skin of the face and the cornea supplied. HSV and HZV are the most common cause of corneal infections in the USA and other industrialized countries. Non-neurotropic viruses such as CMV are also widely prevalent and remain latent in a range of cell types such as endothelia and bone marrow derived cells. Persistence of non-DNA viruses is also a cause of recurrent infection such as Ebola, which can cause late ocular inflammation in individuals who are not killed by this virus.

Insect-borne ocular infections

Trachoma is a major example of an organism that thrives in unsanitary environments, where it is carried by flies that breed on human waste and feed in large numbers on fluids

associated with mucous membranes of the mouth, nose and eyes. Bacteria and infected cells in the tears and mucus are carried by these flies from infected to uninfected eyes. They are not true vectors of *Chlamydia* as there is no life cycle requirement; in contrast, Culex mosquitoes are vectors of Flaviviruses that include Zika virus, Dengue virus and West Nile virus that can all cause ocular disease. *Onchocerca volvulus*, the cause of river blindness and *Loa loa* (eyeworm) are highly adapted human parasites that require an intermediate blackfly or mosquito vector and are found primarily in Africa.

Host defences at the ocular surface

Physical barriers

Blinking is a very effective cleansing mechanism and eyelashes can trap microbes and larger debris, preventing access to the globe. In addition, the eyelids contain sebaceous glands that secrete lactic acid and fatty acids in a low pH environment, which have a direct inhibitory effect on bacterial replication. Tears also contain antimicrobial compounds, including lacritin, lactoferrin, lipocalin and lysozyme, β-defensins and other anti-microbial peptides. The tear film also includes secretory immunoglobulin A (IgA), IgG and complement that contribute to protection against bacterial invasion (see Chapter 7). Protection is also afforded by neutrophils in the tears, which greatly increase in numbers when the eyelids are closed at night.

Epithelium

The corneal epithelium has a glycocalyx on the apical surface comprised of large proteoglycans mucins that include MUC1, MUC4 and MUC16 (see Chapter 4, p. 208). This mucin layer provides a physical barrier that restricts bacterial adherence to corneal epithelial cells, in addition to releasing mucins into the tear film (Fig. 8.1). Epithelial cells form tight junctions, which generate a very effective barrier against subsequent penetration of the corneal stroma; therefore, infections generally occur as a consequence of traumatic injury or alteration of the corneal surface microenvironment. Epithelial cells also secrete antimicrobial peptides including β-defensins, cathelicidin (LL-37) and calprotectin (S100A8/A9) and by degradation of cytokeratin-6A in corneal epithelial cells. These keratin-derived antimicrobial peptides (KDAMPs) have distinct non-αβ amphipathic conformations with a central 10-residue glycine-rich hydrophobic strip that is central to its broad bactericidal activity.

Resident macrophages and dendritic cells in the cornea

The normal mammalian cornea was thought to be devoid of myeloid (bone-marrow derived) cells as the predominant cells in the corneal stroma are keratocytes that produce cornea-specific collagens and proteoglycans; however, with improved immunostaining methods and by examining whole mount cornea preparations rather than histological sections, an entire network of macrophages and dendritic cells was revealed. Most bone marrow-derived cells in the normal corneal stroma have characteristics of macrophages, although dendritic cells are more prominent in the peripheral, limbal region of the cornea and in the basement membrane of the epithelium (Bowman's membrane). These cells extend pseudopodia ('periscopes') to the apical surface to detect microbes or microbial products (Fig. 8.2A–E). Nanotubes associated with macrophages appear to connect distant cells in the corneal stroma, especially during inflammation (Fig. 8.2F).

Commensal bacteria

Commensal bacteria have not been reliably detected on the corneal surface; however, a compelling study by St Leger et al. (2017) demonstrated that the conjunctiva of healthy mice is colonized by *Corynebacterium mastiditis*, which elicit IL-17a production by resident γδ T cells. IL-17a induces release of anti-microbial peptides by conjunctival epithelial cells that prevent infection with *P. aeruginosa* and *Candida albicans*. *C. mastiditis* is also prevalent in the conjunctiva of healthy individuals.

Pathogen recognition receptors and recruitment of neutrophils

Most nucleated cells can recognize and respond to microbial products, although macrophages and dendritic cells are most efficient. These cells express multiple copies of cell surface and endosomal receptors that recognize bacterial and fungal proteins, carbohydrates, lipids and nucleic acids (termed pathogen recognition receptors, PRRs) (see Chapter 7, p. 389). Ligand binding initiates intracellular signalling events that result in production of pro-inflammatory and chemotactic cytokines. In contrast to T and B cell antigen recognition receptors, which are generated following gene rearrangement, PRRs are invariant receptors encoded in the germ line genes that do not undergo gene rearrangement and are able to respond to microbial products immediately after receptor engagement. Furthermore, whereas mature T and B cells primarily recognize specific peptides (and in some cases well defined carbohydrates), the invariant receptors associated with innate immunity recognize conserved proteins, lipids and nucleic acids primarily associated with microbes. Though not a focus of the current chapter, these receptors can also recognize endogenous self-antigens, Danger Associated Molecular Patterns (DAMPs see Chapter 7) including silica, uric acid, asbestos crystals and advanced glycation end products, which activate the innate immune response. Self-antigens also induce

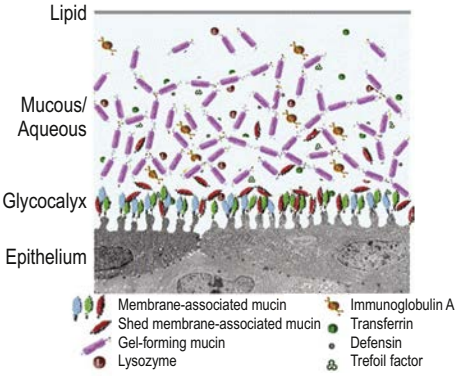


Fig. 8.1 Ocular surface mucins. The aqueous phase of the tears contains MUC5AC *(purple)* and other tear components. The glycocalyx has extended membrane-associated MUC1 *(blue)*, MUC4 *(green)* and MUC16 *(red)*. Extracellular domains are released into the tear film. (From Gipson, 2004.)

autoimmune responses that can cause ocular inflammation, including uveitis (see Chapter 7).

Among the best-characterized innate immune receptors are cell surface and endosomal toll-like receptors (TLRs), cell surface c-type lectins and intracellular nucleotide-binding oligomerization domain (NOD)-like receptors (NLRs).

Toll-like receptors (TLRs)

TLR family members are single transmembrane receptors that recognize structurally conserved microbial products; furthermore, activation of these receptors leads to production of pro-inflammatory and chemotactic cytokines that mediate recruitment of neutrophils, macrophages and lymphocytes to the site of infection. As shown in Fig. 8.3, TLRs are located on the cell surface and in endosomes, and can recognize lipids, proteins and nucleic acids, primarily via leucine-rich repeat regions. Lipid-binding TLRs include TLR2, which forms heterodimers with TLR1 or TLR6 to bind lipopeptides, and TLR4/MD-2 that recognize the lipid A moiety of lipopolysaccharide (LPS). TLR5 and TLR11 recognize proteins and are activated by bacterial flagellin (TLR5), uropathogenic *Escherichia coli* or *Toxoplasma gondii* profilin (TLR11). TLR3, TLR7, TLR8 and TLR9 are located on endosomal membranes and bind viral and bacterial nucleic acids. With the exception of TLR3, all TLRs stimulate the cells through the MyD88 common adaptor molecule, leading to NF-κB translocation to the nucleus and expression of genes encoding pro-inflammatory cytokines and chemotactic cytokines (chemokines). TLR3 and TLR4 activate the TRIF pathway, which induces IRF3 transcription and production of type I interferons that mediate antiviral responses. TLR4 activation involves accessory molecules, including lipopolysaccharide (LPS) binding protein and CD14, which combine to extract single endotoxin molecules from the outer membrane and form monomeric endotoxin, and MD-2, which is the receptor for the lipid A moiety of LPS. CD14 also functions as a chaperone that transports TLR4 from the plasma membrane to endosomes in order to activate the TRIF pathway.

NOD-like receptors

NLRs comprise an intracellular family of pathogen recognition molecules, which also activate NFkB and production of pro-inflammatory and chemotactic cytokines. NOD2 has been well characterized and shown to recognize muramyl dipeptide on degraded bacterial cell wall peptidoglycan, and can therefore respond to invading Gram-negative and Gram-positive bacteria (Box 8.1). However, as Gram-positive bacteria have more peptidoglycan in the cell wall, NOD2 activation occurs following infection of *Staphylococci* or *Streptococci*.

NLRs include NLRP3, which recognizes bacterial toxins and crystals, and NLRC4, which recognizes flagellin of Gram-negative bacteria such as *P. aeruginosa* (Fig. 8.4). Once activated, these NLRs form a large, multi-protein complex called an inflammasome, which facilitates caspase-1 cleavage of IL-1β, IL-18 and IL-33 from the inactive pro-form, to the bioactive, mature form of these cytokines. Caspase-1 also cleaves the pro-form of the pore forming protein Gasdermin-D and the N-terminal peptides oligomerize in the plasma membrane to form pores, resulting in release of these cytokines and necrotic cell death, also termed pyroptosis.

C-type lectins: In addition to bacterial products, host cells recognize fungal cell wall components by activation of c-type lectins on the cell surface (Fig. 8.5). Dectin-1 recognizes β-glucan, whereas Dectin-2 and Dectin-3 recognize alpha mannans and are activated after either clustering of Dectin-1 or heterodimerization of Dectin-2 and Dectin-3, which leads to production of pro-inflammatory and chemotactic cytokines.

Neutrophils are the first responders to invading microbes: neutrophils are the most abundant leucocytes in the blood of normal individuals, comprising ~70% total white cells. Neutrophils in the blood constitutively express receptors for chemokines, specifically CXCR1 that binds IL-8/CXCL8, and CXCR2, which binds CXCL1, CXCL2 and CXCL5 (although murine neutrophils express CXCR1, only human neutrophils produce IL-8 and murine neutrophils respond to CXCL1, 2 and 5 through CXCR2). Neutrophils also constitutively express adhesion molecules that bind to receptors on vascular endothelial cells (Box 8.2).

Neutrophils (and other leucocytes) are recruited from capillaries (including limbal blood vessels) to infected tissues through a sequence of events in which pro-inflammatory cytokines produced at the site of

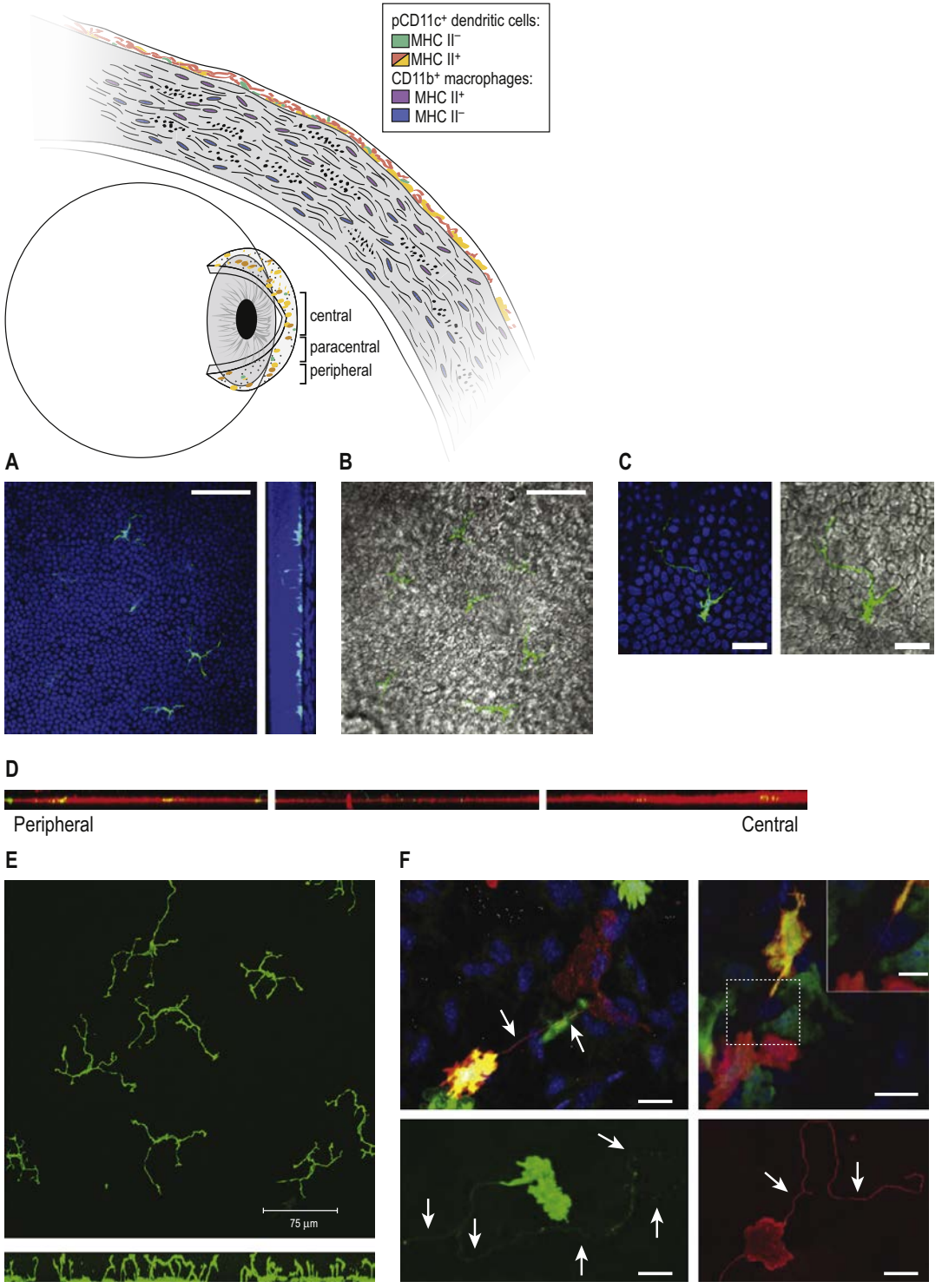


Fig. 8.2 (**A–D**) Resident myeloid cells in healthy human corneal stroma and epithelium. The healthy mammalian cornea contains dendritic cells at the epithelial cell surface with more cells at the periphery compared with the central cornea. (**A,D**) Confocal microscopy shows dendritic cells on the basement membrane of the epithelium. (**E**) CX3CR1 GFP expression by dendritic cells in the corneal epithelium of naive transgenic mice. (**F**) Nanotube formation (*arrows*) in inflamed corneas 24 hours after abrasion and exposure to bacterial endotoxin. *Green*, CX3CR1; *red*, MHC class II; *yellow*, overlap. *MHC*, Major histocompatibility complex. (Parts A–D, from Knickelbein et al., 2009, Part F, from Chinnery, 2008.)

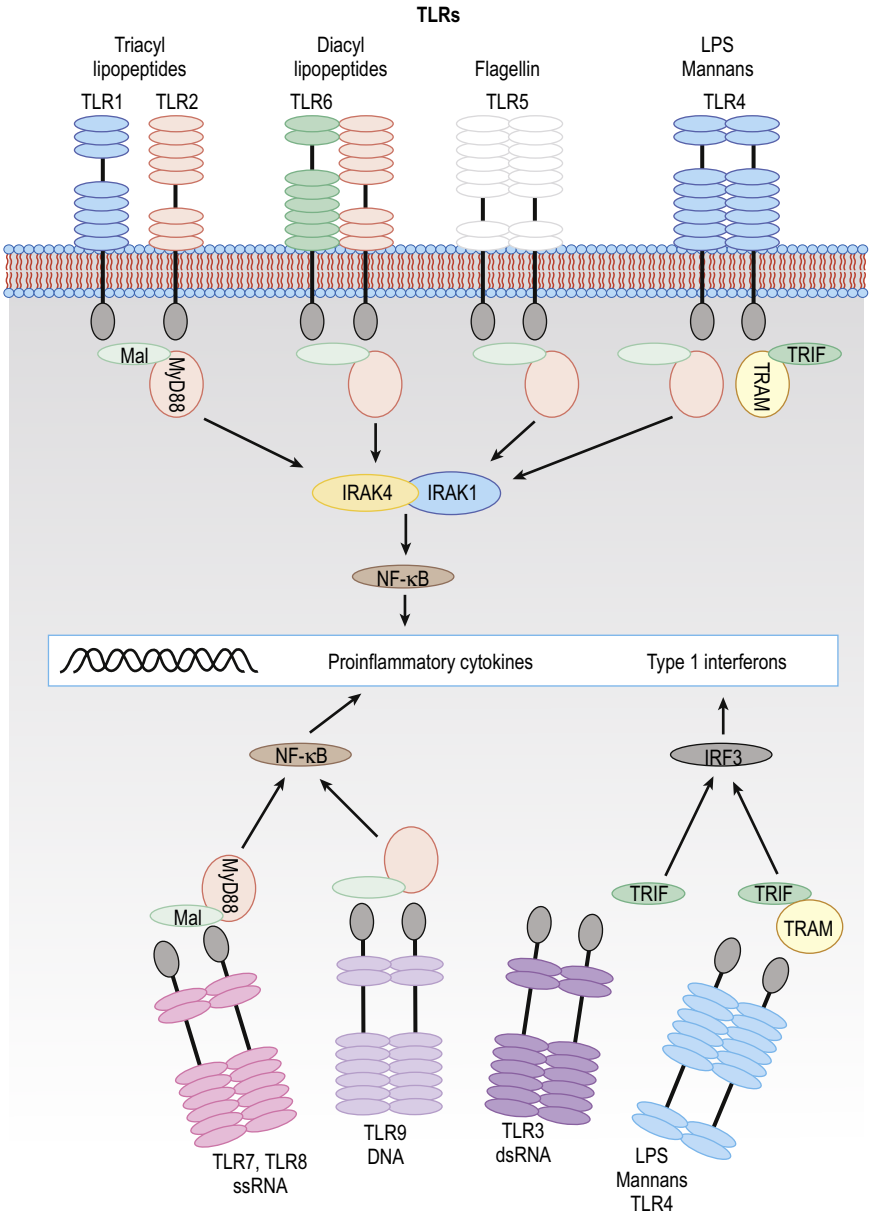


Fig. 8.3 Toll-like receptors *(TLR)* present on the cell surface recognize microbial proteins, mannans and lipoproteins, and signal through MyD88 and NFkB to produce pro-inflammatory cytokines. In contrast, TLRs in endosomes recognize nucleic acids and signal either through MyD88 or through TRIF and IRF3 to produce Type I IFNs. TLR4 on the plasma membrane activates MyD88 but can be endocytosed to activate the TRIF pathway. *ss*, Single stranded; *ds*, double stranded. (From O'Neill et al., 2013.)

infection induce expression of adhesion molecules on vascular endothelial cells. Selectins mediate tethering, intracellular adhesion molecules (ICAM-1, 2 and VCAM-1) binding to integrins on the neutrophils and chemokines mediate their transmigration across the vascular endothelium (Fig. 8.6A). Once in the tissue, migration of neutrophils to the site of infection is also dependent on a chemokine gradient, especially in an avascular tissue such as the cornea, where cells need to migrate from peripheral, limbal vessels.

Box 8.1 **NOD2 and inflammation**

Mutations in NOD2 are associated with susceptibility to autoimmune diseases that include Crohn's Disease, which is a common and painful form of inflammatory bowel disease, and Blau's syndrome, which is manifest by multiple autoimmune disorders, including a severe form of uveitis. Although not completely understood, some of the polymorphisms of NOD2 result in hyperresponsiveness to MDP from otherwise harmless commensal bacteria in the intestine and skin.

The response of neutrophils to bacteria and yeast and conidia (spores) involves phagocytosis, degranulation and neutrophil extracellular trap formation (NETs, Fig. 8.6B) (see also Chapter 7, p. 385). When neutrophils are unable to ingest the much larger fungal hyphae or *Acanthamoeba*, they can bind to the pathogen surface and release cytotoxic components, including reactive oxygen and nitrogen species, anti-microbial peptides, serine proteases and matrix metalloproteinases. They can also undergo NETosis, releasing these cytotoxic components in the context of a DNA/histone-rich net. Neutrophils also express pathogen recognition molecules, and ligand activation stimulates production pro-inflammatory and chemotactic cytokines that exacerbate neutrophil infiltration and recruit macrophages and T cells. Release of NETs by neutrophils also contributes to the loss of epithelial cells and keratocytes and corneal damage during the early stages of infection. The distribution of microglia in the retina and macrophages and dendritic cells in the uveal tract is discussed in Chapter 1.

Adaptive immunity: Although most bacterial infections are extracellular and are dealt with rapidly by innate immunity, intracellular pathogens such as mycobacteria, protozoa and especially viruses survive longer and can induce an adaptive immune response. Thus, they stimulate T- and B-cell responses, which control infection and regulate the severity of infection. For example, CD4 cells play an important regulatory role in herpes simplex keratitis, whereas CD8 cells regulate herpes latency in trigeminal ganglia and long-term exposure to airborne fungal spores induces systemic T-cell responses that regulate the severity of subsequent corneal infection. Toxoplasmosis and onchocerciasis are examples of chronic infections in which adaptive immunity plays an important role in determining disease severity and outcome. Other factors that affect adaptive immunity and increased susceptibility to microbial infections include immunosuppressive drugs and HIV infection (see also Chapter 7). Commensal organisms including bacteria (such as nasal cavity *Staphylococcus aureus*), viruses (e.g. *cytomegalovirus*) and fungi (e.g. *Pneumocystis carinii*) are initiators of opportunistic infection in immunosuppressed individuals that include HIV infected patients and patients on immunosuppressive drugs.

Contact lenses are a major risk factor for microbial keratitis

Approximately 134 million people worldwide wear contact lenses and lens wear is the most common risk factor for corneal infections in the industrialized world. Contact lenses inhibit epithelial cell proliferation and migration and suppress limbal stem cell production of basal corneal epithelial cells. Soft contact lenses, especially extended wear lenses, alter the microenvironment of the ocular surface by reducing the flow and effectiveness of tears, thereby trapping microbes at the surface of the epithelium. Furthermore, poor lens care hygiene Including lens cases and lens care solutions facilitates growth of bacteria and fungi, often forming multi-organism biofilm, which are more resistant to antibiotics and lens care solutions. Lens care solutions were found to be ineffective in killing clinical and environmental isolates of *Fusarium*, resulting in over 300 cases of keratitis in the USA, Europe and Singapore in 2005–2006. Similarly, the increased incidence of *Acanthamoeba* keratitis in Chicago in 2007 was initially thought to be due to a lens care solution, resulting in withdrawal of the product; however, later findings showed the outbreak was due to changes in chlorination of the Chicago river. A more recent outbreak in 2011 in the south-east of England continuing from an earlier outbreak in 2004 similarly failed to identify contact lens wear as a direct risk factor and that poor hygiene was a significant factor.

Viral infections of the eye

DNA viruses

Adenovirus primarily causes respiratory disease; however, depending on the serotype, adenovirus also causes a self-limiting conjunctivitis or severe epidemic keratoconjunctivitis (serotypes 3, 7, 8 and 19) or pharyngoconjunctival fever (serotypes 1, 2, 3, 5, 7 and 14). Herpes viruses are endemic and infect a large proportion of the population. Most of these DNA viruses have the potential to establish latency and to persist for the lifetime of the host. There are nine human herpes viruses (HHV 1–9) with varying levels of virulence. Cytomegalovirus (CMV) HHV-5 has the potential to infect many cells but has a preferential selectivity for myeloid cells and endothelial cells, which have a common ontogeny from the bone marrow. Patients presenting with chronic recurrent anterior uveitis frequently have CMV infections, which is probably due to latent virus persisting in uveal tissue myeloid or endothelial cells (Voigt et al., 2017). CMV infection, like

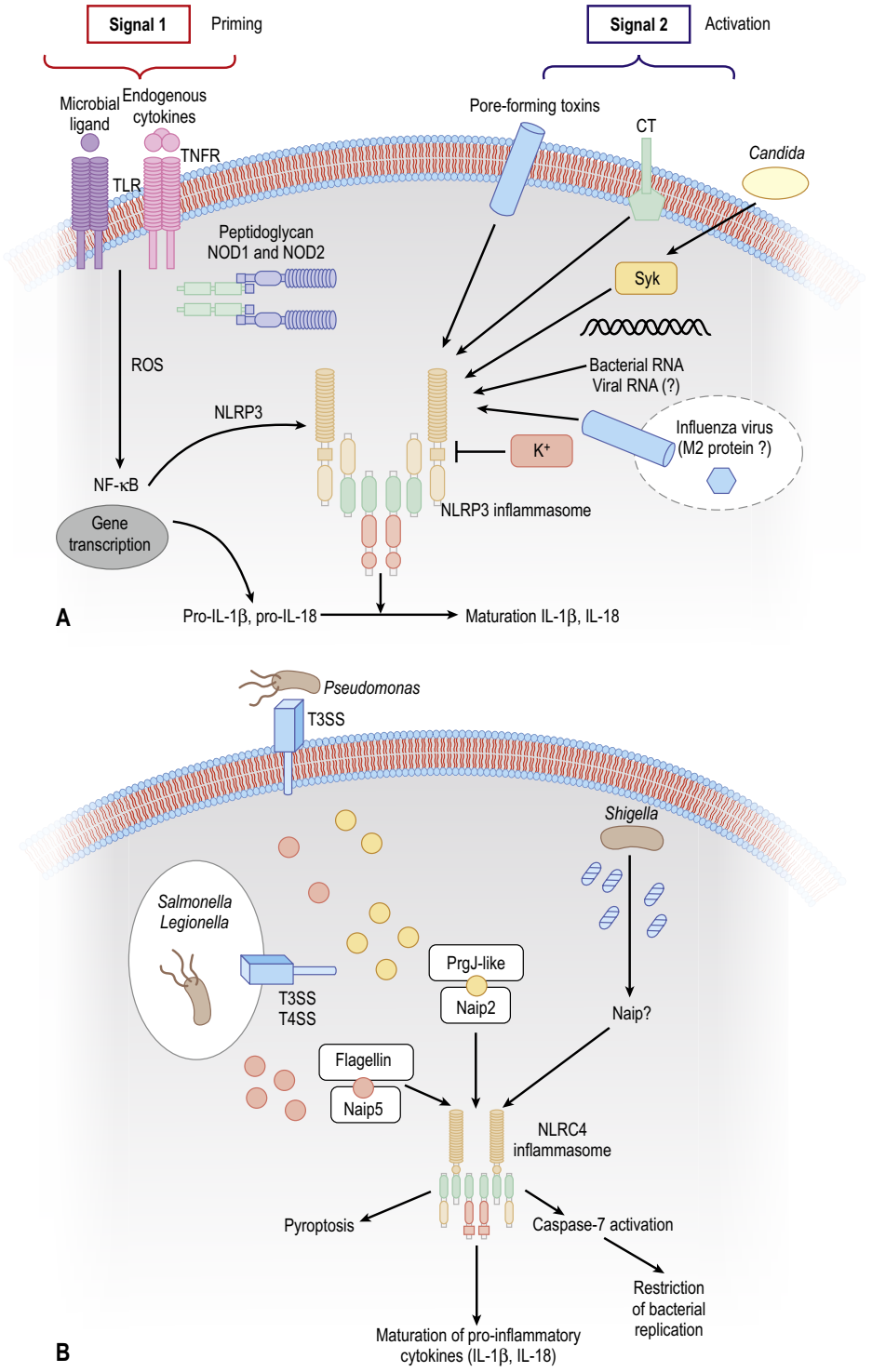


Fig. 8.4 NOD-like receptors NLRP3 (**A**) and NLRC4 (**B**) activation leads to cleavage of pro-IL-1b. Signal 1 can be toll-like receptors *(TLR)* or C-type lectin receptor activation that leads to gene expression of the IL-1 pro-form through NFkb. Signal 2 then leads to assembly of the NLRP3 or NLRC4 inflammasome complexes that then activate caspase 1, which cleaves pro-IL-1b to the bioactive, 17kD form, which is secreted. *CT*, *TNFR*, cholera toxin, tumour necrosis factor receptor. (From Franchi et al., 2012.)

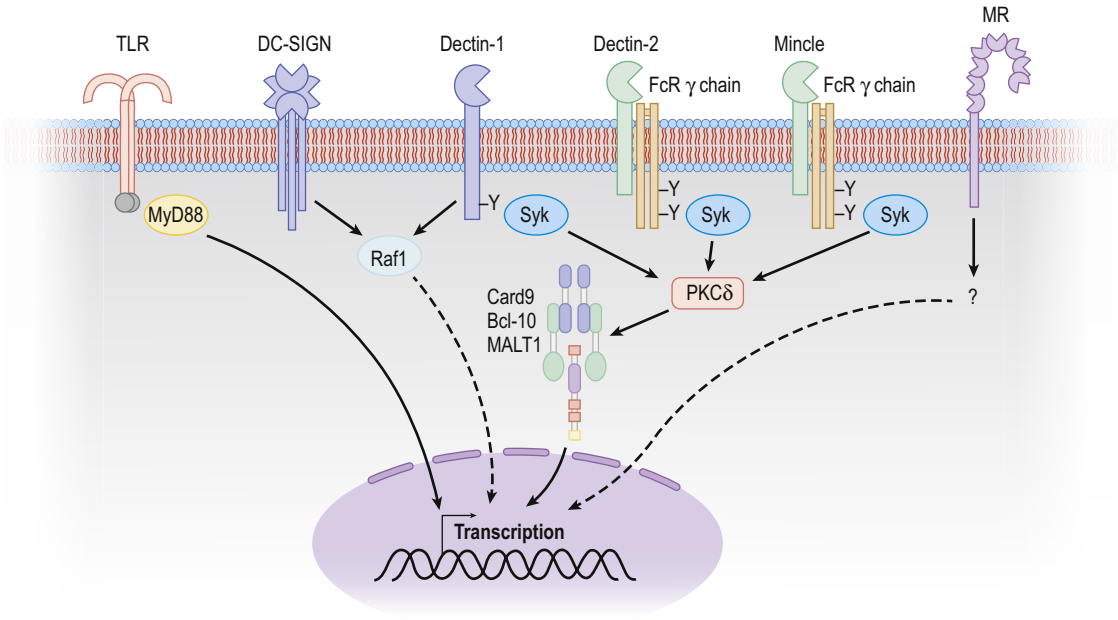


Fig. 8.5 C-type lectins recognize fungal cell wall carbohydrates and mediate antifungal immunity, including phagocytosis and cell signalling through the Syk and the Card9 complex to activate NFkB. *MR*, mannose receptor;*TLR*, toll-like receptor. (From Hardison et al., 2012.)

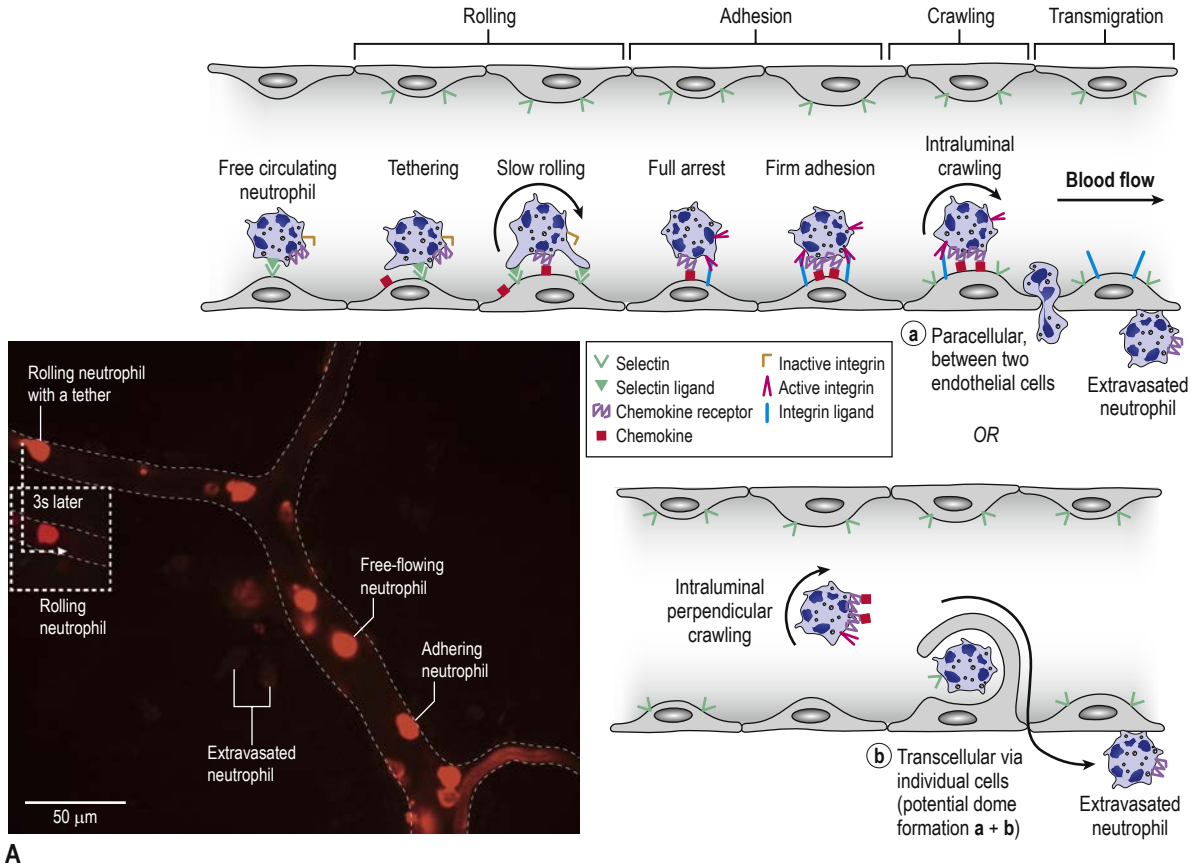
Box 8.2 **Chemotactic cytokines (chemokines)**

CXC chemokines such as IL-8 are directly (and specifically) chemotactic for neutrophils, whereas pro-inflammatory cytokines such as IL-1α, IL-1β and Tumour Necrosis Factor (TNF-α) can induce increased expression of vascular cell adhesion molecules on capillary endothelial cells in the limbus. Expression of these adhesion molecules has a critical role in tethering, binding and facilitating neutrophil transmigration into the corneal stroma (Fig. 8.6). This is a general mechanism for extravasation of leukocytes such as lymphocytes, where expression of specific adhesion molecules and chemokines on vascular endothelial cells are recognized by specific receptors on difference cell types, thereby coordinating the cellular recruitment to the tissue.

the other herpes virus infections, is widespread but is usually subclinical and remains latent throughout the lifetime of the host, thus promoting recurrent attacks of inflammation due to reactivation of the virus even in immunocompetent individuals. However, reactivation of CMV is more likely to occur and to spread to the retina (see later) when the immune system is compromised, which occurs in the elderly and in patients with immunodeficiency. Most commonly, ocular CMV is a granulomatous recurrent bilateral chronic anterior uveitis; however, in severely immunosuppressed individuals such as untreated HIV positive patients, a severe blinding retinitis can develop (see below and also Chapter 9, p. 502).

Herpes simplex virus

Herpes simplex virus-1 and -2 (HHV-1, HHV-2) are neurotropic viruses, defined as targeting neurons after initial infection of the epithelium (e.g. of cornea, skin or mucous membranes) where the virus can remain latent for the lifetime of the host. However, stimuli such as ultraviolet radiation or stress can lead to escape from latency and retrograde migration of the virus along the trigeminal ganglia to the cornea and skin, where it causes recurrent herpes simplex keratitis and skin blisters. HSV-1 is the main cause of oral, skin and ocular infections, whereas HSV-2 causes genital infections. These viruses are endemic worldwide, with the World Health Organization in 2017 estimating that 3.7 billion people under 50 years old (67% total) are infected with HSV-1 and that HSV-2 infects some 417 million individuals between the ages of 15 and 49 years. Most oral and genital herpes infections are mild or asymptomatic, whereas skin infections can cause painful blisters. Transmission is mostly oral or sexual (HSV-2). Primary herpetic corneal infections (herpes simplex keratitis) can cause severe disease due to infiltrating granulocytes, monocytes and lymphocytes; however, following initial primary infection, the virus enters sensory nerve fibres in the cornea and migrates in a retrograde manner to the trigeminal ganglia, which provides sensory



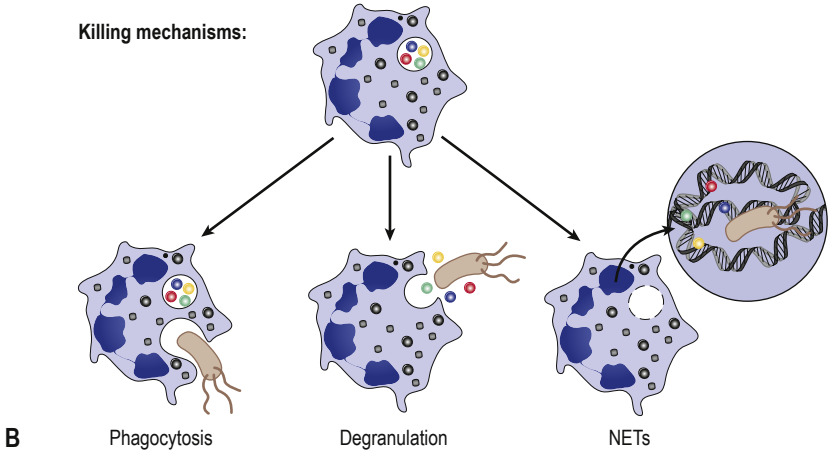


Fig. 8.6 Neutrophil recruitment to infected and inflamed tissues. (**A**) Multistep process of neutrophil migration from capillaries to sites of infection. (**B**) Antimicrobial activity of neutrophils occurs: phagocytosis, degranulation and release of reactive oxygen and proteolytic enzymes. A third mode of killing is formation of neutrophil extracellular traps *(NETs)* where neutrophils release DNA, histones and other microbicidal proteins as they undergo necrosis. (From Kolaczkowska et al., 2013.)

innervation to the cornea (see Chapter 1). HSV-1 can then enter a latent state that generally lasts for years, where viral DNA is present in neurons, but no infectious virus is produced. Following exposure to ultraviolet light, stress or immune suppression, HSV-1 is reactivated and is axonally transported into the corneal epithelium where the virus is shed in the tears and can induce an intense inflammatory and immune response. HSV-1 lesions and can be identified by characteristic dendritic-like lesions (Fig. 8.7A).

Host response to primary herpes simplex infection. TLR9 expression in corneal epithelial cells is required for the initial host response to HSV-1 infections and can be induced by HSV-1 DNA alone. HSV-1 infected cells produce type 1 interferons (IFN-α/β), which inhibit viral replication by activating natural killer cells and CD8 cells (see Chapter 7). $CD4^+$ Th1, Th17 and T regulatory (Treg) cells also play an important role in limiting the primary response to HSV-1, with the suppressive activity of Treg cells balancing the pro-inflammatory and pro-angiogenic activity of Th1 and Th17 cells.

Angiogenesis and lymphangiogenesis in herpes simplex keratitis. Blood and lymph vessel formation of the normally avascular cornea is a characteristic feature of HSK (Fig. 8.7B) and is important in initiating an adaptive immune response to the virus. In this case, lymphatic vessels transport viral antigens to draining nodes, and blood vessels transport mature HSV-1–specific T cells to the cornea. However, formation of new blood vessels (angiogenesis) also impairs visual acuity, although angiogenesis is regulated by vascular endothelial cell growth factors and receptor antagonists.

Herpes zoster virus (HZV)

HHV-3 (also known as *Varicella zoster* virus) is also a neurotropic virus and establishes latency after infection in infancy and childhood in a similar manner to herpes simplex. Primary infection occurs in the lungs where the virus proliferates and can infect adjacent nerve cells. In childhood, HZV presents as a widely distributed exanthem with a typical rash (chickenpox rash). In adults, HZV commonly occurs as a late reactivation of a latent infection acquired in childhood (~10% of cases). Infection in this case is usually restricted to a single dorsal root ganglion or cranial nerve and commonly presents as shingles where infected surface epithelial cells in the distribution of the sensory neurone develop into a crusting vesicular rash, which can be extremely painful. Herpes zoster ophthalmicus (HZO) involving the neurons in the trigeminal ganglion and is a common presentation of HZV infection. Involvement of the trigeminal (Vth) cranial nerve (see Chapter 1, p. 73) implicates the other tissues served by the Vth nerve, including ocular and orbital structures. HHV-3 can then lead to adverse sequelae, including uveitis, glaucoma, keratitis, secondary cataract, ocular hypotension resulting from severe uveitis, lid and ocular muscle defects, and the orbital apex syndrome (visual loss caused by optic neuritis and multiple oculomotor pareses). Other herpes viruses can cause viral retinitis in immunocompromised patients (see later).

Epstein–Barr virus (EBV)

Also known as HHV-4, EBV is the virus that causes infectious mononucleosis (glandular fever). EBV infects epithelial cells and B cells, where it persists as a latent infection for the entirety of the patient's life. It is highly contagious, with ~90% of adults displaying evidence of previous subclinical or symptomless infection in the form of positive serology (anti-EBV antibodies). Patients may present with ocular symptoms such as mild uveitis or non-specific retinal signs, but definitive linkage to EBV infection is difficult because of the large number of infected individuals without symptoms. EBV has been implicated in conditions such as retinal vasculitis and multiple sclerosis. More definite causative linkage occurs as a consequence of metastatic dissemination, resulting in haematological malignancies to the eye and other tissues.

RNA viruses

Measles virus

Rubeola virus (Measles virus *morbillivirus, Mv*) is a single-stranded RNA virus, with a negative sense nucleotide sequence (see Chapter 3, p. 139). The measles virus envelope contains two proteins that mediate its entry into cells: haemagglutinin (H) and membrane fusion (F) proteins. Through these, it binds the host protein CD46 (a complement regulatory protein), the lymphocyte activating protein (SLAM, CD150) and an adhesion protein (nectin/PVRL4 –primary vitreoretinal lymphoma4 protein). Measles is highly contagious and can be fatal during the acute stage because the virus can also cause encephalitis. The spread of dis-information regarding pathogenicity and risk has led to reduced levels of measles vaccination and to recent outbreaks of the disease in Europe and the USA. Measles infection in the third world is a major cause of childhood morbidity and mortality.

Measles virus can cause conjunctivitis and keratitis during the acute infectious stage of disease, which can result in permanent corneal scarring with blindness. In malnourished children with vitamin A deficiency, measles keratitis can be the trigger for spontaneous corneal ulceration and perforation of the globe with total loss of vision. It is estimated that up 60,000 cases of childhood blindness occur as a result of measles infections in malnourished children. Furthermore, CNS involvement with measles results in encephalitis and the measles virus can cause optic neuritis and retinitis (see later).

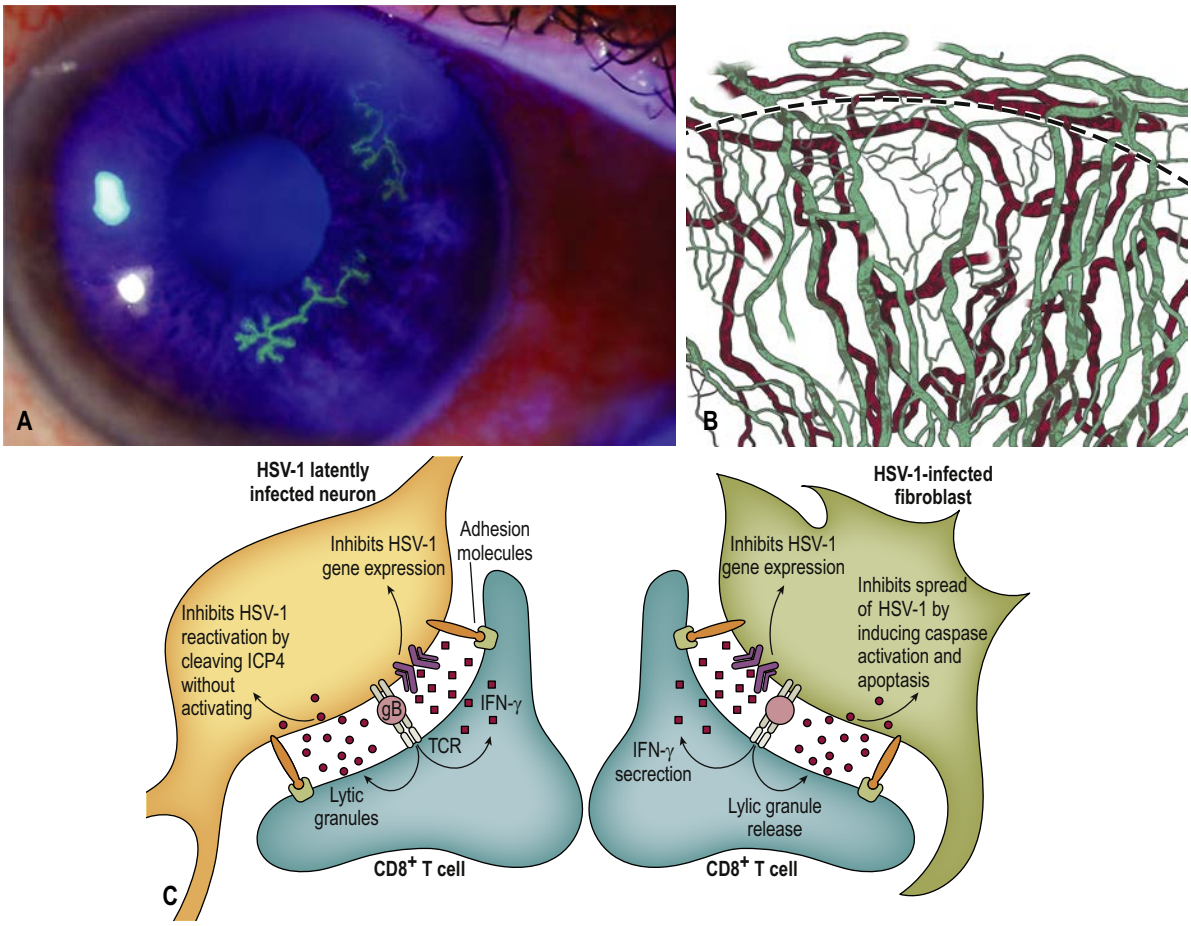


Fig. 8.7 (**A**) Herpes simplex keratitis *(HSK)* showing characteristic dendritic-shaped lesion. HSK is also characterized by pronounced corneal angiogenesis. (**B**) HSV-infected whole mount murine cornea showing growth of capillaries (CD31) and lymphatic vessels (LYVE-1) from the limbus to the cornea. (**C**) HSV-specific $CD8^+$ T cells, interactions with HSV-1 latently infected neurones and lytically infected fibroblasts. IFN-γ and lytic granules inhibit HSV-1 reactivation from latency through a non-lytic mechanism; therefore, virus emerges only when T-cell function is compromised. However, lytic granules induce apoptotic cell death of infected fibroblasts. (From Knickelbein et al., 2008.)

Rubella virus

Rubella virus (also known as rubeola or German measles) belongs to the Togavirus family and causes a self-limited, mild illness characterized by a low-grade fever, enlarged lymph nodes and a typical rash. It is a positive sense, single-strand RNA virus That is transmitted through respiratory secretions (coughing, sneezing, etc.). It is universally infective but can be controlled effectively by vaccination. Its major effect on the eye is through maternal vertical transmission to the fetus or directly in early infancy where it can cause numerous defects including microphthalmia, congenital glaucoma and cataracts and retinitis, manifested as salt and pepper retinopathy. These ocular symptomscan present together with central nervous system and cardiac symptoms.

Ebola virus

Ebola virus (EBOV) is a single-stranded anti-sense RNA virus and member of the Filoviridae, which was identified in 1976 in Zaire during an outbreak of haemorrhagic fever. It was originally thought to be transmitted to humans from

non-human primates. EBOV is highly virulent with mortality rates ranging from 50% to 90% depending on which of the five viral strains is causing the infection. Transmission is almost exclusively by contact with bodily fluids and viral entry occurs through cuts or abrasions. In 2019, after 2500 individuals were infected and almost 1700 killed in the Democratic Republic of the Congo, the WHO declared Ebola to be a Global Health Emergency. This designation allows for increased resources to control the epidemic.

Myeloid cells, including dendritic cells and tissue macrophages, are the major target of EBOV, which binds to the dendritic cell intercellular adhesion molecule specific-3 integrin (DC-SIGN), TAM-receptor tyrosine kinases or T-cell immunoglobulins and mucin domain-1 protein (TIM-1). These phosphatidylserine (PS) receptors bind to the viral envelope and mediate uptake into endosomes. In this pathway, viral envelopes mimic apoptotic cells that express PS on the outer layer of the plasma membrane. The virus can inhibit dendritic cell maturation, thereby impairing development of a robust adaptive immune response. EBOV can induce a cytokine storm that contributes to haemorrhagic fever manifestations, including disseminated intravascular coagulopathy, and resultant organ failure.

In individuals who survive the initial infection, the virus is retained in immune privileged tissues, including the eye and the testis. Up to 40% of survivors can experience disease reactivation weeks or months after recovery and virus can be isolated from joint fluid, semen, CSF and ocular fluids. The resurgent virus causes severe tissue damage: in a study of 96 survivors of the 2014–2016 outbreak in West Africa, almost 40% were blind as a result of anterior uveitis and others had cataract or vitreoretinal pathology. It is unclear whether this RNA virus actually becomes latent, i.e. non-replicating but viable, or persists as a chronic infection with a very low replication rate (see later). Viral persistence is at least in part a result of the ability of Ebola viral protein 35 (VP35) to inhibit production of type I interferon that is required for viral clearance.

Although several vaccines have been developed, the vaccine being used in recent outbreaks is the recombinant vesicular stomatitis virus–Zaire Ebola virus (rVSV-ZEBOV) where the VSV expresses an EBOV protein. The vaccine was developed in Canada and is licensed to Merck where it has been used in recent outbreaks, given not only to patients, but also those in contact with infected individuals.

Vector-borne viruses

Flaviviruses

Flaviviruses are positive-strand RNA viruses with a 10.5 kb genome that encodes 10 proteins: 3 structural and 7 non-structural proteins that mediate viral replication, budding and interactions with host cell proteins. These include Zika, Dengue, West Nile and Yellow Fever viruses. These viruses all cause ocular manifestations as described in Table 8.2.

Zika virus

The World Health Organization declared Zika virus a *Public Health Emergency of International Concern* in 2016 following multiple cases of neurological disease in Brazil and other countries in South America and in Puerto Rico and the southern USA. Transplacental infection resulted in microcephaly and ocular damage, mainly macular and optic nerve abnormalities causing blindness. Although the incidence of disease has dropped since its peak in 2015 and 2016 and there were no outbreaks in 2018, this mosquito-borne virus

Table 8.2 **Comparison of the major ocular findings among the different flaviviruses**

Ocular Complication	ZIKV	DENV	JEV	WNV	YFV	KFDV
Conjunctivitis/keratitis	+	+	–	–	–	+
Macular mottling	+	+	–	–	–	–
Chorioretinal atrophy	+ focal pigmentary clumping	+	–	–	–	+
Optic nerve abnormalities	+	+	–	+	+	–
Cataract	Hypoplasia, cupping, pallor	–	–	–	–	–
Microphthalmia	+	–	–	–	–	–
Iris coloboma	+	–	–	–	–	–
Uveitis	+	+	–	+	–	+
Chorioretinitis	+	+	–	–	–	+
Retinal hemorrhage	+	+	+	–	–	+

ZIKV, Zika virus; *DENV,* Dengue virus; *JEV,* Japanese encephalitis virus; *WNV,* West Nile virus; *YFV,* Yellow Fever virus; *KDFV,* Kyasanur Forest Disease virus.

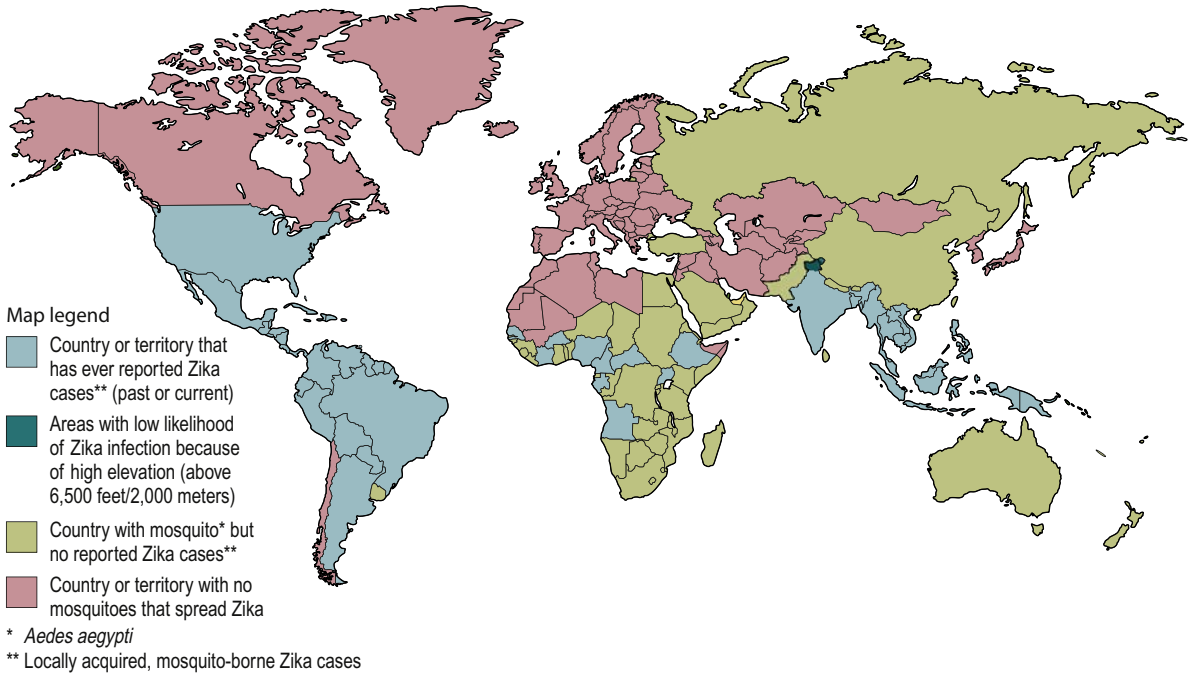


Fig. 8.8 Zika virus prevalence. (From Centers for Disease Control, 2017.)

still poses a risk. Fig. 8.8 shows past and recent prevalence of Zika virus worldwide.

eLink 8.1 Further information on the epidemiology of Zika virus is available here https//:expertconsult.inkling.com

Transmission. Zika virus is an RNA-type Flavivirus, a family that also includes Dengue, West Nile, Yellow fevers and Japanese encephalitis virus and all transmitted through the *Aedes* genus of mosquitoes (these viruses are also termed arboviruses, meaning they are transmitted by arthropod vectors). Zika virus can also be transmitted vertically from the infected mother to the fetus by the transplacental route. However, a unique feature of Zika virus is that it can also be transmitted sexually because it can directly infect primary spermatocytes and Sertoli cells where it can persist (so far detectable 188 days post infection). Sexual transmission played an important role in the Zika virus outbreak in Rio de Janeiro (Fig. 8.9A).

Clinical manifestations. Acute infection usually causes a mild flu-like illness that resolves within a few days. However, the outbreaks in Polynesia and South America also caused fever, rash, myalgia and arthralgia, with severe cases resulting in multi-organ failure, meningitis and encephalitis, and Guillain–Barre Syndrome (Fig. 8.9B), an infection-triggered autoimmune disorder with peripheral and autonomic nerve damage leading to muscle weakness, pain, altered sensation and in severe cases respiratory failure. Transmission of infection to the fetus leads to significant CNS damage.

eLink 8.2 A description of Zika virus fetal damage is available here https//:expertconsult.inkling.com

As a consequence, infants with microcephaly exhibit learning disabilities, mental retardation and behavioural problems, with more severe cases also having epileptic seizures.

Ocular abnormalities. Acutely infected healthy individuals infected with Zika virus frequently develop a non-purulent conjunctivitis that eventually subsides. Patients may also develop anterior and posterior uveitis.

Ocular damage in both microcephalic and non-microcephalic infants born to presumed Zika-infected mothers, particularly in the first trimester, exhibited bilateral retinal lesions, predominantly pigment mottling and often severe chorioretinal atrophy (Fig. 8.11A). Optic disc hypoplasia (Fig. 8.11B) microphthalmia, congenital glaucoma, iris colobomas, cataracts and intraocular calcifications have also been observed.

eLink 8.1 EPIDEMIOLOGY OF ZIKA VIRUS

Zika virus was first described in a rhesus monkey in the Zika rain forest in Uganda in 1947 and was later documented in several cases in sub-Saharan Africa and south-east Asia. The first epidemic was reported in 2007 in Yap island in Micronesia where >75% of the population was infected and presented with rash, arthralgia and conjunctivitis that was attributed to Zika virus. A subsequent outbreak in French Polynesia in 2013 affected 29,000 individuals. The first outbreak of Zika virus in the Americas was in 2015 in Bahia, Brazil, and the second was in Puerto Rico in 2015, with an estimated 37,000 cases. According to a 2018 report from the Centers for Disease Control (CDC), there have been 5,716 cases in the USA mainland since 2016, including 230 proven endogenous cases that were transmitted by mosquitoes in the USA. 108 cases were reported between January and August 2018 mostly from Florida and Texas.

A 2017 World Health Organization report noted that 31 of 84 countries that have vector-borne transmission reported microcephaly and other congenital malformations, with 6% reported in the USA compared with 46% in Brazil.

eLink 8.2 ZIKA-INDUCED FETAL ABNORMALITIES

Vertical transmission of Zika virus to the fetus follows infection of cells of the placenta, including trophoblasts, endothelial cells and placental villi macrophages. Zika virus also infects human neural progenitor cells and fetal neurons, but not mature neural cells. These observations are consistent with the findings that developing fetuses in the first trimester of pregnancy are more susceptible to Zika virus infection which results in neural cell death and developmental abnormalities. Further, *in vitro* studies show that Zika virus preferentially infects progenitor cells of the neural cortex, which inhibits cell proliferation and differentiation and increase inflammation and cell death. Infection of microglia cells results in production of pro-inflammatory cytokines that likely contribute to cell death of neuronal progenitor cells. However, in the Brazil outbreak where 182 women at different stages of pregnancy exhibited signs of infection, 42% of fetuses had clinical or brain imaging abnormalities. Congenitally infected infants and neonates develop hypoplasia of the brainstem and cerebellum, microcephaly, hearing loss and ocular abnormalities (see later).

Microcephaly is defined as head circumference that is less than the third percentile for gestational age and sex (Fig. 8.10) and occurs as a consequence of impaired brain development, with skull size correlating with underlying brain size. As a direct approach to see whether Zika virus can cause microcephaly, Li et al. infected embryonic mice and found that Zika directly infects and replicates in neuronal progenitor cells in the brain, resulting in smaller brains, with a thinner cortex, consistent with microcephaly.

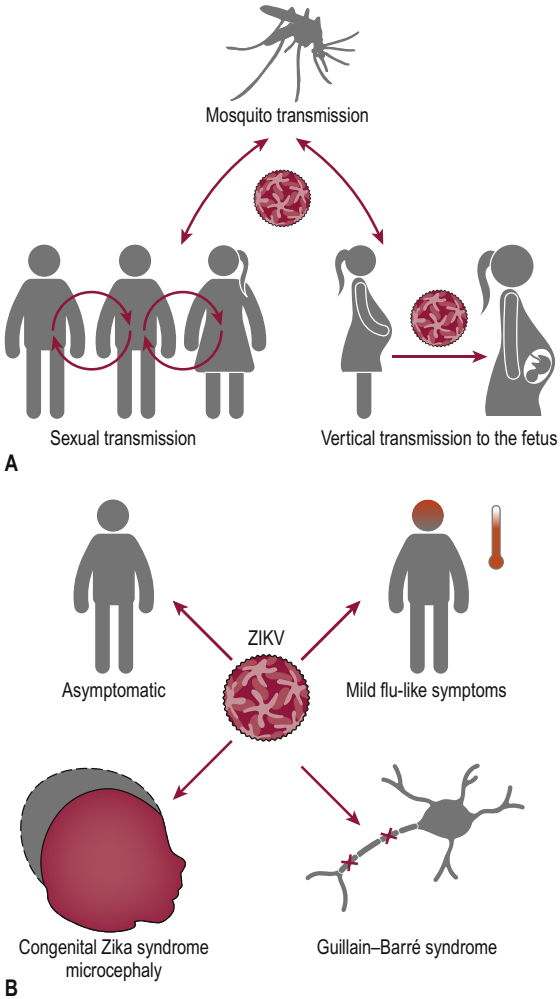


Fig. 8.9 Zika virus transmission. (**A**) Zika virus is transmitted by mosquitoes, by sexual intercourse, and transplacentally. (**B**) Most infections of otherwise healthy adults are asymptomatic or cause mild, flu-like symptoms; however, Zika virus can also cause neurological disease, including Guillain–Barré syndrome. Congenital infections result in neuronal and ocular abnormalities, including microcephaly. (From Pierson and Diamond et al., 2019.)

eLink 8.3 Further information on the replication and immunology of Zika virus infection is available here https//:expertconsult.inkling.com

Treatment and vaccines. An increased understanding of the adaptive immune response to Zika virus and the other Flaviviruses is important in the development of vaccines for susceptible populations, in addition to vaccines aimed at blocking dissemination of the virus during an epidemic. Several approaches are underway, including vaccines based on non-infectious sub-viral particles, inactivated (infectious) virus and attenuated live chimeric vaccines. A DNA–plasmid vaccine encoding membrane and envelope proteins is also underway. In contrast to dengue virus, Zika virus has a single serotype, making vaccine development relatively uncomplicated. Therapeutic approaches to Zika virus have multiple targets, including caspase 3 to prevent cell death, peptide targets to prevent viral entry, chloroquine that disrupts viral autophagy and viral release after endocytosis and the anti-cancer drugs temoporfin and tribavirin (ribovirin) that inhibit viral replication. Development of new and more targeted therapeutics is underway.

Dengue virus, West Nile virus and Japanese encephalitis virus (JEV)

Viral latency

Latent viral infection can occur in any cell type, although certain cells provide the correct environment to promote latency. A pattern of infection has emerged where infection occurs in the prenatal or neonatal period or in infancy, after which the virus becomes latent and persists for the life of the host. The outcome depends in part on the virulence of the virus (Ebola is much more likely to kill than CMV) and on the ability of the host immune response to contain the infection (HSV-1 latency is maintained in the trigeminal ganglion by CD8 cells). When the host immune response is impaired as occurs in AIDS patients, elderly patients or patients taking immunosuppressive drugs, the ability to contain the virus(es) is lost and the virus reactivates and infects tissues that are not infected in immunocompetent individuals. It is remarkable that so many healthy individuals maintain viruses in a latent state and survive to old age without experiencing a reactivation, a concept that does not just apply to viruses but to pathogens in general (Forrester et al., 2018).

Each pathogen adopts a distinct approach to achieve latency. In the course of primary HSV-1 corneal infection, the virus infects the epithelial cells and enters the axons of sensory neurons, after which it is transported in a retrograde manner to the cell bodies, located in the trigeminal ganglia. Viral DNA is either inserted into the DNA nucleus or is present in the nucleus or cytoplasm as an episome (see Chapter 3, p. 139). This episomal stage of the virus is regulated by HSV-1 latency-associated transcripts (LAT), which maintain the virus in its latent stage and inhibit apoptotic cell death by blocking caspase activity. Also, HSV-1 specific $CD8^+$ T cells are found in close association with infected neurons and can form immunological synapses in which the T-cell receptor and CD8 molecules are in the same proximity as Class I receptors on the neurons

eLink 8.3 ZIKA VIRUS REPLICATION AND IMMUNOLOGY

Replication

Viral envelope proteins enter the host cell following receptor-mediated endocytosis, the envelope is degraded in lysosomes and the released RNA is translated on the membrane of the endoplasmic reticulum. Viral proteins undergo post-translational modifications in the Golgi apparatus and the assembled virions are released by exocytosis. Identification of the host cell surface receptors for Zika virus remains under investigation, but candidates to date include the AXL receptor tyrosine kinase on multiple cell types, DC SIGN on dendritic cells and the phospatidylserine receptor (PS) TIM-1 on skin fibroblasts, endothelial cells and fetal neurons, which recognizes PS in the viral envelope and mediates uptake into endosomes.

Immune response to Zika virus

Innate immunity: The critical response to viral infections in general is production of Type I family of interferons that include IFN-alpha and IFN-beta and Type III IFNs, primarily IFN-lambda. IFN-alpha and IFN-beta bind to the broadly expressed heterodimeric receptor IFNAR1/2 to produce a number of proteins that can restrict viral replication, whereas IFN-lambda binds the more restricted IFNARL1/IL-10R2. Zika virus can subvert the IFN response by inhibiting signal transduction pathways. Murine models of infection have yielded considerable insight into the pathogenesis of Zika virus. Immune-competent mice are refractive to infection; however, mice lacking type I IFN receptors are susceptible to infection, which results in neuronal symptoms and death due to unrestricted growth of the virus. Similarly, pregnant mice that are either IFNAR deficient or by neutralizing type I IFNs are highly susceptible to Zika virus in the first trimester, resulting in infection of trophoblasts, fetal brain damage and death. Infection at later stages of pregnancy had no adverse effects on fetal development, thereby reproducing observations in infected individuals. Infection of pregnant non-human primates resulted in similar outcomes as described for humans. Given the sylvatic cycle predicted for Zika virus spread to humans, these observations have important implications.

Adaptive immunity: IgM is generated within days of infection, but neutralizing IgG does not develop until at least one week post-infection. There is also cross-reactivity of neutralizing antibodies between Zika virus and other Flaviviruses, including dengue, and anti-human monoclonal antibodies to dengue can also neutralize Zika virus. CD8 cells are also important in the host defence to Flaviviruses. In murine models, cytotoxic CD8 T cells can restrict Zika virus growth, as depletion of CD8 cells results in increased morbidity and mortality.

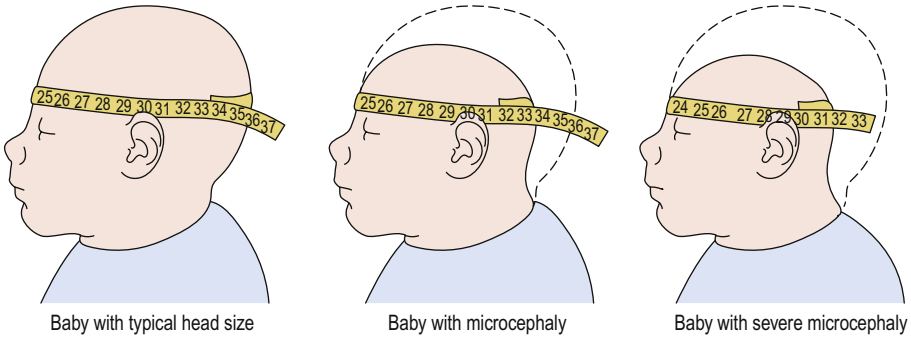


Fig. 8.10 Microcephaly caused by vertical transmission of Zika virus. Infants born to infected mothers can have severe neurological and physical defects, including microcephaly, resulting from decreased brain tissue and impaired development of the skull. (From Freitas et al., 2016)

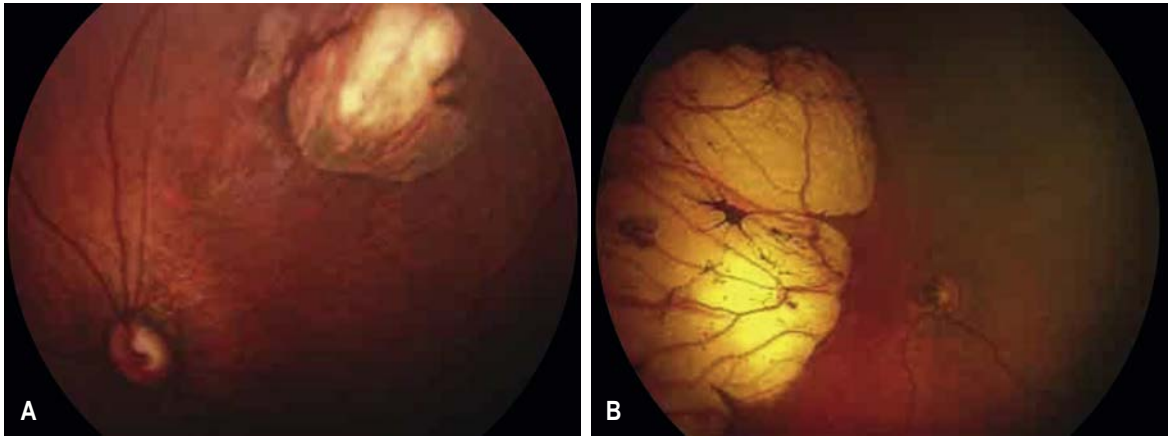


Fig. 8.11 Zika virus retinal abnormalities. (**A**) Peripheral temporal chorioretinal scar and atrophy with pigmentary, perilesional mottling in a 1-month-old infant with microcephaly. (**B**) Optic disc hypoplasia and large nasal chorioretinal atrophy and scar in a 20-day-old microcephalic infant. (From Freitas et al., 2016.)

(Fig. 8.7C). Although CD8 production of IFN-γ is important, release of perforin in CD8 cell lytic granules is essential to maintain latency without killing the neurons.

HSV-2 and VZV adopt a similar mechanism to achieve latency in neurons and can be treated using drugs such as Acyclovir (and derivatives), which targets viral DNA polymerase. The location of CMV episomes is less well known since there are definitive LATS. However, it is generally accepted that myeloid cells are the target for latent infection, including bone marrow $CD34^+$ myeloid progenitor cells. Viral transfer in progenitor cells allows for long-lived persistence of latency and infect myeloid cells as they mature and populate the tissues as resident macrophages and dendritic cells, where they provide a continuing source of virus that can reactivate *in situ* and cause inflammation and tissue damage. This mechanism may explain how CMV infection induces chronic recurrent anterior uveitis (see earlier).

In CMV, there are three forms of latent infection that are characterized by the expression of specific viral proteins (EBNA 1-3C, LMP1-2B) and a non-coding RNA (EBER). EBV can mutate through the three stages in B cells but only develop as far as Latency I/II in epithelial cells in which only some of the required proteins are expressed. In myeloid cells, it is likely that Latency III, in which all latency proteins are expressed, occurs in B-cell progenitors and thus allows for the lifelong persistence of latent infection. During the acute infection, this occurs in the salivary glands probably in tertiary lymphoid organs (see Chapter 7, p. 420) and in the bone marrow but may also take place in long-lived memory B cells in the tissues in adulthood.

The question of latency in RNA viruses such as EBOV and Zika is less clear since they depend much more on the downstream functions of the infected cell. As indicated previously, latency, dormancy and persistence are terms that can be confused; however, a working definition might depend on whether the organism is truly non-replicating but viable, in which case it would be truly latent and a persistent, whereas a very slowly replicating organism might be considered to generate a subclinical/low-grade chronic persistent infection.

Viral retinitis

The most common causes of viral retinitis include the herpes virus family members herpes simplex virus, varicella zoster virus and human cytomegalovirus (CMV). Less common but significant causes of viral retinitis are lymphocytic choriomeningitis virus (LCMV), Epstein–Barr virus, rubella, measles, West Nile virus, dengue and the Chikungunya virus, which is considered an emerging pathogen. Primary CMV infection is controlled by $CD8^+$ T cells and, as with all herpes viruses, CMV infects neurons where it remains persistent over the lifetime of the host. CMV infection usually occurs during a period form *in utero* to infancy and can cause congenital infection, such as chorioretinitis, hydrocephalus and microcephaly. LCMV is a single-stranded RNA arenavirus and humans are infected by inhalation or ingestion of mice, particles contaminated with mouse faeces, urine or saliva. Although viral cytotoxicity is a significant cause of direct retinal damage, an exacerbated inflammatory and immune response can also occur in the absence of regulation and can contribute to the severity of retinal damage (see Chapter 7). Indeed, experimental models of CMV and LCMV retinitis in which the virus is injected into the vitreous show rapid dissemination to the lymph nodes and spleen, production of virus-specific T cells and invasion of the retina by cytotoxic T cells, which can cause significant tissue damage and retinal disease.

As indicated, HIV patients are at higher risk for CMV, HSV and VZV in particular producing diseases such as acute retinal necrosis and progressive outer retinal necrosis (see Chapter 9, p. 502. The advent of highly active anti-retroviral therapy (HAART) has reduced the incidence of viral retinitis; however, a long-term study of CMV retinitis patients with AIDS taking HAART showed that these patients still have an increased risk of progression of retinitis with complications and visual loss.

Bacterial infections of the eye

The ocular surface and bacterial infections

Bacteria that cause conjunctivitis and blepharitis include *Streptococcus pneumoniae* and *Staphylococcus aureus*. These infections result in conjunctival inflammation and irritation but do not affect vision. In contrast, bacterial infections of the cornea are a major cause of blindness and visual impairment worldwide. Once bacteria gain entry to the corneal stroma, they can replicate quickly, and induce a rapid neutrophil infiltrate that contributes to vision loss.

***Pseudomonas aeruginosa*:** *P. aeruginosa* is the most common cause of bacterial keratitis worldwide. These Gram-negative bacteria are ubiquitous in fresh water and contact lens wear and corneal injury are the major risk factors. Although *P. aeruginosa* induces a rapid host response in the cornea, most clinical isolates express virulence factors that counter this response and facilitate bacterial survival. However, *P. aeruginosa* produces exotoxins as part of the Type III Secretion System (T3SS), which comprises a needle-like structure that injects bacterial exotoxins directly into host cells. The T3SS is expressed in clinical isolates and plays an essential role in bacterial survival *in vivo*. As shown in Fig. 8.12, the needle structure protrudes from the surface of the bacterium and forms a pore in the host-cell membrane through which ExoS, ExoT or ExoU are injected. ExoU causes rapid cell lysis and severe corneal disease as a result of its phospholipase activity, whereas ExoS and ExoT are related proteins that have ADP ribosyl transferase and GTPase activating protein (GAP) enzymatic activity. Most functions of ExoS and ExoT have yet to be determined; however, ExoS ADPRT directly inhibits NADPH oxidase assembly by animal models show that expression of specific T3SS exotoxins and needle structure is essential for bacterial survival and replication in the cornea. Although almost all clinical and environmental isolates express either ExoS or ExoU, they do not produce both exotoxins. Strains expressing ExoU but not ExoS have been designated cytotoxic strains, whereas ExoS expressing strains are considered invasive as they can replicate in corneal epithelial cells without causing rapid lysis.

The host response to *P. aeruginosa* in the cornea has been extensively studied using murine models, and although adaptive immunity can be demonstrated at later stages of infection, innate immune responses are essential to regulate the rapid growth of these bacteria in the cornea. Experimental evidence supports the concept that *P. aeruginosa* LPS and flagellin activate TLR4 and TLR5 on resident macrophages, which produce CXC chemokines and pro-inflammatory cytokines that mediate neutrophil recruitment from limbal vessels to the corneal stroma. Inhibition or blockade of TLR4/5, intracellular signalling pathways, IL-1R or neutralization of IL-1α/β or CXC chemokines impairs neutrophil infiltration and function, allowing the bacteria to survive, resulting in more severe corneal disease. Expression of these mediators is also elevated in corneal exudates from infected individuals. Suppressor molecules such as members of the TLR family, neuropeptides such as vasoactive intestinal protein (VIP), and even high concentrations of flagellin also have a suppressive effect on the

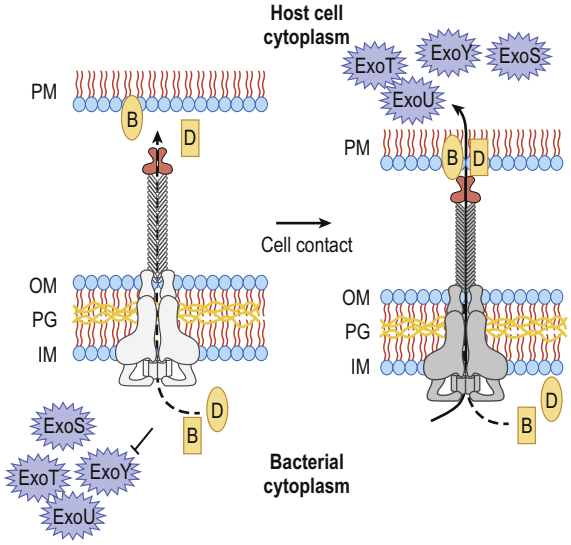


Fig. 8.12 *Pseudomonas aeruginosa* type III secretion system *(T3SS)*. The T3SS spans the bacterial cell envelope including the inner membrane *(IM)*, outer membrane (*OM)* and peptidoglycan layer *(PG)*. Following contact with the host cell plasma membrane *(PM)*, the translocator proteins PopB *(B)* and PopD *(D)* form a pore-forming complex with the needle tip *(PcrV, red)*, and the effector proteins are exported through the bacterial cell envelope and the needle complex and into the cytosol of the targeted cell. To date, four effector proteins have been described in *P. aeruginosa*, ExoS, ExoT, ExoY, and ExoU, although the exact complement of effectors varies among strains and clinical and environmental isolates. Effector proteins are either directly cytolytic (ExoU) or modulate cell function to facilitate bacterial survival in the host cell. (Diagram by A. Rietch, from Pearlman et al., 2013.)

host inflammatory response, although the mechanisms have yet to be fully characterized. Ultimately, a therapeutic approach targets specific inflammatory pathways rather than using broadly acting corticosteroids and may have greater potential to inhibit inflammation without side effects when given together with antibiotics.

Streptococcus pneumoniae is a common cause of bacterial keratitis, especially in developing countries. As a normal commensal in the lungs and upper respiratory tract, *S. pneumoniae* is frequently isolated from the conjunctival sac of healthy individuals and corneal infection is likely to be associated with traumatic injury. *S. pneumoniae* produces several exotoxins; however, the most potent is pneumolysin, which can directly form pores in the host cell membrane, leading to rapid leakage and cell death. Pneumolysin is released as a monomer, but forms a multimer on the host cell membrane, leading to pore formation and rapid cell death. In contrast, there is no apparent correlation between the serotype expressed and severity of infection.

Staphylococcus aureus causes less severe disease than *Streptococcus* or *Pseudomonas* and is associated with peripheral ulcer formation. However, antibiotic resistant isolates, including methicillin resistant *S. aureus* (MRSA), make treating this infection more challenging. *S. aureus* also produces several virulence factors including haemolysin α, β and γ, which facilitate bacteria survival *in vivo*. Clinical observations have suggested that it may possible to predict the type of organism that is infecting the cornea by analysis of the clinical presentation on slit lamp microscopy.

Contact lenses and bacterial infection

The incidence of contact lens-associated infections appears to be increasing partly because of increased contact lens usage but also perhaps because of increased antibiotic resistance to infection (Cheung et al., 2016). Other organisms including *Serratia marcescens* and coagulase negative staphylococci such as *S. epidermidis* can cause a mild corneal inflammation associated with contact lens wear, although secretion of the serralysin toxin may also contribute to disease severity. In many cases of contact lens related inflammation such as contact lens-associated red eye (CLARE) and contact lens peripheral ulcers (CLPU), bacteria cannot be cultured although corneal infiltrates that are probably neutrophil-rich are present. As CLARE and CLPU can be replicated in experimental models using TLR agonists such as LPS, it is likely that products of dead and degenerating bacterial activate TLRs on resident corneal cells, leading to neutrophil recruitment to the stroma.

Bacterial endophthalmitis

Bacterial endophthalmitis is characterized by a pronounced inflammatory response in the vitreous, resulting in extensive tissue damage and vision loss. Bacteria can invade the posterior segment of the eye from exogenous infection, following ocular surgery (most commonly cataract) or after traumatic or penetrating injury. Endophthalmitis can also be initiated haematogenously following infiltration of bacteria via the retinal vasculature from infection elsewhere in the body (endogenous route). Recent studies also indicate that diabetes mellitus may increase the likelihood of exogenous bacterial endophthalmitis because of increased permeability of the blood retinal barrier.

Coagulase negative staphylococci including *S. epidermidis* cause the majority of post-operative endophthalmitis, although streptococci, *S. aureus* and enterococci such as *Enterococcus faecalis* and *Propionibacterium acnes* can also cause infection. *Staphylococci, Streptococci, E. coli* and *Klebsiella pneumoniae* are among the most common causes of exogenous endophthalmitis, whereas *Bacillus*

cereus is more likely to be isolated from cases of post-traumatic endophthalmitis. Animal model studies indicate that bacterial cell wall components activate TLR on Müller cells, microglia, and macrophages in the neural retina and the inner retinal layer, in addition to hyalocytes in the vitreous. Although most infections respond to antibiotics, retinal damage and visual impairment occurs as a result of bacterial toxin production and cellular infiltration.

Trachoma

Trachoma has long been appreciated as the most prevalent microbial cause of blindness worldwide. The World Health Organization (WHO) estimates that ~1.3 million people are blind and 1.8 million have impaired vision as a result of trachoma. Trachoma is endemic in 42 countries with a population of ~200 million, although the highest incidence of disease is in sub-Saharan Africa. In Ethiopia and Sudan, the prevalence was estimated at 50% of children and 9% of adults who have active trachoma. However, in the last 10 years, control programmes have reduced the incidence of infection and the impact of this disease. The long-term goal is elimination of the disease worldwide. Fig. 8.13A shows the global distribution of trachoma in 2012 according to WHO and maps showing disease control over time can be found on the Global Trachoma Atlas website.

Trachoma is caused by *Chlamydia trachomatis*, specifically subgroups (serovars) A, B, Ba and C. These bacteria replicate in epithelial cells where they can present as reticulate bodies that do not have a cell wall. They can also form spore-like elementary bodies, which have cell walls and it is this stage that is transmitted from person-to-person by direct contact, or on the legs and bodies of flies that are attracted in large numbers to the mucous membrane of the eyes, nose and mouth, and can therefore be transmitted from infected to uninfected individuals.

Most people are infected in childhood, when they are first exposed to the bacteria. Longitudinal studies indicate that there is an incubation period following infection with *C. trachomatis*, after which bacteria are detected in the conjunctiva and can be transmitted. Disease is not necessarily dependent on the presence of live bacteria, which are cleared by the host response. However, recurrent infection of the conjunctival epithelium results in inflammation and development of characteristic papillary conjunctivitis, and over years will cause conjunctival scarring. Contraction of scar tissue causes entropion (rolling of the eyelids inward). The eyelashes then scratch the cornea (trichiasis), which can result in corneal opacification and blindness. The predominant disease manifestations are shown in Fig. 8.13B. The inflammatory response is not well understood, although animal models show an important role for IFN-γ.

The World Health Organization initiated the Global Alliance for the Elimination of Blinding Trachoma (GET2020) in 1998 with the goal of eliminating the disease by 2020. The approach is based on the SAFE strategy to control transmission, which is an acronym for Surgery for in-turned eyelids, Antibiotic use, with at least two series of treatments with oral or topical azythromycin (Zithromax, donated by Pfizer), Facial cleanliness to prevent person-to-person and fly- (*Musca sorbens*) to-person transmission and Environmental changes to increase access to clean water and encourage hand washing and the use of latrines. The latter approach is aimed at reducing the number of flies in the community, which breed on faecal material.

Trachoma in Australia. Although trachoma was long ago eliminated in the non-indigenous population in Australia, it has persisted in the indigenous population, largely due to overcrowding, poor sanitation and poor health care. In 2009, 15%–20% of children aged 1–9 years had trachoma. However, the Australian government has made a major commitment to eliminate the disease by making improvements in housing, hygiene and health care. By 2015, the number of infected children had dropped to 4.6% with only a few communities in central Australia still having high rates of the disease.

Fungal infections of the eye

Yeast and moulds can infect the cornea, causing pronounced, long-lasting infections that are very difficult to treat. *Candida albicans* is the most common yeast causing corneal infections, most often following cataract surgery or corneal and endothelial transplants. Candida is also associated with therapeutic contact lenses, steroid use, immunosuppression and corneal surgery. The yeast form of *Candida* infects the cornea, which then germinates to form pseudohyphae in the corneal stroma.

However, worldwide, moulds rather than yeast are the predominant cause of fungal keratitis, with ocular trauma being the predominant risk factor. As moulds such as *Aspergillus* and *Fusarium* species are plant saprophytes and grow abundantly on crops, the highest incidence and prevalence of fungal keratitis are in agricultural regions, especially during harvest seasons when there is an abundance of airborne spores. On a global scale, fungal keratitis accounts for ~65% of all corneal ulcers and is more common than bacteria as the cause of ocular infections. *Fusarium solani* and *F. oxysporum* are the main aetiologic agents of fungal keratitis followed by *Aspergillus species* (*A. flavus, A. fumigatus*), while *Curvularia, Alternaria and Penicillium* are less common causes.

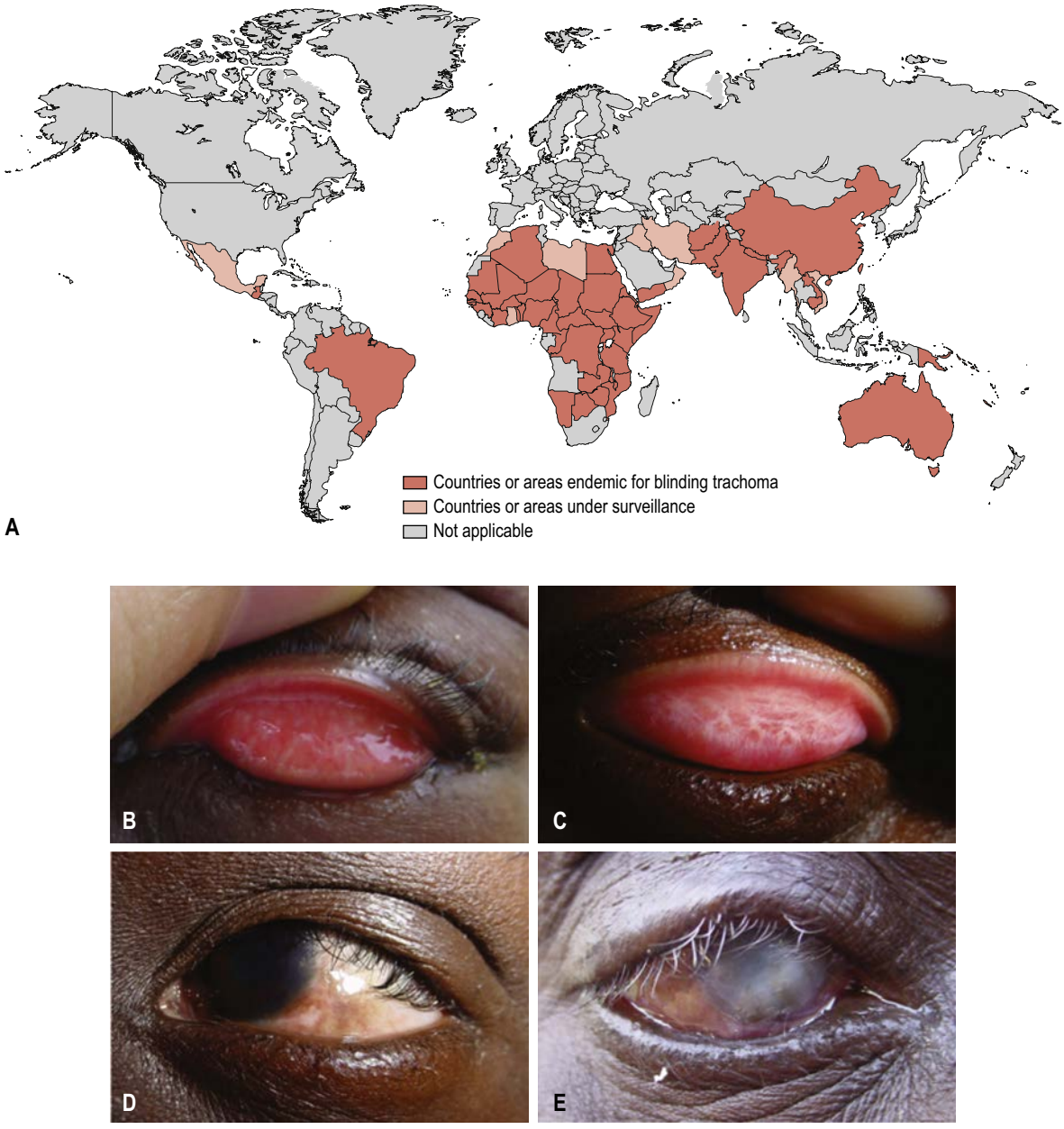


Fig. 8.13 Prevalence of trachoma and major clinical features. (**A**) Prevalence of trachoma worldwide in 2014 (*dark orange*, endemic areas; *light orange*, areas under surveillance; World Health Organization). (**B–D**) Clinical stages of trachoma. (**B, C**) Active trachoma in a child, characterized by a mixed papillary follicular response. (**D**) Tarsal conjunctival scarring. (**E**) Blinding trichiasis. (From Burton and Mabey, 2009.)

Corneal infection with *Aspergillus* and *Fusarium*: Spores or conidia are produced in large numbers on conidiophores attached by stalks to live or decaying plants (Fig. 8.14A). Conidia are dispersed by wind either alone or on plant material and can cause injury to the unprotected cornea and introduce multiple spores directly to the corneal stroma. Corneal trauma is also caused by conidia and conidiophores attached to insects

carrying plant material, tree branches and even whiplash injury from a cow tail in rural communities. Once in the corneal stroma, conidia germinate and hyphae can penetrate the corneal stroma. The resulting inflammatory response that presents as a corneal ulcer is mostly a result of neutrophils recruited to the site of infection and trauma (Fig. 8.14C,D).

The tip of the growing hyphae contains multiple proteases, including collagenases that facilitate hyphal penetration throughout the corneal stroma. If not treated early, hyphae can also penetrate Desçemet's membrane and migrate into the anterior chamber, where they stimulate a pronounced neutrophil infiltrate seen clinically as a hypopyon. Although neutrophils can limit hyphal growth, they may not prevent penetration into the posterior segment, where they can cause endophthalmitis. The resulting inflammatory response damages the retina, is difficult to treat and often requires removal of the entire globe.

Contact lens-associated fungal keratitis: In contrast to rural agricultural areas, contact lens wear is the primary risk factor in industrialized countries. The impact of this disease on the industrialized world was illustrated by several hundred cases in the USA, western Europe and Singapore during the 2005–2006 outbreak of contact lens-associated fungal keratitis. *F. solani* and *F. oxysporum* were the cause of the 2005–2006 contact lens-associated outbreak, which were resolved in large part by full thickness keratoplasty and occasionally by removal of the globe. In contrast to trauma-induced fungal keratitis in which conidia directly cause infection, contact lens-associated fungal keratitis is more likely initiated by hyphae, which form as biofilm on lenses and on lens cases. Fungal biofilms are complexes of hyphae and extracellular carbohydrates that have increased resistance to anti-microbial agents and are an important cause of infection associated with implants and catheters. Following contact with the ocular surface, hyphae in the biofilm can penetrate the cornea stroma through minor epithelial abrasions where they establish infection.

Treatment of fungal keratitis: In contrast to antibiotics used to treat bacterial keratitis, anti-mycotic agents are much less effective as hyphae can penetrate into the deeper stroma, thereby limiting accessibility. Topical natamycin and voriconazole are effective if given very early, but in general, fungal keratitis is notoriously difficult to treat, especially after the hyphae penetrate deeper stromal layers. Therefore, full thickness keratoplasty is required; however, transplant failure is not uncommon if the fungus has not been completely cleared (high risk grafts, see Chapter 4). Even in milder cases that respond to anti-fungal treatment, resolution is associated with fibrosis, resulting in corneal opacity and visual impairment.

Pathogenesis of fungal keratitis: Although *Aspergillus* and *Fusarium* produce mycotoxins that are important in plant disease, there is no direct evidence that they contribute to the pathogenesis of keratitis. Most evidence points to the role of the inflammatory response as a major factor in causing tissue damage. The host response to fungal pathogens involves recognition of cell wall components, including β-glucan and α-mannose, which bind to the c-type lectins Dectin-1 and Dectin-2 on the host cell membrane. β-glucan also activates the lectin-binding site of CR3 that is highly expressed on neutrophils). Fig. 8.14E shows β-glucan expression in an infected post-transplant cornea. C-type lectin activation leads to phosphorylation of spleen tyrosine kinase and translocation of NF-κB to the production of pro-inflammatory and chemotactic cytokines as shown above in Fig. 8.5. However, these responses also activate IFN-gamma and IL-17 producing T cells (Th1 and Th17 cells), which exacerbates the host response by activating receptors on epithelial cells and fibroblasts.

Corticosteroids: Treatment options for inflammation are limited to corticosteroids and non-steroidal anti-inflammatory agents such as cyclosporine. Steroid use is particularly risky, as premature use impairs the host response and results in enhanced fungal growth and tissue destruction. Thus, it will be important to develop more targeted and effective anti-inflammatory agents. Other approaches will be to target essential pathways of survival such as anti-oxidant and iron scavenging pathways, which are effective in animal models of disease.

Protozoan causes of ocular disease

***Acanthamoeba* keratitis**: *Acanthamoeba* keratitis is an emerging cause of corneal infections, causing severe pain and visual impairment, forming a characteristic lesion (Fig. 8.15). Infection is often associated with contact lens wear and occurs when the cornea is exposed to high concentration of free-living amoebae present in water supplies. The increased incidence of keratitis was noted in relation to changes in water treatment such as in an outbreak in Chicago in 2007. Increased incidence of *Acanthamoeba* keratitis was also associated with a contact lens care solution, although in contrast to the *Fusarium* outbreak, this was a result of contamination. *Acanthamoeba* keratitis also occurs independently of contact lens wear, especially in developing countries, where it does not present with the characteristic symptoms.

Acanthamoeba is a free living protozoan that exhibits a biphasic life cycle as a vegetative trophozoite and a physiologically inert cyst form. Trophozoites are the active stage that exist as free-living amoebae and feed on bacteria and fungi. As noted later, some species of ingested bacteria survive, notably *Legionella* species, and replicate in the trophozoites, which can behave as Trojan horses

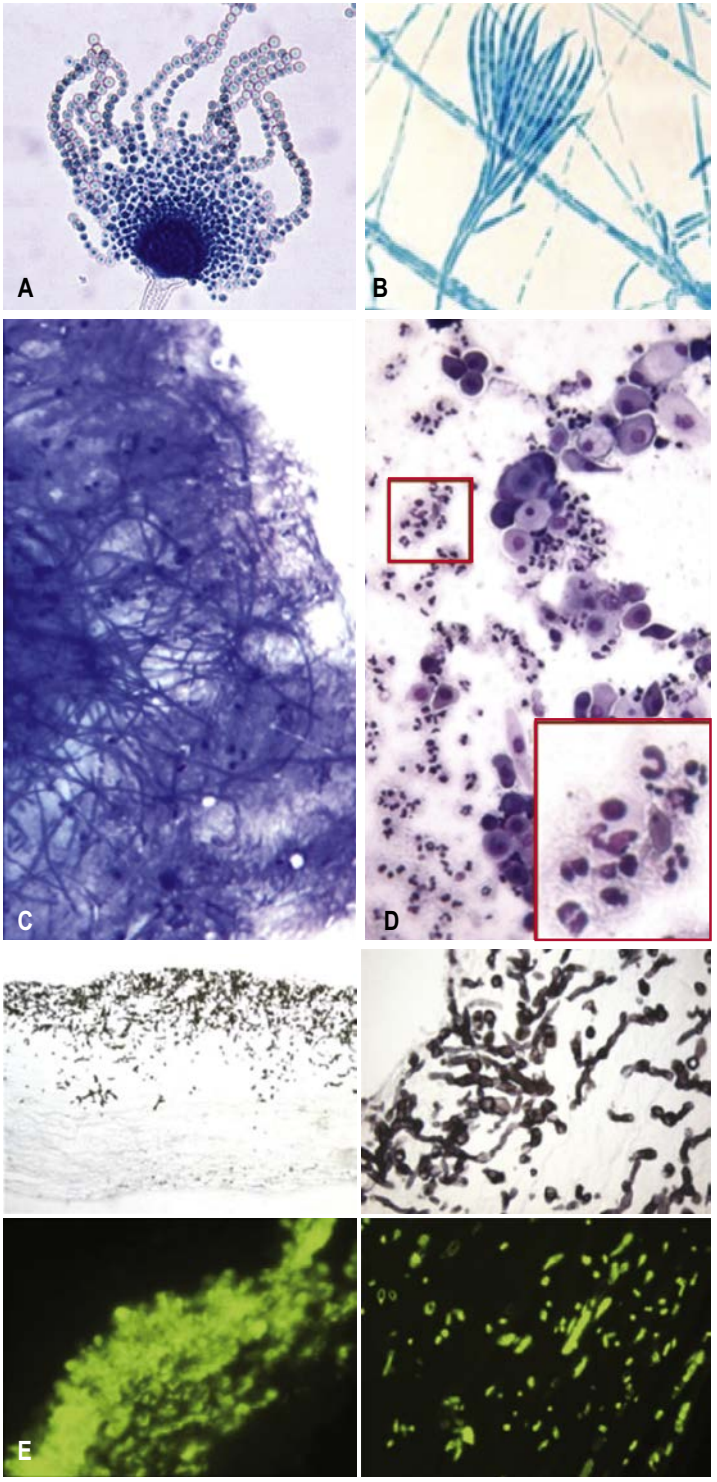


Fig. 8.14 Fungal keratitis caused by *Fusarium* and *Aspergillus* moulds. *Aspergillus flavus* (**A**) and *Fusarium solani* (**B**) conidiophores, which contain multiple conidia (spores) on their stalks, can penetrate the corneal stroma following ocular trauma with plant material. Conidia germinate and form hyphae that can be detected in corneal ulcer smears after Giemsa staining (**C**) *Aspergillus*. Neutrophils are the predominant cell type in corneal ulcers (**D**). Hyphae penetrate the corneal stroma and express the cell wall component β-glucan (**E**) upper panels: silver stain of *Aspergillus*-infected cornea obtained after transplant. Lower panels stained with antibody to β-glucan. (Parts C–E from Karthikeyan et al., 2011.)

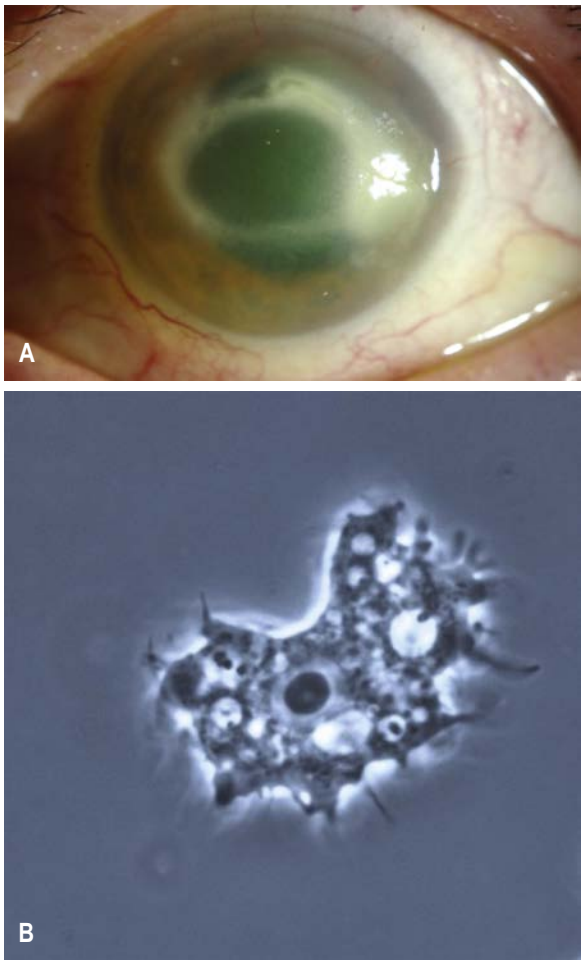


Fig. 8.15 Acanthamoeba keratitis (**A**) Characteristic radial neuropathy. (**B**) *Acanthamoeba castellani* trophozoite in culture. (Part A from Tu and Joslin, 2012.)

when they infect the cornea, releasing bacteria into the tissue. Under adverse physiological conditions, the trophozoites encyst and the cyst wall is extremely resistant to environmental agents, irradiation and the host immune system. They are therefore extremely difficult to treat when this occurs in infected corneas and require the use of highly cytotoxic drugs such as biguanides and chlorhexidine (propamidine isethionate, chlorhexidine digluconate and polyhexamethyl biguanide) given over months. Even after causing severe ocular surface damage, these drugs are not always effective and the patients still require corneal transplantation. *Acanthamoeba* keratitis is most prevalent among contact lens wearers, although it can also be introduced into the corneal stroma by ocular trauma. *Acanthamoeba* have been recovered from soil, air, chlorinated swimming pools, hot tubs, tap water and contact lens solutions. Wearing contact lenses when showering or swimming in fresh water ponds increases the risk. Although the underlying conditions for infection are not known, it probably involves trophozoites being trapped between the contact lens and the cornea.

***Acanthamoeba* virulence factors:** The first step in the pathogenesis of infection is adhesion to the ocular surface, which is mediated by carbohydrate–lectin interactions. Acanthamoeba expresses a 400-kDa transmembrane mannose-binding protein (MBP) comprised of 130kD subunits and has characteristics typical of a cell surface receptor. MBP is associated with virulence as it is expressed at higher levels by pathogenic strains and immunization of experimental animals results in protection from infection. A second virulence factor is the mannose-induced protein (MIP) 133, which is serine protease with collagenase activity and cytopathic effect on corneal epithelial cells.

Host response to *Acanthamoeba*: There are very few studies on the host response of infected individuals to this pathogen. Serum antibodies to *Acanthamoeba* antigens are commonly detected in uninfected individuals, which is consistent with environmental exposure by inhalation or ingestion. In animal models, IgA in the tears is protective and trophozoites can be killed *in vitro* by neutrophils or macrophages. Both cell types have been detected by analysis of post-transplant corneas; however, as corneas are examined only after long-term treatment, it is difficult to determine which cells are important. It seems reasonable to postulate that activation of pathogen recognition receptors stimulate infiltration of these cells to the cornea. These have not been identified but could be a result of collagen breakdown products such as Pro-Gly-Pro tri-peptides, which activate neutrophil chemokine receptors directly. In addition, release of endogenous bacteria could trigger this pathway and induce an inflammatory response. This concept does not address the chronic infection seen in this disease, which most likely relates to the inability of neutrophils and other cells to kill *Acanthamoeba* after they encyst in the cornea. The pro-inflammatory cytokine IL-17A is also important in regulating parasite growth in the cornea in a murine model.

***Acanthamoeba* as a Trojan horse:** As some bacteria are not killed by *Acanthamoeba*, trophozoites inadvertently act as a host of intracellular bacteria that are also associated with human disease. For example, *Acanthamoeba* is best characterized as a host of *Legionella sp.* (mostly *L. pneumophila*), which is associated with Legionnaire's disease; however, *Acanthamoeba* keratitis clinical isolates also harbour *Pseudomonas*, (mostly *P. aeruginosa*), *Mycobacteria* and *Chlamydia*, and in *L. pneumophila and P. aeruginosa* is dependent on expression of bacterial virulence factors.

Toxoplasmosis

Toxoplasma gondii is the most common cause of infectious retinochoroiditis worldwide, causing up to 50% of all posterior uveitis in some countries in the industrialized world, with higher levels in developing countries. *T. gondii* is an obligate intracellular protozoa and is among the most ubiquitous pathogens worldwide, with >70% of the global population seropositive. This parasite has been found in almost all mammals and can cause infections in domesticated farm animals (especially pigs) and in pets. Cats are the definitive hosts and are the only mammals that support the sexual stage of this parasite (Fig. 8.16). Transmission therefore occurs by ingesting the oocysts in food that has been contaminated with cat faeces, but individuals can also be infected following ingestion of tissue cysts (bradyzoites) in undercooked meat or in congenital toxoplasmosis by maternal transmission during pregnancy (often in the third trimester). In humans and other intermediate hosts, the tachyzoite form can invade and undergo rapid asexual replication in any nucleated cell. Following oral infection, this occurs in enterocytes, where they rapidly replicate and cause cell death. This evokes an IFN-γ -induced Th1 T-cell response; however, some parasites can cross the epithelial barrier, infect circulating monocytes and are disseminated to multiple organs, including the brain and the retina. Despite the presence of a functional innate immune response, the parasites evade the immune system in these immune-privileged sites but remain semi-dormant as viable bradyzoites within sequestered tissue cysts. However, in immune compromised individuals, the parasites can again replicate and cause multi-focal brain and retinal lesions. Although neutrophils can kill extracellular organisms, *T. gondii* can also infect neutrophils, and can regulate their ability to produce IL-1β by blocking NF-κb activation and by inhibiting caspase-1 activity.

Host response: Most of our current understanding of the control of *T. gondii* infection comes from murine models. *T. gondii* proteins activate TLR2 and TLR4, and TLR11 on monocytes and dendritic cells to produce IL-12, and natural killer (NK) cells to produce IFN-γ, which then activate $CD4^+$ and $CD8^+$ T cells. However, there is also an IFN-γ-independent mechanism of parasite killing, which is mediated by CD40–CD40L dependent autophagy.

Ocular toxoplasmosis: *T. gondii* induces an inflammatory response in the uveal tract and in the retina, causing uveitis and retinochoroiditis, although specific mediators of inflammation have yet to be identified. *In vitro*, the parasites can replicate in Müller cells, retinal epithelial cells and retinal endothelial cells. These cells then produce pro-inflammatory and chemotactic cytokines that mediate recruitment of neutrophils, macrophages and lymphocytes. Animal models of toxoplasmosis have shown that depletion of both $CD4^+$ and $CD8^+$ T-cell subsets, but not either one alone, will lead to recurrence of the disease, implying that both cell types are required. Infection of the retina and choroid, i.e. at the blood–retina barrier (BRB), may occur by several routes, one well described route being the Trojan horse where toxoplasma-infected monocytes, when activated, are able to cross the BRB and become latent within the retinal tissue (see earlier comment about latency in viruses). In immunocompetent individuals, *Toxoplasma* uveitis is self-limiting; however, recurrent lesions are associated with development of retinochoroiditis and loss of vision. It is likely that reactivation of latent tissue cysts is the basis for recurrence, with the stimulus probably related to the host response.

Helminths can cause ocular infection

Helminths are a rare but well recognized cause of ocular disease in the industrialized world, the most common being toxocariasis. Other less well recognized, or unidentified, helminths may cause an unusual retinal infection known as diffuse unilateral subacute neuroretinitis in which the organism migrates around the fundus of the eye in the subretinal space, causing a pigmentary retinopathy and low-grade retinal inflammation, with progressive loss of visual field and eventually visual acuity.

Toxocariasis

Toxocara canis and *Toxocara cati* are nematodes transmitted by dogs and cats, respectively, where the adult worms are found in the intestines. Eggs are secreted in the faeces and are ingested by the secondary hosts, which include pigs and rodents. Larvae migrate through the intestine and can be carried to any part of the body via the blood stream. The cycle is completed when the dog or cat eats the secondary host. Human infection also occurs following ingestion of contaminated food, although humans are a dead-end in terms of the worm's life cycle. However, the larvae still migrate through the intestine and are disseminated, which is termed visceral larva migrans.

Ocular infections are among the most common manifestations, mostly causing retinitis that appears as a large inferior white intraocular vitreoretinal mass, although the live worm can sometimes be lodged in the anterior chamber. Treatment of ocular disease is with oral thiabendazole or albendazole, which is better tolerated.

Onchocerciasis (river blindness)

The filarial nematode *Onchocerca volvulus* is the causative organism of Onchocerciasis, also known as river blindness,

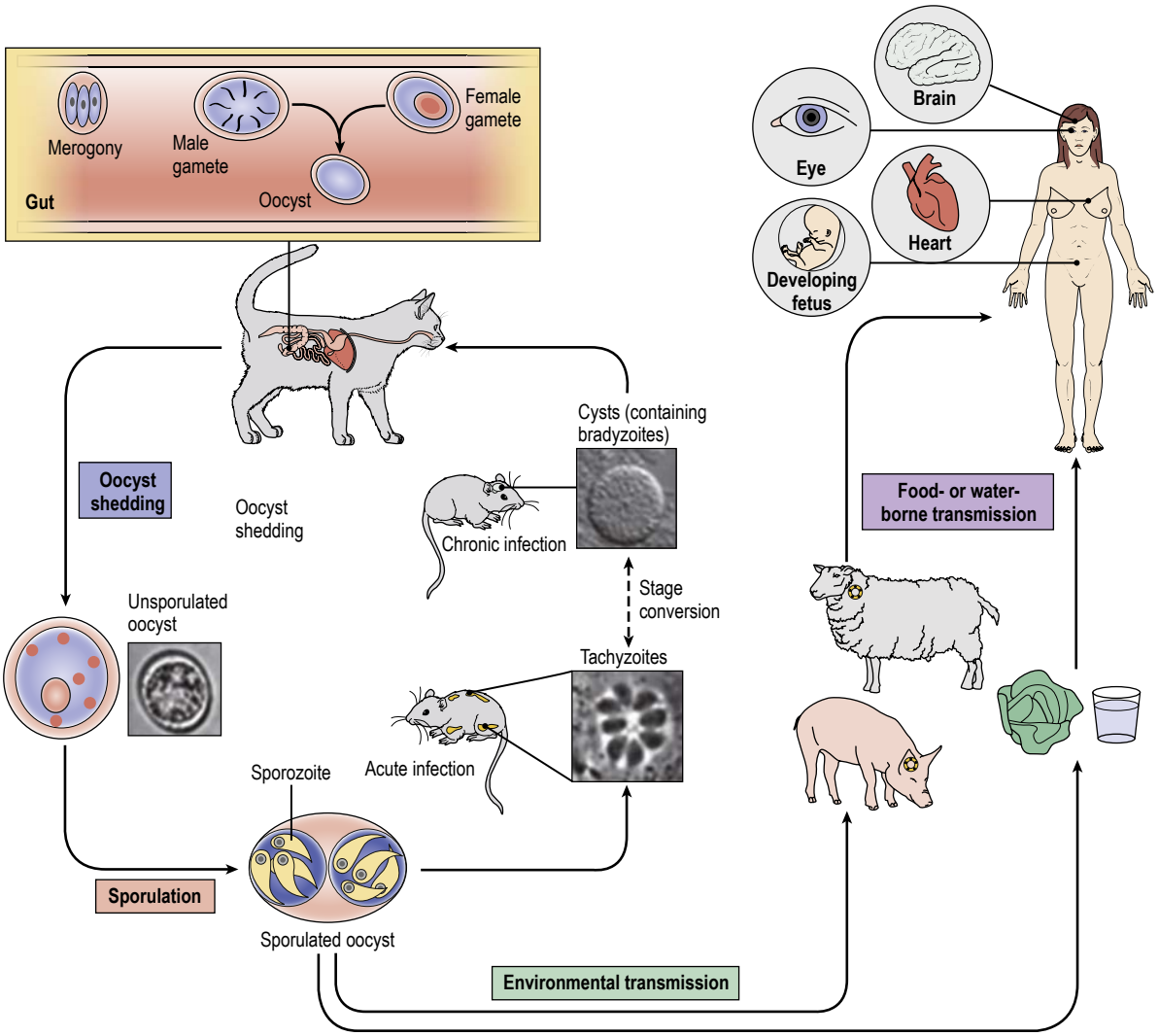


Fig. 8.16 The complex life cycle of *Toxoplasma gondii*. Humans are accidental hosts and represent a dead end to continuing the life cycle. Cats are the only definitive host permissive for sexual replication in which male and female gametes are formed in the cat enterocytes. *T. gondii* gametes form oocytes, which survive in the environment and are ingested by intermediate hosts, including rodents, sheep, pigs and humans. Fast-replicating tachyzoites and slow-replicating bradyzoites are generated in epithelial cells, and can cause inflammation and tissue damage, including in the brain and the eye. (From Hunter and Sibley, 2012.)

which has caused devastating blindness throughout sub-Saharan Africa, central and South America, and in the Arabian Peninsula. Until control efforts were initiated by the World Health Organization in the 1970s, the physical, social and economic impact of blindness and visual impairment caused by this infection resulted in villages adjacent to rivers, with the most fertile soil, being abandoned. The blindness develops as a result of anterior and posterior segment ocular inflammation, including sclerosing keratitis, chorioretinitis and optic neuritis.

Life cycle and parasite burden: The 'river' connection relates to the *Simulium* blackfly vector that breeds in fast-flowing rivers. First-stage larvae (microfilariae, L1) in infected individuals are ingested during the female blackfly's blood meal and migrate through the gut and thorax to the salivary gland having undergone two moults to the third-stage larvae (L3). During the second blood meal, the L3-stage larvae enter the bloodstream, develop into stage four larvae (L4) and then into adult males and females that can live for over 10 years. Adult worms are sequestered

in collagenous nodules and females produce millions of microfilariae during their lifetime and can survive in the skin for up to 30 months; individuals can be very heavily infected. Microfilariae are ~300 μm long, exhibit rapid motility and are readily detected in the anterior chamber by slit lamp microscopy. Fundus examination can also detect microfilariae in the retina.

***Pathogenesis*:** Given the large parasite burden of infected individuals, a strong immune response might be expected; however, live microfilariae appear to cause minimal tissue damage either to the skin or the eyes. As with other chronic parasitic infections in which the host and parasite have co-evolved, the infected host generates a predominantly immunosuppressive response that involves T-regulatory cells, IL-10 and TGF-β, which facilitates the longevity of adult and larval worms and continuation of the life cycle. In marked contrast to the suppressive response to live *O. volvulus*, when the larvae die either by natural attrition or following chemotherapy, infected individuals experience tissue damage, which manifests as onchodermatitis and stromal keratitis (Fig. 8.17) Epidemiological and experimental data indicate that almost all immunopathology associated with onchocerciasis is directly or indirectly related to the local death of microfilariae.

Endosymbiotic Wolbachia bacteria: Although immunopathology has long been associated with dead and degenerating microfilariae, the finding that *O. volvulus* and other filariae harbour intracellular *Rickettsia*-like bacteria led to an increased understanding of the disease. *Wolbachia* are ubiquitous in the ovaries of insects and other arthropods, where they influence sex ratio and are transmitted transovarially. Filariae have obligate developmental stages in insects (*O. volvulus* in blackflies, lymphatic filariae in mosquitoes), which probably explains why they are the only group of nematodes that harbour *Wolbachia*. Using antibodies to the *Wolbachia* surface protein (WSP), *Wolbachia* were found to be abundant in the nematode hypodermis in the gravid uterus (Fig. 8.18) and in microfilariae. Thus, larvae migrating through the skin and the eye also harbour bacteria and, when they die, bacteria are released into the tissues. Several lines of evidence support the notion that *Wolbachia* is responsible for much of the immunopathology associated with ocular onchocerciasis: first, *Wolbachia* organisms and DNA are found in the blood of filarial infected individuals following chemotherapy; second, *O. volvulus* from individuals depleted of *Wolbachia* by antibiotic treatment do not induce corneal inflammation; third isolated *Wolbachia* induce corneal inflammation through TLR2 activation, which is the same as soluble extracts from the whole worms. *Wolbachia* are also detected in microfilariae in the skin of infected individuals. Analysis of the *Wolbachia* genome showed they do not have the LPS synthase gene and therefore cannot produce LPS, but they can produce lipopeptides. A synthetic *Wolbachia* lipopeptide was able to reproduce these findings, showing corneal inflammation that is dependent on TLR2 and TLR6.

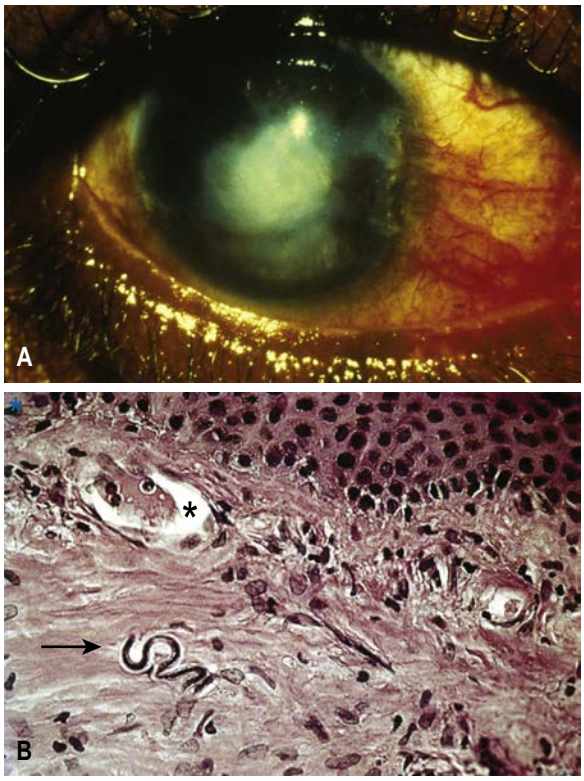


Fig. 8.17 (**A**) Late stages of ocular onchocerciasis (river blindness) caused by the parasitic nematode *Onchocerca volvulus*. (**B**) A single microfilaria *(arrow)* in an area of the corneal stroma with no inflammation, whereas cellular infiltrates and corneal blood vessels *(asterisk)* are present at other sites. (From Pearlman, 2012.)

Onchocerciasis control: In the 1970s, the World Health Organization launched the Onchocerciasis Control Program (OCP), based on aerial spraying of insecticide together with chemotherapy. Although early drugs including diethylcarbamazine had serious side effects, ivermectin (Mectizan) was found to kill the microfilariae (although not the adults) and was effective if given annually. After several treatments, transmission was significantly reduced in the initial target area in west Africa and Ivermectin was provided free by Merck, Inc. This programme was expanded as the African Program for Onchocerciasis Control (APOC) and a similar ivermectin distribution programme was initiated in the Americas and in the Arabian peninsula. These programmes have

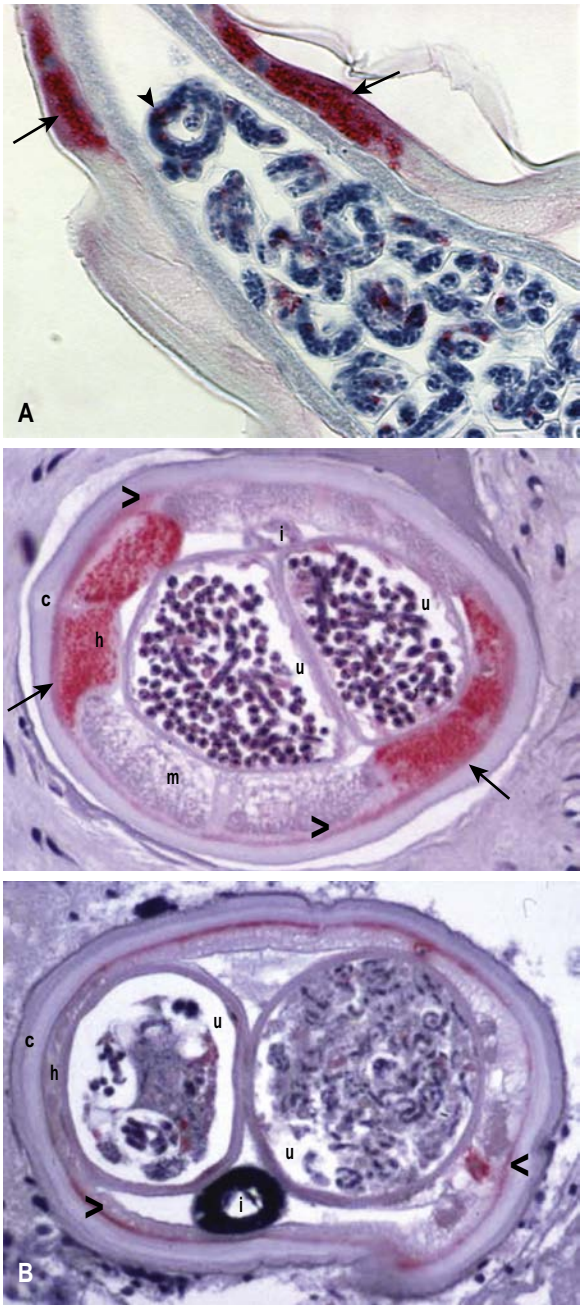


Fig. 8.18 (**A**) Endosymbiotic *Wolbachia* bacteria in the hypodermis and in the developing microfilariae in the uterus of an adult female filarial worm (*Brugia malayi;* longitudinal sections incubated with antibody to the *Wolbachia* surface protein and identified using alkaline phosphatase (red), indicated by the *arrows* for the hypodermis, and the *arrowhead* for microfilariae). (**B**) Cross-section of adult female worms following treatment with ivermectin (IVR) alone showing *Wolbachia (upper panel)* or with IVR and doxycycline that kills *Wolbachia (lower panel)*. (Part A Photomicrograph by A. Hise, from Pennisi, 2002. Part B from Hoerauf, 2001.) c: cuticle; h: hypodermis; u: uterus; m: muscle; i: interstitium.

met with considerable success and improved the lives of millions of individuals. Remaining challenges are to maintain the reduced transmission, especially in areas of civil strife, and to examine the effectiveness of additional therapeutic such as albendazole and antibiotics that target *Wolbachia*.

In targeting *Wolbachia*, studies showed that doxycycline and related antibiotics reduced not only the microfilarial load, but also the adult worm burden, effectively curing rather than only treating the infection. The caveat has been that in contrast to IVR that is given annually, doxycycline requires multiple doses over several days. The challenge remains to discover more potent antibiotics that can be administered *en masse*. The Bill and Melinda Gates Foundation has invested in this approach to develop new antibiotics as an academia/industry Anti-*Wolbachia* Consortium (A-Wol.org). This group reported in 2018 that a new antibiotic, AWZ1066, rapidly and selectively kills *Wolbachia* in murine models of filariasis, reducing the number of adults and microfilariae. They also showed a role for the macrolide veterinary antibiotic, tylosin A. If clinical trials are successful, these drugs will have a major effect on the burden of onchocerciasis.

Loa loa (loiasis, eye worm)

Loa loa is a filarial nematode with a similar life cycle as *O. volvulus*, except the adults migrate through subcutaneous tissues and the microfilariae are in the blood. The intermediate vectors are large Chrysops flies rather than blackflies and transmission is limited to regions of western Africa. Disease manifestations are caused by the migrating adults that can live for up to 17 years in a human host, and migrate through subcutaneous tissues at an estimated 1 cm per minute. Adult worms also migrate through the conjunctiva (not the cornea), where they can be surgically removed (Fig. 8.19). Localized inflammation known as Calabar swellings may occur in subcutaneous tissues of the skin, but they do not cause major tissue damage. *Loa loa* does not harbour *Wolbachia* and infection is treated with corticosteroids and diethylcarbamazine.

Conclusions

In summary, microbial infections of the eye are a major cause of blindness and visual impairment worldwide. Infections can affect the anterior and posterior segments and disease pathogenesis is the net result of both microbial virulence and the host response. Different pathogens have evolved multiple strategies for survival in the eye and the host is not always discriminating in its ability to kill the microbes without causing tissue damage. However, the ocular surface barriers perform an outstanding function in

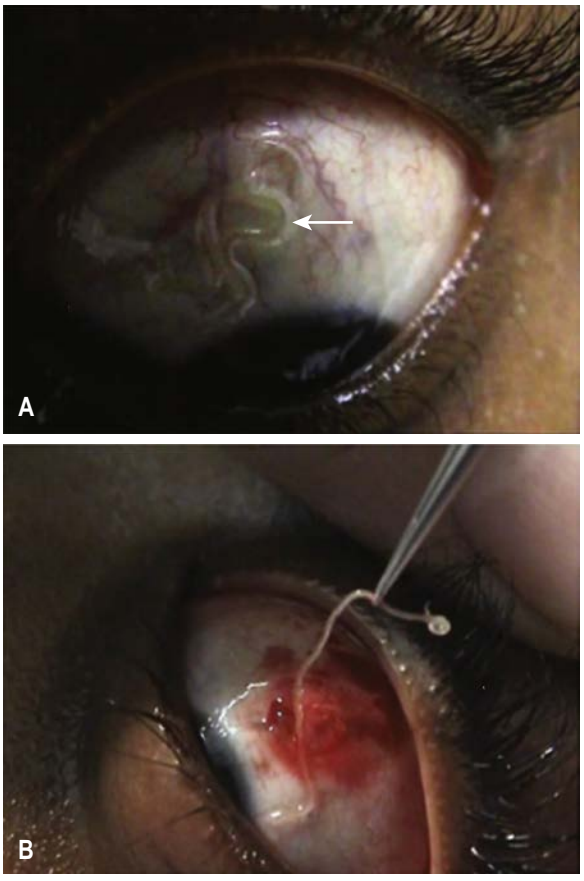


Fig. 8.19 Loiasis (eyeworm). Adult *Loa loa* migrating through the conjunctiva (**A**), where it was then surgically removed (**B**). (From Shah and Saldana, 2010.)

protecting us from most environmental microbes under normal physiological conditions and infection occurs primarily when the ocular surface is compromised by trauma or extended contact lens wear.

Future treatment for many of these blinding eye infections is primarily dependent on the basic research required to characterize the pathogenesis of disease and to identify specific targets for immune intervention beyond the use of corticosteroids that have numerous unwanted side-effects. These could include blocking pathogen recognition molecules or neutrophil chemokines when given together with effective antibiotics. However, many of these microbes, bacteria in particular, have become resistant to antibiotics and there is a pressing need to develop new classes of anti-microbial agents. To this end, anti-microbial peptides have some potential as they are biological rather than pharmacologic agents, their specificity is well understood and they can be readily synthesized as a native sequence or with substituted amino acids that may improve efficacy. Targeting microbial scavenging for essential metals such as iron and zinc may also prove to be a feasible approach.

eLink 8.4 A full reading list is available online here https//:expertconsult.inkling.com

References

Cheung, N., Nagra, P., Hammersmith, K., 2016. Emerging trends in contact lens-related infections. Curr. Opin. Ophthalmol. 27, 327–332. https://doi.org/10.1097/ICU.0000000000000280.

CDC, 2017. Centers for Disease Control and Prevention. National Center for Emerging and Zoonotic Infectious Diseases (NCEZID), Division of Vector-Borne Diseases (DVBD).

de Paula Freitas Jr., B., de Oliveira Dias, J., Prazeres, J., Sacramento, G.A., Ko, A.I., Maia, M., et al., 2016. Ocular findings in infants with microcephaly associated with presumed zika virus congenital infection in salvador, Brazil. JAMA Ophthalmol. 134, 529–535.

Forrester, J.V., 2018. Ebola virus and persistent chronic infection: when does replication cease?. Ann. Transl. Med. 6, S39. https://doi.org/10.21037/atm.2018.09.60.

Gorshkov, K., Shiryaev, S.A., Fertel, S., Lin, Y.W., Huang, C.T., Pinto, A., et al., 2018. Zika virus: origins, pathological action, and treatment strategies. Front. Microbiol. 9, 3252.

Hong, W.D., Benayoud, F., Nixon, G.L., Ford, L., Johnston, K.L., Clare, R.H., et al., 2018. AWZ1066S, a highly specific anti-Wolbachia drug candidate for a short-course treatment of filariasis. PNAS 116, 14.

Hoerauf, A., Mand, S., Adjei, O., Fleischer, B., Büttner, D.W., 2001. Depletion of wolbachia endobacteria in Onchocerca volvulus by doxycycline and microfilaridermia after ivermectin treatment. Lancet 357 (9266), 1415-6 14-1419.

Oliver, G.F., Carr, J.M., Smith, J.R., 2018. Emerging infectious uveitis: Chikungunya, dengue, Zika and Ebola-A review. Clin. Exp. Ophthalmol.

Pearlman, E., Leal Jr., SM., 2012. Review. Cytokine. 58 (1):107-111 460.

Pennisi, E. Worms May Not Act Alone in River Blindness. Science, Mar. 8, 2002 https://www.sciencemag.org/news/2002/03/worms-may-not-act-alone-river-blindness.

Pierson, T.C., Diamond, M.S., 2018. The emergence of Zika virus and its new clinical syndromes. Nature 560, 573–581.

Schmitt, H., Sell, S., Koch, J., Seefried, M., Sonnewald, S., Daniel, C., et al., 2016. Siglec-H protects from virus-triggered severe systemic autoimmunity. J. Exp. Med. 213, 1627–1644.

St Leger, A.J., et al., 2017. An ocular commensal protects against corneal infection by driving an interleukin-17 response from mucosal gammadelta T cells. Immunity 47, 148–158 e145. https://doi.org/10.1016/j.immuni.2017.06.014.

Ung, L., Bispo, P., Shanbhag, S., Gilmore, M., Chodosh, J., 2019. The persistent dilemma of microbial keratitis: global burden, diagnosis, and antimicrobial resistance. Surv. Ophthalmol. 64, 255–271.

Voigt, V., Wikstrom, M.E., Kezic, J.M., Schuster, I.S., Fleming, P., Makinen, K., et al., 2017. Ocular antigen does not cause disease unless presented in the context of inflammation. Sci. Rep. 7, 14226.

Voigt, V., et al., 2018. Cytomegalovirus establishes a latent reservoir and triggers long-lasting inflammation in the eye. PLoS Pathog. 14, e1007040. https://doi.org/10.1371/journal.ppat.1007040.

Further reading

Alizadeh, H., Apte, S., El-Agha, M.S., Li, L., Hurt, M., Howard, K., et al., 2001. Tear IgA and serum IgG antibodies against Acanthamoeba in patients with Acanthamoeba keratitis. Cornea 20, 622–627.

Ansari, Z., Miller, D., Galor, A., 2013. Current thoughts in fungal keratitis: diagnosis and treatment. Curr. Fungal. Infect. Rep 7, 209–218.

Bhosai, S.J., Bailey, R.L., Gaynor, B.D., Lietman, T.M., 2012. Trachoma: an update on prevention, diagnosis, and treatment. Curr. Opin. Ophthalmol 23, 288–295.

Borkar, D.S., Fleiszig, S.M., Leong, C., Lalitha, P., Srinivasan, M., Ghanekar, A.A., et al., 2013. Association between cytotoxic and invasive Pseudomonas aeruginosa and clinical outcomes in bacterial keratitis. JAMA. Ophthalmol 131, 147–153.

Bowler, G.S., Shah, A.N., Bye, L.A., Saldana, M., 2011. Ocular loiasis in London 2008–2009: a case series. Eye 25, 389–391.

Burton, M.J., Mabey, D.C., 2009. The global burden of trachoma: a review. PLoS Negl. Trop. Dis. 3, e460.

Carnt, N., Robaei, D., Minassian, D.C., Dart, J.K.G., 2018. Acanthamoeba keratitis in 194 patients: risk factors for bad outcomes and severe inflammatory complications. Br. J. Ophthalmol 102, 1431–1435.

Carnt, N., Hoffman, J.M., Verma, S., Hau, S., Radford, C.F., Minassian, D.C., et al., 2018. Acanthamoeba keratitis: confirmation of the UK outbreak and a prospective case-control study identifying contributing risk factors. Br. J. Ophthalmol 102, 1621–1628.

Chinnery, H.R., Pearlman, E., McMenamin, P.G., 2008. Cutting edge: Membrane nanotubes in vivo: a feature of MHC class II+ cells in the mouse cornea. J. Immunol 180, 5779–5783.

de Paula Freitas, B., de Oliveira Dias, J.R., Prazeres, J., Sacramento, G.A., Ko, A.I., Maia, M., et al., 2016. Ocular findings in infants with microcephaly associated with presumed Zika virus congenital infection in Salvador, Brazil. JAMA. Ophthalmol 134, 529–535.

Franchi, L., Munoz-Planillo, R., Nunez, G., 2012. Sensing and reacting to microbes through the inflammasomes. Nat. Immunol. 13, 325–332.

Garg, P., Kalra, P., Joseph, J., 2017. Non-contact lens related Acanthamoeba keratitis. Indian. J. Ophthalmol 65, 1079–1086.

Gipson, I.K., 2004. Distribution of mucins at the ocular surface. Exp. Eye Res. 78, 379–388.

Gorshkov, K., Shiryaev, S.A., Fertel, S., Lin, Y.W., Huang, C.T., Pinto, A., et al., 2018. Zika Virus: origins, pathological action, and treatment strategies. Front. Microbiol 9, 3252.

Hardison, S.E., Brown, G.D., 2012. C-type lectin receptors orchestrate antifungal immunity. Nat. Immunol 13, 817–822.

Hauser, A.R., 2009. The type III secretion system of Pseudomonas aeruginosa: infection by injection. Nat. Rev. Microbiol 7, 654–665.

Hoerauf, A., Mand, S., Adjei, O., Fleischer, B., Buttner, D.W., 2001. Depletion of wolbachia endobacteria in Onchocerca volvulus by doxycycline and microfilaridermia after ivermectin treatment. Lancet 357, 1415–1416.

Hong, W.D., Benayoud, F., Nixon, G.L., Ford, L., Johnston, K.L., Clare, R.H., et al., 2019. AWZ1066S, a highly specific anti-Wolbachia drug candidate for a short-course treatment of filariasis. Proc. Natl. Acad. Sci. U. S. A. 116, 1414–1419. https://wwwnc.cdc.gov/travel/page/zika-information.

Hunter, C.A., Sibley, L.D., 2012. Modulation of innate immunity by Toxoplasma gondii virulence effectors. Nat. Rev. Microbiol 10, 766–778.

Iovieno, A., Ledee, D.R., Miller, D., Alfonso, E.C., 2010. Detection of bacterial endosymbionts in clinical acanthamoeba isolates. Ophthalmology 117, 445–452 452 e441–443.

Jabs, D.A., Ahuja, A., Van Natta, M., Lyon, A., Srivastava, S., Gangaputra, S., 2010. Course of cytomegalovirus retinitis in the era of highly active antiretroviral therapy: five-year outcomes. Ophthalmology 117, 2152–2161 e2151–2152.

Joslin, C.E., Tu, E.Y., McMahon, T.T., Passaro, D.J., Stayner, L.T., Sugar, J., 2006. Epidemiological characteristics of a Chicago-area Acanthamoeba keratitis outbreak. Am. J. Ophthalmol 142, 212–217.

Karthikeyan, R.S., Priya, J.L., Leal Jr., S.M., Toska, J., Rietsch, A., Prajna, V., et al., 2013. Host Response and bacterial virulence factor expression in pseudomonas aeruginosa and streptococcus pneumoniae corneal ulcers. PLoS. One 8, e64867.

Karthikeyan, R.S., Leal Jr., S.M., Prajna, N.V., Dharmalingam, K., Geiser, D.M., Pearlman, E., et al., 2011. Expression of innate and adaptive immune mediators in human corneal tissue infected with Aspergillus or fusarium. J. Infect. Dis 204, 942–950.

Knickelbein, J.E., Watkins, S.C., McMenamin, P.G., Hendricks, R.L., 2009. Stratification of Antigen-presenting Cells within the Normal Cornea. Ophthalmol Eye Dis 1, 45–54.

Kolaczkowska, E., Kubes, P., 2013. Neutrophil recruitment and function in health and inflammation. Nat. Rev. Immunol. 13, 159–175.

Leal Jr., S.M., Pearlman, E., 2012. The role of cytokines and pathogen recognition molecules in fungal keratitis - Insights from human disease and animal models. Cytokine 58, 107–111.

Lee, E.J., Truong, T.N., Mendoza, M.N., Fleiszig, S.M., 2003. A comparison of invasive and cytotoxic Pseudomonas aeruginosa strain-induced corneal disease responses to therapeutics. Curr. Eye. Res 27, 289–299.

Lietman, T.M., Pinsent, A., Liu, F., Deiner, M., Hollingsworth, T.D., Porco, T.C., 2018. Models of Trachoma Transmission and Their Policy Implications: From Control to Elimination. Clin. Infect. Dis 66, S275–S280.

Lima, T.S., Lodoen, M.B., 2019. Mechanisms of Human Innate Immune Evasion by Toxoplasma gondii. Front. Cell. Infect. Microbiol 9, 103.

Lima, T.S., Gov, L., Lodoen, M.B., 2018. Evasion of Human Neutrophil-Mediated Host Defense during Toxoplasma gondii Infection. mBio 9.

Miller, F.C., Coburn, P.S., Huzzatul, M.M., LaGrow, A.L., Livingston, E., Callegan, M.C., 2019. Targets of immunomodulation in bacterial endophthalmitis. Prog. Retin. Eye. Res 73, 100763.

Neelam, S., Niederkorn, J.Y., 2017. Pathobiology and Immunobiology of Acanthamoeba Keratitis: Insights from Animal Models. Yale. J. Biol. Med 90, 261–268.

Oliver, G.F., Carr, J.M., Smith, J.R., 2018. Emerging infectious

uveitis: Chikungunya, dengue, Zika and Ebola-A review. Clin. Exp. Ophthalmol.
O'Neill, L.A., Golenbock, D., Bowie, A.G., 2013. The history of Toll-like receptors – redefining innate immunity. Nat. Rev. Immunol. 13, 453.
Pearlman, E., Johnson, A., Adhikary, G., Sun, Y., Chinnery, H.R., Fox, T., et al., 2008. Toll-like receptors at the ocular surface. Ocul. Surf 6, 108–116.
Pearlman, E., Sun, Y., Roy, S., Karmakar, M., Hise, A.G., Szczotka-Flynn, L., et al., 2013. Host defense at the ocular surface. Int. Rev. Immunol. 32, 4–18.
Pierson, T.C., Diamond, M.S., 2018. The emergence of Zika virus and its new clinical syndromes. Nature 560, 573–581.
Saint Andre, A., Blackwell, N.M., Hall, L.R., Hoerauf, A., Brattig, N.W., Volkmann, L., et al., 2002. The role of endosymbiotic *Wolbachia* bacteria in the pathogenesis of river blindness. Science 295, 1892–1895.
Shah, A.N., Saldana, M., 2010. Images in clinical medicine. Ocular loiasis. N. Engl. J. Med. 363, e16.
Sharma, S., Garg, P., Rao, G.N., 2000. Patient characteristics, diagnosis, and treatment of non-contact lens related Acanthamoeba keratitis. Br. J. Ophthalmol 84, 1103–1108.
Subauste, C.S., 2019. Interplay Between Toxoplasma gondii, Autophagy, and Autophagy Proteins. Front. Cell. Infect. Microbiol 9, 139.
Sun, Y., Karmakar, M., Taylor, P.R., Rietsch, A., Pearlman, E., 2012. ExoS and ExoT ADP ribosyltransferase activities mediate Pseudomonas aeruginosa keratitis by promoting neutrophil apoptosis and bacterial survival. J. Immunol 188, 1884–1895.
Sun, Y., Karmakar, M., Roy, S., et al., 2010. TLR4 and TLR5 on corneal macrophages regulate Pseudomonas aeruginosa keratitis by signaling through MyD88-dependent and -independent pathways. J. Immunol 185, 4272–4283.
Suryawanshi, A., Cao, Z., Sampson, J.F., Panjwani, N., 2015. IL-17A-mediated protection against Acanthamoeba keratitis. J. Immunol 194, 650–663.
Szliter, E.A., Lighvani, S., Barrett, R.P., Hazlett, L.D., 2007. Vasoactive intestinal peptide balances pro- and anti-inflammatory cytokines in the Pseudomonas aeruginosa-infected cornea and protects against corneal perforation. J. Immunol 178, 1105–1114.
Tu, E.Y., Joslin, C.E., 2012. *Microsporidia* and *Acanthamoeba*: the role of emerging corneal pathogens. Eye (Lond.) 26, 222–227.
Wuest, T.R., Carr, D.J., 2010. VEGF-A expression by HSV-1-infected cells drives corneal lymphangiogenesis. J. Exp. Med. 207, 101–115.

Chapter 9

Pathology

Contents

Introduction
Cell and tissue damage
- Agents or conditions that can cause direct damage to ocular cells and tissues
- Mechanism of cell death

Healing and repair
- Cornea
- Iris
- Lens
- Retina
- Choroid
- Sclera
- Optic nerve

Inflammation
- What happens in acute inflammation?
- How does infection cause inflammation?
- Inflammation may occur without infection
- Granulomatous inflammation
- Non-granulomatous inflammation
- Inflammation due to autoimmune disease

Metabolic disease
- Diabetes
- Thyroid eye disease
- Disorders of amino acid metabolism

Ageing, degeneration and dystrophies
- Ageing
- Degeneration
- Dystrophies

Neoplasia
- Pathogenesis of neoplasia
- Hamartomas
- Choristomas
- Teratoma
- Benign epithelial tumours
- Malignant epithelial tumours
- Malignant melanoma
- Tumours of peripheral nerves
- Tumours derived from muscle
- Vascular tumours
- Other connective tissue tumours
- Lymphoid tumours
- Lacrimal gland tumours
- Metastatic tumours
- Disorders misdiagnosed as neoplasms

Introduction

A knowledge of pathology is the key to good clinical practice and the deeper this knowledge, the better the practice. Less well recognized is that a knowledge of pathology also underpins much of the work of the ophthalmic and vision scientist and offers wide scope for research ideas. Whilst the study of tissues and cells from pathological processes remains the mainstay of ophthalmic pathology, there have been recent advances in less invasive liquid biopsy particularly in tumour pathology to detect traces of cancer DNA in blood or other fluids. Examples of this in eye pathology include the detection and characterization of circulating melanoma cells (CMCs) which can aid in the diagnosis, prognosis and disease monitoring of uveal melanoma patients. Previously discarded aqueous humour can be used to detect chromosomal changes in retinoblastoma. This chapter describes basic aspects of disease processes with reference to specific entities relevant in ophthalmology.

Cell and tissue damage

Cells and tissues die as part of normal physiological homeostasis and different tissues have different susceptibility to death.

Indeed, different processes occur as cells die and are ultimately cleared away. Similarly, tissues have different capability for regeneration and this applies to all ocular tissues.

Agents or conditions that can cause direct damage to ocular cells and tissues

There are a wide variety of noxious agents and processes or deficiencies capable of causing damage to the eye and some of these are described in this chapter.

Reduced oxygen supply

Central retinal artery occlusion causes reduced oxygen flow to the retina, which becomes opaque due to oedema.

Physical agents

- *Mechanical trauma* is the commonest form of ocular injury. Blunt trauma describes mechanical injury in which the globe remains intact. Anteroposterior deformation separates the delicate attachments between the intraocular structures, e.g. the zonular fibres, rupture of which causes lens dislocation. Separation of the attachment of the ciliary muscle to the scleral spur leads to collapse of the trabecular meshwork and secondary glaucoma (angle recession glaucoma). Penetrating and perforating injuries are caused by a wide variety of weapons, tools, sporting equipment and domestic utensils and almost inevitably lead to severe disorganization within the ocular structures.
- *Extremes of temperature* such as cryotherapy are used surgically to induce an adhesive scar by freezing and thawing in the peripheral retina during the prophylaxis or treatment of retinal detachment. Cryoablation is also used to destroy the ciliary body in cases of intractable glaucoma, the aim being to suppress aqueous formation.
- *Light exposure:* excess exposure to ultraviolet light may lead to an overproduction of free radicals by photons (e.g. hydrogen peroxide, superoxides) with insufficient free radical scavengers (e.g. vitamin A, superoxide dismutase and glutathione transferase) with resultant tissue damage. For example, damage may occur to the corneal epithelium in snow blindness. Lasers used in ophthalmology cause tissue damage dependent on the wavelength, the site of absorption and the quantity of energy released. In panretinal photocoagulation, which is used to treat diabetic vasoproliferative retinopathy, individual burns destroy localized patches of the outer third of the retina and the underlying retinal pigment epithelium (RPE) cells. Reactive proliferation of RPE at the edge of the burn leads to clinically visible pigmentation around a white circle, which is the result of a 'scar' formed by glial cells. Transpupillary thermotherapy (TTT) uses infrared light to heat (to 40°C) choroidal melanomas and bring about tumour cell necrosis. Several types of laser are used in ophthalmology to treat various eye diseases and newer lasers such as the diode laser and the nanolaser are continuing to be developed (see Table 9.1). However, there is little evidence that these are superior to the widely used, robust argon laser (Moutray, 2018)

Table 9.1 **Lasers used in ophthalmology**

Type of Laser	Wavelength (nm)	Mechanism of Action	Application in Ophthalmology
Argon	488-514	Photocoagulation	Coagulation from choriocapillaris to inner nuclear layer
Frequency-doubled Nd-YAG	532	Photocoagulation	Safer than argon laser for macular treatment
Diode Laser	810	Photocoagulation	ROP; retinal photocoagulation; destruction of ciliary body and TPT of choroidal lesions
Photodynamic therapy	689	Photoradiation	Choroidal vascular pathologies; choroidal haemangiomas; leaking choroidal naevi; choroidal melanomas
Nd-YAG	1064	Photodisruption	Posterior capsulotomy in posterior capsular opacification; Peripheral iridotomy in glaucoma
Excimer	193	Photoablation	Refractive surgery (PRK, LASIK); Phototherapeutic keratectomy
Femtosecond Laser	1053	Photodisruption	Refractive surgery (LASIK flap creation; femtosecond-assisted cataract surgery)

Nd-YAG, Neodynium yitrium aluminium garnate; *TPT*, transpupillary thermotherapy.

- *Ionizing radiation*: each of the three types of ionizing radiation in common use in medical practice has applications in ophthalmology:
  - charged particles (electrons and α-particles)
  - uncharged particles (neutrons)
  - electromagnetic radiation (X-rays and γ-rays).

Proton beam therapy is now an accepted choice for the external treatment of ocular melanomas, while γ-emitters such as ^{106}Ru and ^{60}Co can be applied topically to the sclera over a melanoma in the form of a radioactive plaque. X-rays and γ-rays are used for the external beam treatment of retinoblastoma.

The unit (gray; Gy) is a measure of the amount of energy absorbed in the target tissue. For melanomas, a dose of up to 110 Gy is required at the base of the tumour if a ^{106}Ru plaque is applied and 40–60 Gy external radiation is required for the treatment of a retinoblastoma. With such high doses, cataract is inevitable if the lens is not shielded. Radiotherapy in and around the eye and orbit can have short- and long-term side-effects (Box 9.1).

Chemical agents

Toxic chemicals of any sort will damage the delicate ocular tissue when applied externally. Acid and alkali burns are encountered most commonly in clinical practice, but detergents can also cause significant damage.

- *Acid burns:* acids (hydrochloric, nitric and sulphuric) and acidic fluids coagulate proteins, so that diffusion through the cornea and the sclera into the eye may be limited. Necrosis of epithelial and stromal cells leads to scarring in the conjunctiva with eyelid distortion (entropion and ectropion). Corneal scarring requires keratoplasty to restore vision.
- *Alkali burns:* alkalis, such as ammonia and sodium hydroxide, pass through tissue easily and the high pH is sufficient to destroy the cells of the lens, uveal tract and retina as the alkaline fluid diffuses through the vitreous as far as the optic nerve.
  - *Iron*: Iron is necessary for life, but excess iron can be toxic to tissues. Iron is thought to damage tissues primarily by generating oxygen free radicals. Iron can be deposited in the cornea in keratoconus. It also has a role in cataractogenesis and may play a role in pathogenesis of age-related macular degeneration.
  - *Copper:* Copper can be toxic to a range of tissues including the eye where it causes retinal degeneration and inflammation.

Toxins

Toxins are generally proteins derived from bacteria, plants and animals (e.g. snakes). Some have been purified and have value as therapy such as botulinum toxin. However, most toxins from organisms cause severe tissue damage, such as the exoenzyme S and exotocin A produced by *Pseudomonas aeruginosa*, which may cause a severe lytic keratitis (see Chapter 8, 480). Tobacco and alcohol abuse may lead to toxic-nutritional optic neuropathy with damage to the papillomacular bundle. Similarly, many drugs used to treat systemic disease can have toxic side-effects for the retina, e.g. ethambutol for tuberculosis.

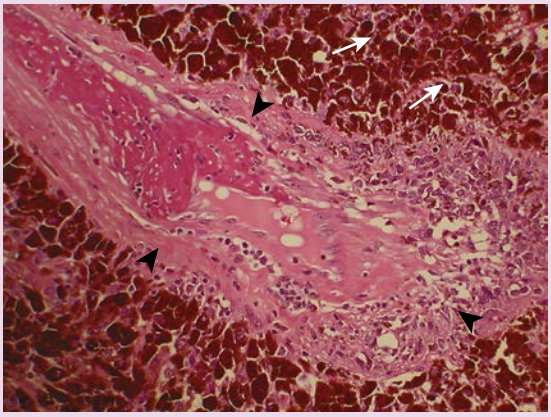
Viruses

Viruses that commonly cause ocular damage include herpes simplex virus and herpes zoster virus (see Chapter 8, p. 474).

Box 9.1 **Complications of irradiation therapy: effects on cells and tissues**

Radiation from any source has a profound effect on nuclear DNA, with fracture of chromosomes, dislocation and translocations. The consequence is tissue destruction and suppression of cell division, which is essential for the ablation of tumour tissue. The side-effects for non-tumour tissues are:

- Endarteritis: infiltration of the vessel wall by inflammatory cells and proliferation of spindle cells (myofibroblasts, intimal fibroplasia) within the internal elastic lamina. There is endothelial cell swelling and fibrin deposition. Narrowing of the vessel lumen contributes to tumour destruction by ischaemic necrosis (see figure. White arrows indicate necrotic melanoma cells. Black arrowheads show a blood vessel with inflammatory cells in the wall and fibrin in the lumen.). A long-term secondary complication of radiation arteritis is dilation of the capillary bed (telangiectasia). Radiation vasculopathy causes leakage of plasma constituents in retinal vessel walls.
- Loss of hair, teeth and glandular tissue. Irradiation of the orbit damages lacrimal gland tissue and leads to a dry eye.
- Massive necrosis of normal tissue when the radiation dose is excessive.
- An increased risk of mutations, with malformation in offspring and the induction of a second malignant tumour.



Dysregulated immunity

Several autoimmune diseases, such as rheumatoid arthritis and multiple sclerosis, affect the eye and will be discussed under the section on inflammation (p. 506). Reduced immunity may lead to opportunistic infections such as cytomegalovirus retinitis.

Nutritional deficiencies

Vitamin A deficiency leads to night blindness as it is a precursor to the formation of the photopigment rhodopsin. Vitamin A also has a role in mucus production with reduction in goblet cells and lacrimal gland secretion such that later xerophthalmia and keratomalacia occur.

Vitamin B_{12} deficiency is implicated in tobacco/alcohol optic neuropathy.

Trace metals such as iron, zinc and copper play an important role in retinal physiology. Iron deficiency can affect the retina and in severe cases causes retinal haemorrhage. Zinc deficiency is associated with poor dark adaptation. Excess copper is toxic to the retina. Copper deficiency is associated with optic neuropathy.

Genetic abnormalities

It is not surprising that errors occur during complex genetic activities. Germ cells and proliferating somatic cells (including stem cells) are susceptible to such errors. When abnormalities arise in germ cells, the abnormality may be transmitted to the next generation. Abnormalities in somatic cells may result in cell death or malignant transformation. Genetic abnormalities may involve whole chromosomes, parts of chromosomes, gene clusters or single genes.

Chromosomal abnormalities.

- Polypoidy occurs when chromosomal numbers are increased by an exact multiple of normal. This may be seen in hypertrophied muscle cells and ageing cells but does not usually occur in gametogenesis as it is incompatible with life.
- Aneuploidy occurs when the number of chromosomes is increased, usually by one (trisomy). This may occur in Down syndrome.
- Structural abnormalities may arise during replication and result in duplication and deletion of gene clusters of single genes.

Single gene defects. These are caused by defects in single genes. They usually affect the structure of the resulting protein product and can have wide-ranging effects. They can be inherited as an autosomal dominant (e.g. neurofibromatosis type 1, see Box 9.2), autosomal recessive (e.g. Wilson's disease, see Box 9.3; Tay–Sachs disease) or sex-linked (e.g. some forms of retinitis pigmentosa) disorder. Whilst most genetic disorders occur in nuclear DNA, it should be remembered that mitochondria also contain DNA in chromosomes, which are self-replicating and encode enzymes involved in oxidative phosphorylation. When mutations

Box 9.2 **Neurofibromatosis type 1 (NF1) and the eye**

This is an autosomal dominant condition with almost complete penetrance as a result of a mutation in the *NF-1* gene on chromosome 17. It is characterized by café-au-lait spots and neurofibromas on the skin. Ocular abnormalities include the following:

- *Eyelids*: café-au-lait spots and neurofibromas can affect the lids. Plexiform neurofibromas may cause cosmetic and functional problems.
- *Orbit*: proptosis may occur because of optic nerve gliomas or orbital plexiform neurofibromas. Bony defects may occur in the orbital floor or roof.
- *Anterior segment*: Lisch nodules, melanocytic hamartomatous lesions of the iris, are a defining feature of NF1. Enlarged corneal nerves and conjunctival, episcleral and scleral involvement by neurofibromas may occur.
- *Glaucoma*: the major cause of glaucoma is infiltration of the angle by a neurofibroma obstructing aqueous flow. Goniodysgenesis may also occur.
- *Retina and choroid*: choroidal hamartomas consist of a mixture of melanocytic and neuronal elements. Hamartomas of the retina and retinal pigment epithelium may also occur.
- *Optic nerve*: optic pathway gliomas are the cause of most visual morbidity associated with NF1. They occur in 15% of patients and are usually juvenile pilocytic astrocytomas. They can involve the optic nerve, chiasm, optic tract and hypothalamus.

Box 9.3 **Wilson's disease**

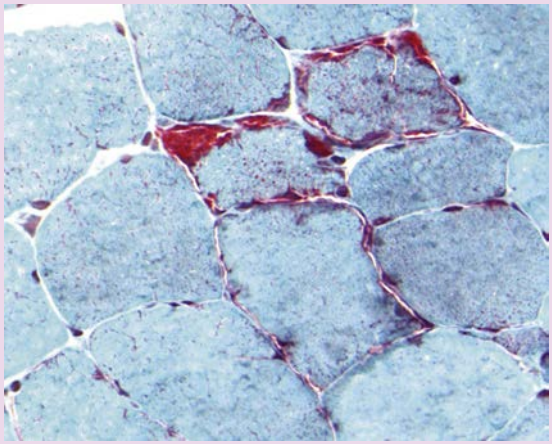
This is an autosomal recessive condition caused by mutations in the Wilson's disease protein (*ATP7B*) gene, which encodes a cation transport enzyme that transports copper into bile and incorporates it into ceruloplasmin. Abnormal transportation of copper results in its accumulation, principally in the liver and brain. This leads to cirrhosis and neurological and psychiatric symptoms.

The ophthalmic findings, which rarely cause significant visual loss, are:

- *Kayser–Fleischer rings* are rings in the cornea caused by deposition of copper in Descemet's membrane.
- *Sunflower cataracts* are caused by brown or green pigmentation beneath the anterior lens capsule.

Box 9.4 **Mitochondrial disorders and the eye**

Inherited mitochondrial diseases are rare and mainly affect the central nervous and musculoskeletal system. Abnormal mitochondria are found in the periphery of skeletal muscle fibres and have a characteristic 'ragged-red' appearance with Gomori trichrome stain as shown in the figure.



There are significant ophthalmic signs and symptoms in the following conditions:

- *Leber's hereditary optic atrophy*: point mutations occur in mitochondrial DNA resulting in abnormal ATPase 6. Young males lose vision because of demyelination in the optic nerve; the papillomacular bundle is most severely affected.
- *Chronic progressive external ophthalmoplegia (CPEO)*: this is a slowly progressive, bilateral symmetrical ocular muscle dystrophy that starts in late childhood or early adulthood. Ptosis, external ophthalmoplegia and occasionally a pigmentary retinopathy occur.
- *Kearns–Sayre syndrome*: initial signs and symptoms are similar to CPEO but pigmentary retinopathy is more common with the greatest effect on the macula. Cardiac conduction abnormalities present after the development of ophthalmoplegia.
- *MELAS syndrome (mitochondrial encephalomyopathy, lactic acidosis and stroke-like episodes)*: onset is usually in childhood with muscle weakness, recurrent headaches, vomiting and seizures. Stroke-like episodes involve temporary weakness on one side of the body. Eye signs include external ophthalmoplegia, atypical pigmentary retinopathy and nuclear cataract.

occur in mitochondrial chromosomes, the defect is passed on via the ovum because the sperm does not have cytoplasmic constituents; thus, mitochondrial disorders are normally maternally transmitted (see Chapter 3, p. 144 and Box 9.4).

Mechanism of cell death

Cells can die through a variety of processes, some of which cause more 'bystander' tissue damage than others. The nature of the injurious agent (e.g. if it is an organism, its virulence) or set of conditions (e.g. if it is a nutritional deficiency, then the level of deficiency) determines the outcome for the cell or tissue. Mild injury may be reversible. Severe injury will result in death either by necrosis or by apoptosis.

Necrosis

Necrosis describes death of a cell or group of cells, typically following severe hypoxia, physical or chemical injury. It is always pathological. There is a rapid depletion of intracellular energy systems. The cell membranes and intracellular organelles are disrupted and typically there is an inflammatory reaction.

Apoptosis

Unlike necrosis, apoptosis or programmed cell death is important in both health and disease. It is a rapid process usually affecting single cells within a population of healthy cells. Unlike necrosis, apoptosis is an active process involving protein cleavage and DNA degradation. The molecular events occurring in apoptosis are outlined in Fig. 9.1.

Sublethal cell injury

Various injurious agents will cause reversible cell damage without cell death. This may result in the following cellular changes:

Hydropic swelling. This occurs when cells are incapable of maintaining ionic and fluid homeostasis. This leads to a loss of function of plasma membrane energy-dependent ion pumps. The cell and organelles become swollen, often with cytoplasmic vacuolation. Blunt trauma may cause widespread retinal oedema (commotio retinae), which may be caused by transient spasm of the retinal vessels, producing ischaemia and damage to endothelial cells with leakage into the tissue. However, it is equally likely to result from transient interruption of axoplasmic flow in the ganglion cell processes.

Atrophy. This is a decrease in cell size or number resulting in shrinkage of affected tissues and organs. Blunt trauma may also cause shearing of photoreceptors, leading to atrophy of the photoreceptors and focal hypertrophy of the RPE (see later), which extends into the retina (pseudoretinitis pigmentosa).

Lesser degrees of damage may allow the cell time to adapt so that it is able to withstand injury. Mechanisms of adaptation include:

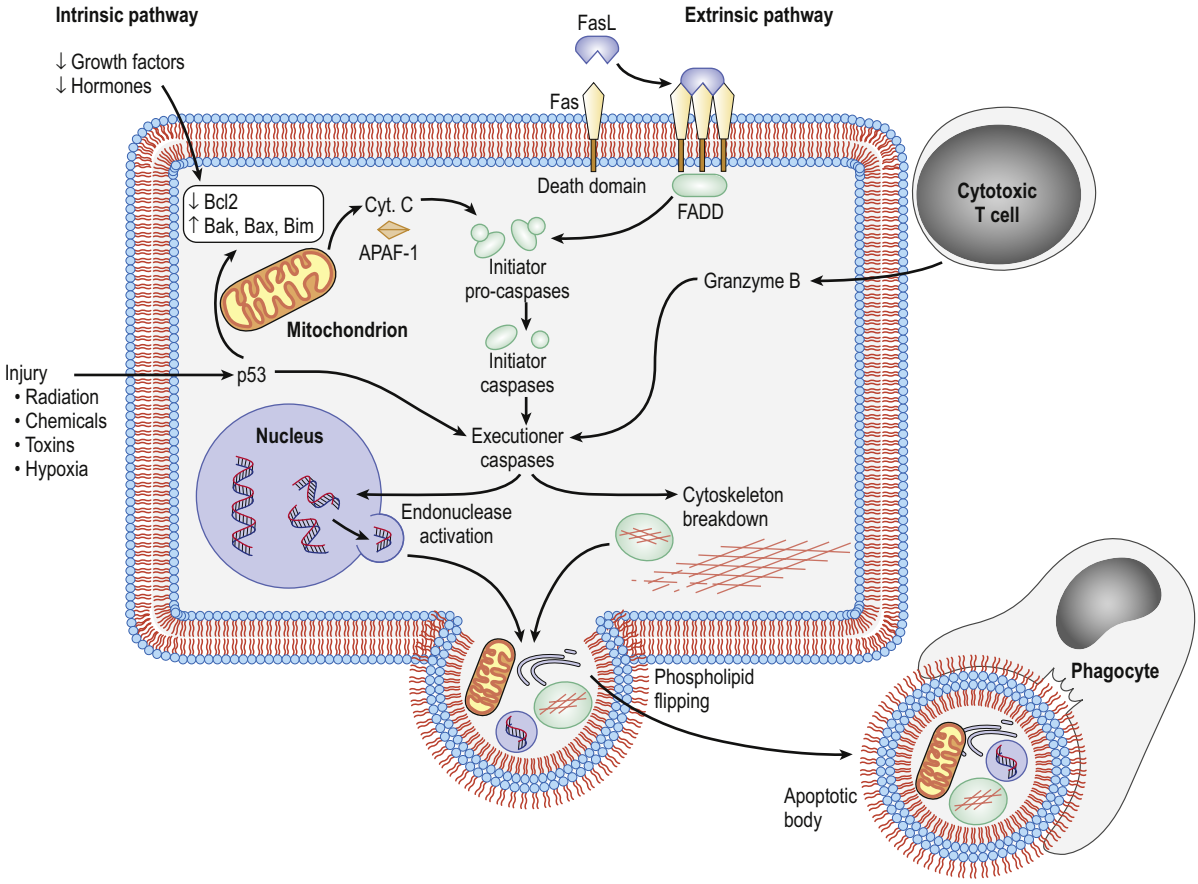


Fig. 9.1 Mechanisms of apoptosis. Apoptosis can be induced by intrinsic or extrinsic pathways. Forms of cell injury or reduction in growth factors or hormones can activate the intrinsic or mitochondrial pathway either by directly influencing the balance of anti-apoptotic proteins (Bcl2 family) and pro-apoptotic proteins (Bak, Bax and Bim) or by inducing activation of p53, which acts on these proteins. An excess of pro-apoptotic proteins increases mitochondrial permeability and allows leakage of other pro-apoptotic proteins (e.g. cytochrome C) into the cytoplasm. Cytochrome binds apoptosis activating factor-1 *(APAF-1)* and this complex can cleave initiator caspases. The extrinsic pathway is initiated by engagement of cell surface death receptors with the appropriate ligand. Cell surface death receptors include Fas and Type1 TNF receptor. Fas ligand *(FasL)* cross-links several receptors, allowing the death domains to come in contact, forming a binding site for Fas-associated death domain *(FADD)*. This in turn can activate pro-caspases. The execution phase is then carried out by a cascade of further caspases. Cytotoxic T cells can directly activate executioner caspases via granzyme B. Executioner caspases also activate endonucleases, which break up nuclear chromatin. The cell breakdown products are packaged into the apoptotic bodies. Flipping of membrane phospholipids allows rapid phagocytic recognition of apoptotic bodies, avoiding an inflammatory response.

Hypertrophy. This occurs when the cells increase in size, often as a result of increase in functional demand. This can occur in the RPE following injury (see earlier).

Metaplasia. This is a change from one type of differentiated tissue to another. In dry eye, there is a reduction in tear production and this results in a change of the normal conjunctival epithelium with goblet cells to stratified squamous type.

Healing and repair

The eye, like any other tissue, responds to injury by an initial acute inflammatory response followed by vascularization and wound closure with a fibroblastic scarring response. These responses may fundamentally alter the architecture and specific function in the tissue and can have serious effects on vision. For example, the wound response of the retina to

retinal hole formation and detachment is to stimulate glial proliferation, termed proliferative vitreoretinopathy (PVR). This proliferation does not, however, restore retinal and choroidal integrity but has the opposite effect by contracting and shortening the retina, to the point that retinal reattachment may be surgically impossible. Similar serious consequences result from scarring in other ocular tissues.

Cornea

Healing in the cornea usually leads to a corneal opacity (scar). The corneal epithelium regenerates at the limbus (limbal stem cells, see Chapter 4, p. 211) and spreads rapidly across the cornea. Bowman's layer does not regenerate. Stromal keratocytes transform to fibroblasts and myofibroblasts to heal stromal wounds. Transparency is lost because the collagen fibres are not aligned properly (see Chapter 4, p. 217). Desçemet's membrane, the basement membrane of the corneal endothelium, does not regenerate. The corneal endothelium fills in defects by sliding and in so doing deposits secondary layers in Desçemet's membrane: the membrane is elastic and there is often recoil at the edge of a deficit.

Iris

The presence of fibrinolysins in the aqueous inhibits fibrin clot formation and scar tissue does not appear to form in the iris stroma; this, plus the continuing flow of aqueous from posterior to anterior chamber of the eye, leads to the persisting patency of defects made by iridectomy and iridotomy procedures (with beneficial clinical results). Proliferation of the iris pigment epithelium may occur in response to trauma. Stromal melanocytes disrupted by trauma or inflammation can lead to colour changes in the iris.

Lens

The lens epithelium responds to some forms of trauma and other metabolic disturbances by undergoing fibrous metaplasia, but in general the lens does not heal in the true sense but rather responds to trauma by becoming opaque (cataract).

Retina

Damaged nerve cells are replaced by glial cells (gliosis), which are derived from perivascular astrocytes and Müller cells. In the RPE there is proliferation and metaplasia to fibrous tissue. A combination of glial cells and metaplastic RPE cells are found in preretinal membranes, particularly when such cells migrate via a retinal hole (small breaks in the retina that can occur as a result of high myopia, vitreous separation and certain other conditions; fluid accumulating under the retina results in retinal detachment). Fibrosis/gliosis due to PVR (see earlier) occurs in cases of retinal detachment and is the leading cause of failure in retinal detachment surgery.

Choroid

The melanocytes of the choroid do not proliferate in response to trauma; scar tissue in the choroid is derived from scleral fibroblasts.

Sclera

Scars are formed by proliferation of episcleral fibroblasts. Secondary fibrosis may contribute to failure of drainage procedures following glaucoma surgery.

Optic nerve

Trauma is followed by axonal loss and demyelination with reactive proliferation of glial cells and connective tissue cells in the meninges.

Inflammation

What happens in acute inflammation?

Inflammation is the dynamic process by which living tissues react to injury (Fig. 9.2). The injurious agent may be physical, chemical, infective or immunological. The classic signs of acute inflammation are redness, heat, swelling, pain and loss of function. These signs can be explained by changes occurring at the microscopic level. Hyperaemia is associated with microvascular changes and accounts for redness and heat. Following injury, there is an initial period of vasoconstriction, followed by capillary and then arteriolar dilatation. This is because of the direct effect of injury on the vessels in addition to release of chemical mediators from damaged cells. Exudation then occurs, accounting for swelling. This is the increased passage of protein-rich fluid through the vessel wall into the interstitial tissue. This increase in fluid results in dilution of toxins and provides protective antibodies, fibrin and various factors to promote healing. Passage of protein is mediated by direct endothelial injury and chemical mediators, including histamine, bradykinins and leukotriene. Fluid movement results from increased filtration pressure partly because of hyperaemia and loss of proteins from the capillaries. Following exudation, leucocytes (neutrophils and monocytes) migrate to the site of injury (Fig. 9.2). The cells migrate in response to an increasing concentration gradient of the chemotactic agents. Chemotactic agents are released from other leucocytes (cytokines), complement components (C3a), arachidonic acid derivatives (leukotrienes) or pathogenic bacteria. Once within the tissues, leucocytes clear the

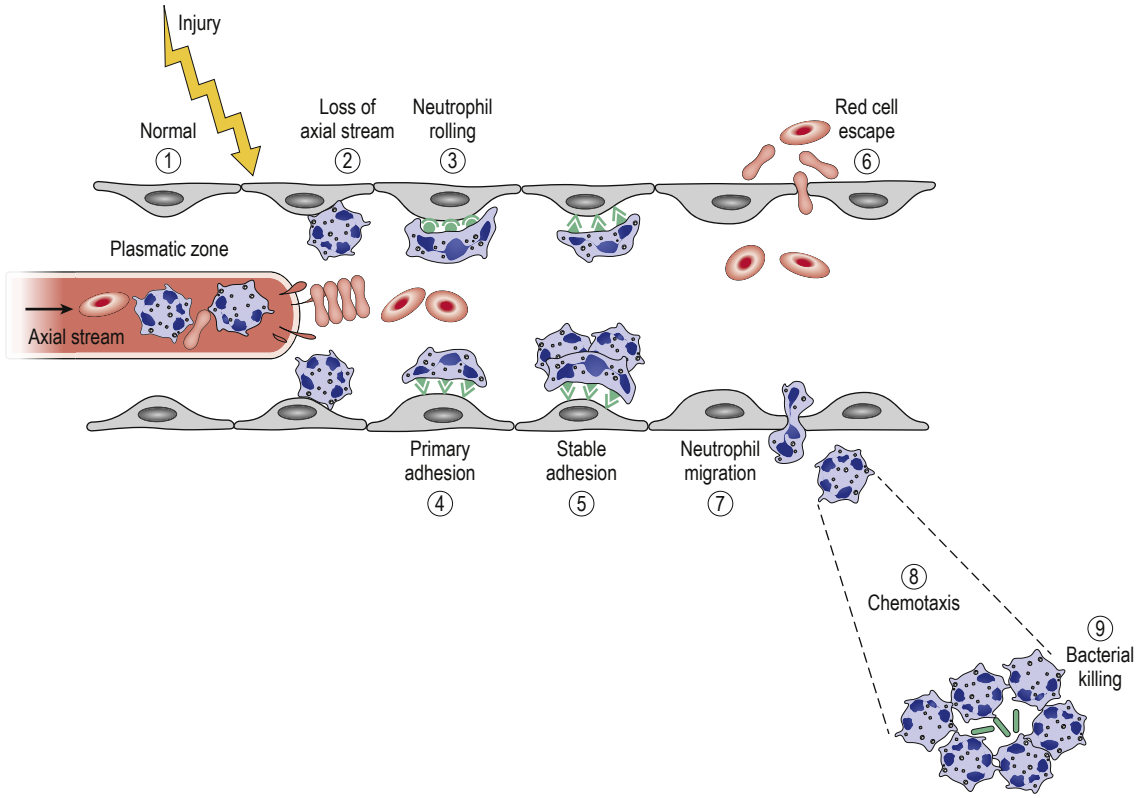


Fig. 9.2 Acute inflammation. In the normal situation the cellular constituents of the blood move in a central column known as the axial stream with the plasma constituents peripherally (**1**). Following injury, vascular changes lead to exudation of fluid and slowing of blood flow with loss of the axial stream of cells and margination of neutrophils (**2**). Neutrophil rolling then occurs as a result of endothelial selectins recognizing carbohydrate groups on the neutrophils (**3**). Following this loose adhesion, ICAM-1 and VCAM-1 expressed on the endothelial cells form a primary adhesion by binding integrins expressed on the neutrophils (**4**). A more stable adhesion forms with aggregation of neutrophils (**5**). Red cells begin to escape passively through gaps between endothelial cells (**6**). Transendothelial migration and extravasation of neutrophils occur due to integrin adhesion to endothelial PECAM-1 and ICAM-1 (**7**). There is enzymatic degradation of the basement membrane and neutrophils then travel along a chemotactic gradient to the site of injury (**8**). At the site of injury neutrophils will engulf opsonized bacteria and kill them either by oxygen-dependent formation of free radicals or by oxygen-independent release of lysozyme (**9**).

injurious agent by phagocytosis. This process involves opsonization of bacteria by complement components before engulfment within the leucocyte. Once within the leucocytes, the lysosomes fuse with the phagosome and bacteria may be killed by oxygen-dependent formation of free radicals or by activation of lysosomal enzymes. Various chemical mediators are also released from inflammatory cells or are present within plasma. Those released from inflammatory cells include histamine, serotonin, prostaglandins and leukotrienes as well as a range of cytokines. These mediators may result in vasodilatation (serotonin, prostaglandins), increased vascular permeability (histamine, leukotrienes) and lymphocyte proliferation and macrophage activation (cytokines). Those present in plasma are interrelated cascade systems, including the clotting cascade, fibrinolysis, complement system and bradykinin system. The outcome of acute inflammation depends on various factors, including the infecting organism and the extent of tissue necrosis. Inflammation may resolve, suppurate, repair with organization and scarring, or progress to chronic inflammation.

Inflammation in the eye has special features characteristic of the tissue involved. For instance, in the cornea a small traumatic erosion may resolve without scarring. However, bacterial keratitis may lead to formation of a corneal abscess that may ultimately heal with extensive scarring. Ongoing chronic inflammation may occur with herpes simplex stromal keratitis. The 'privileged' status of some ocular tissues such as the retina and optic nerve (and the brain) modifies their response to trauma and infection and thus modifies the overall inflammatory response. However, the other ocular tissues, although classically considered 'privileged', often respond to injury and infection in much the same way as other tissues such as the

skin, lung and gut and whether they display immune privilege is now being questioned (Forrester at al., 2018).

How does infection cause inflammation?

Infection occurs when microorganisms invade the sterile internal body tissues and then multiply. The local reaction to infections is usually inflammatory and is evoked by cell damage and death. There are many classes of organisms which cause inflammation (see Chapter 8) and each does so in different ways. Some of the pathology induced by these organisms is briefly described here.

Bacteria

Intraocular bacterial infections are rare but highly destructive. The pathogenicity of bacteria varies according to the species (see Chapter 8, p. 480) and this is reflected in the nature of the immune response to the organism.

Pyogenic infection. Pyogenic bacterial infection of the intraocular compartment is termed endophthalmitis and may be exogenous or metastatic. Panophthalmitis (in which the infection involves the whole of the ocular and periocular tissues) describes a rapid and devastating tissue destruction, which may be complete by 48 hours (Fig. 9.3). Gram-positive cocci are among the commoner organisms that produce a purulent infection when introduced into the eye either by accidental or surgical trauma or by haematogenous spread from another site. The vitreous is an ideal medium for bacterial proliferation, in that the sites for ingress of inflammatory cells (e.g. retinal vessels, the optic disk and pars plana) are some distance from the proliferating pathogens. (Therapy may also be less than optimal because the blood retinal barrier impedes the diffusion of antibiotics; see Chapter 6, p. 353).

Neutrophils are the predominant cell type in metastatic endophthalmitis, i.e. when bacteria enter the eye via the bloodstream, as in bacterial endocarditis or after contamination of an indwelling intravenous catheter. During bacterial killing, neutrophils may release lytic enzymes that can lead to the total destruction of all layers of the retina.

The corneoscleral envelope normally provides an excellent barrier to invasion by bacteria. However, if the cornea is in some way compromised, as for instance during the inappropriate use of steroids in corneal ulceration, the prolonged and incorrect use of contact lenses or inadequate wound closure after intraocular surgery, direct bacterial entry through the cornea may occur.

Several factors increase the risk of endophthalmitis (Box 9.5).

Chronic infection. Mycobacteria (tuberculosis, leprosy), actinomycetes (skin and lung infection) and spirochaetes (syphilis, yaws) are responsible for chronic destruction of tissue. *Borrelia burgdorferi*, a spirochaete transmitted by ticks, may cause arthritis, neurological disease and conjunctivitis (Lyme disease). *Bartonella* spp., which causes cat scratch disease, has also been implicated as a cause of neuroretinitis. Pathogens that produce a chronic infection induce humoral and cellular responses in the host (see Chapter 7); the latter is characterized by macrophages, lymphocytes and plasma cells. Many pathogenic organisms, such as *Mycobacterium tuberculosis*, have the capacity to survive within the host macrophage. In the case of *M. tuberculosis*, macrophages attempt to limit the spread of the organism by accumulating around the dead and dying (necrotic) cells killed by the organism. The macrophages then become more elongated, taking on an epithelial morphology (epithelioid macrophages). This collection of epithelioid macrophages is known as a granuloma (Fig. 9.4). Fusion with neighbouring macrophages forms a characteristic multinucleate giant cell (Langhans' cell). The central mass of dead tissue within the granuloma appears cheese-like macroscopically: hence, the terms 'caseous necrosis,' 'caseating' and 'caseation'.

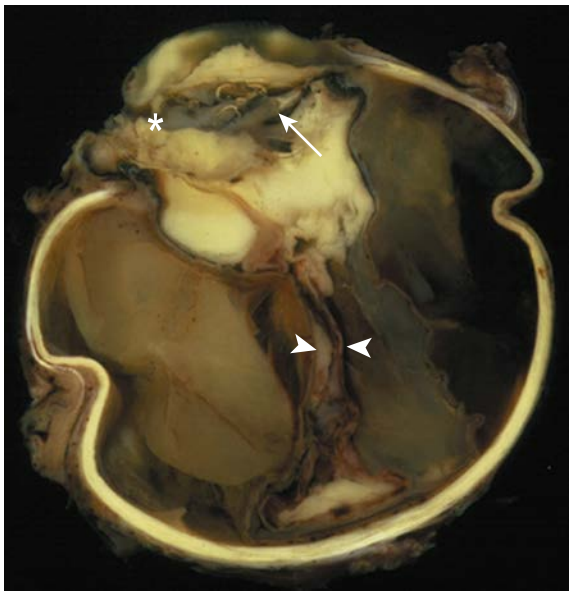


Fig. 9.3 Enucleation was required after a cataract operation when an intraocular lens implant *(arrow)* was complicated by a dehiscence of the wound *(*)*. Gram-positive cocci were found in the vitreous abscess. The retina *(arrowheads)* was detached by haemorrhage.

Viruses

The following viruses most commonly cause ocular disease: herpes simplex virus, herpes zoster virus (varicella group) and cytomegalovirus.

Box 9.5 **Factors that increase the risk of endophthalmitis**

- The major factor in non-accidental post-surgical endophthalmitis is active host infection, e.g. bacterial conjunctivitis.
- The major factor in accidental post-injury endophthalmitis is retention of infected foreign material, especially organic material.
- Poor surgical technique, inadequate wound closure and leaking blebs, e.g. after drainage procedures, can cause endophthalmitis.
- Contaminated intraocular lenses are a common cause of late low-grade bacterial endophthalmitis.
- Debility, chronic corneal ulceration and immunosuppression are rare causes of endophthalmitis.
- Contact lens wear with poor hygiene remains a cause of severe intraocular infection if corneal ulceration occurs.

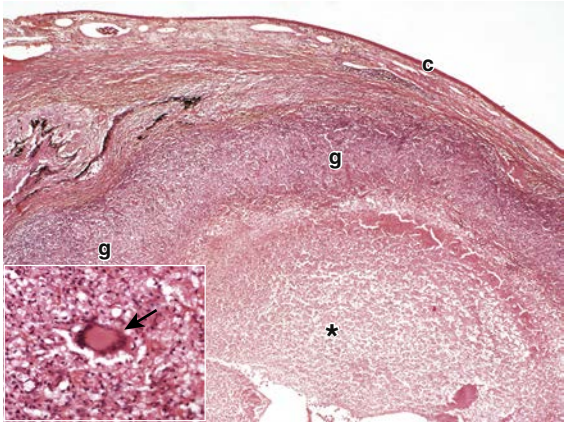


Fig. 9.4 A tuberculoma involving the anterior chamber. The inflamed cornea *(c)* is anterior. There is caseous necrosis *(*)* filling the anterior chamber, which is surrounded by granulomatous inflammation *(g)*. The granulomatous inflammation is shown on higher power in the insert, where epithelioid macrophages have fused to form a multinucleated giant cell *(arrow)*.

Herpes simplex keratitis. Type 1 herpes simplex virus causes superficial corneal ulceration, which is finger-like or dendritic within the epithelium. Primary herpes simplex infection usually occurs through the oral mucosa, the lips or the skin of the face. This is followed by transneural and, subsequently, latent viral infection in the neurones of the trigeminal ganglion and the sympathetic ganglia. The virus is morphologically detectable by ultrastructural investigation within the corneal cells (keratocytes and epithelial cells) and nerves (Fig. 9.5). A variety of possible triggers

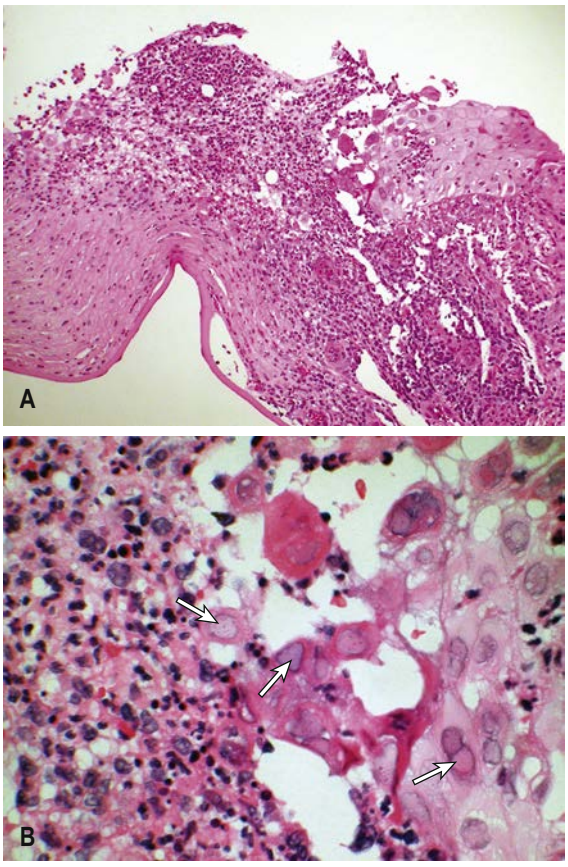


Fig. 9.5 (**A**) Herpes simplex ulcer in the cornea showing acute inflammation in the ulcer base. (**B**) At higher power the epithelial cells at the edge of the ulcer contain intranuclear inclusion of herpes viral particles *(arrows)*.

(ultraviolet light, cold) reactivate the virus, which migrates along the sensory nerves to the corneal epithelium. The virus then invades, replicates and spreads within the epithelium. Effective antiviral therapy has dramatically reduced the complications of herpes simplex keratitis. These include chronic fibrosis and scarring (disciform keratitis) with stromal vascularization and persisting chronic inflammation: leakage of lipid-rich plasma into the corneal stroma leads to pale yellow deposits (secondary lipid keratopathy).

Herpes zoster ophthalmicus. Herpes zoster virus infects the ganglia and branches of sensory nerves such as the trigeminal nerve. When the virus is reactivated, it replicates and produces vesicle formation in the skin in the distribution of the affected nerve or its branches. The eyelids, conjunctiva, cornea and uveal tract are involved in the inflammatory process. A lymphocytic infiltrate appears around the long and short ciliary nerves and is present in the choroid and ciliary body.

Box 9.6 **Polymerase chain reaction and the eye**

Polymerase chain reaction (PCR) involves the selective amplification of specific segments of DNA. The basic reaction involves repetitive cycles of DNA synthesis. Each cycle consists of three steps:

- *Denaturation:* the first step involves denaturation of the target nucleic acid, which renders it single stranded.
- *Annealing:* denaturation is followed by annealing of synthetic oligonucleotide primers specifically designed to hybridize to the target nucleic acid region.
- *Extension:* the third step involves extension from the annealed primer catalysed by a DNA polymerase enzyme. In a typical PCR analysis, 20–40 cycles are carried out with successive products becoming templates for subsequent cycles such that there is exponential amplification of the target region.

The basic technique may be adapted for various applications.

1. *Detection of genetic mutations and deletions (e.g. Leber's hereditary optic atrophy)*
   - Direct sequencing of the PCR product
   - Detection of single-strand conformation polymorphisms (SSCP)
   - Detection of restriction fragment length polymorphisms (RFLP).
2. *Detection of pathogens*
   - PCR is useful for detecting various pathogens, particularly viral pathogens (e.g. hepatitis B, cytomegalovirus, herpes simplex virus).
3. *Detection of changes in gene expression*
   - Reverse transcription PCR is based on the comparison of the amount of PCR product generated with the amount produced from a known concentration or copy number of control amplification targets in the same reaction. This technique can be useful for studying clonality in lymphoma.

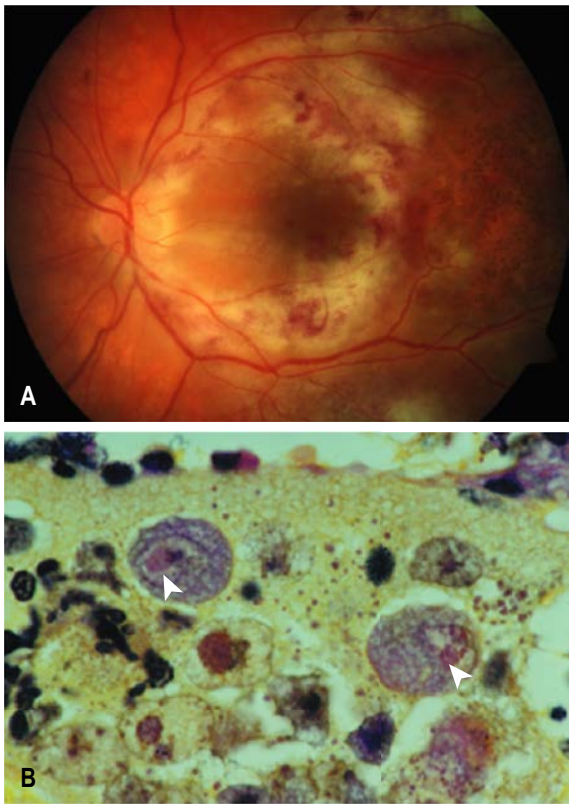


Fig. 9.6 (**A**) Fundoscopic image of acute CMV retinitis showing areas of retinal necrosis along the retinal arcades associated with small areas of haemorrhage. (**B**) Histology of CMV retinitis shows the necrotic retina contains enlarged cells in which characteristic owl-eye intranuclear inclusions are present *(arrowheads)* (phloxine tartrazine stain). (**A** courtesy Dr Graeme Williams, Glasgow.)

Acute retinal necrosis. Acute retinal necrosis can be unilateral or bilateral and usually occurs in immunocompetent individuals. It is usually caused by infection of retinal neural cells with herpes simplex or varicella zoster virus. Morphological distinction of these viruses can be difficult in retinal biopsies and identification of viral DNA using polymerase chain reaction (Box 9.6) may be more helpful in providing an accurate diagnosis. Enucleation specimens show sectorial or massive haemorrhagic retinal necrosis, associated with vitreous exudation and choroidal inflammatory cell infiltration. Intranuclear viral inclusion bodies can sometimes be seen by light microscopy.

Progressive outer retinal necrosis. In immunocompromised individuals, herpes simplex and herpes zoster viruses may cause destruction of the outer retina without the accompanying vitritis, retinal vasculitis or papillitis usually associated with acute retinal necrosis.

Cytomegalovirus retinitis. Before antiretroviral therapy became available, cytomegalovirus retinitis was a common ocular infection in individuals with acquired immunodeficiency syndrome (AIDS). In its fulminant stage it is characterized by progressive areas of retinal necrosis, usually without haemorrhage (Fig. 9.6). It may also present with granular appearance in the peripheral retina or in a perivascular distribution known as 'frosted angiitis'. Infected cells are often enlarged with a characteristic owl-eye inclusion body (Fig. 9.6). Cytomegalovirus infection in immunocompetent individuals is now recognized as one cause of a chronic/recurrent anterior uveitis, often with raised intraocular pressure intraocular pressure (Posner-Schlossman syndrome).

Box 9.7 **Trachoma**

Trachoma histology is described in four stages:

- Stage I: conjunctival oedema, lymphoid follicles and a diffuse infiltrate of lymphocytes, plasma cells and polymorphs.
- Stage II: florid conjunctival inflammation with more follicles, epithelial thickening, pannus and fibrovascular replacement of the conjunctival stroma resulting in formation of papillae.
- Stage III: progressive fibrosis without follicles or significant inflammation.
- Stage IV: severe scarring with internal deformation of the lids (entropion) and trichiasis, which leads to abrasion of the cornea by the lashes. Reduction in tear production because of inflammation of lacrimal gland and duct. Squamous metaplasia of the conjunctival epithelium.

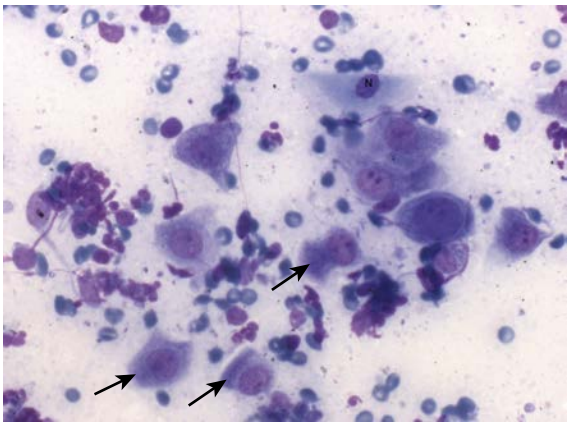


Fig. 9.7 Giemsa-stained conjunctival smear showing conjunctival epithelial cells and pus cells. There are basophilic intracytoplasmic inclusions in some cells *(black arrows)*.

Measles. Measles is re-emerging as a disease in developed countries due to loss of confidence in vaccination programmes but is recognized as a leading cause of blindness worldwide.

Measles can cause conjunctivitis and/or keratitis with corneal ulceration and scarring. Less commonly, measles virus infects the retina or optic nerve causing retinopathy and optic neuritis, which can both lead to blindness.

Chlamydia

Chlamydia trachomatis (see Chapter 8), obligate intracellular bacteria causing trachoma and adult and neonatal inclusion conjunctivitis, is one of the commoner causes of human blindness worldwide. The histological stages of trachoma (see p. 511 for corresponding clinical features) are described in Box 9.7. Inclusion body conjunctivitis, so named because of the intracytoplasmic bodies seen in Giemsa-stained conjunctival scrapings (Fig. 9.7), is much less sight-threatening and is associated with genitourinary infection.

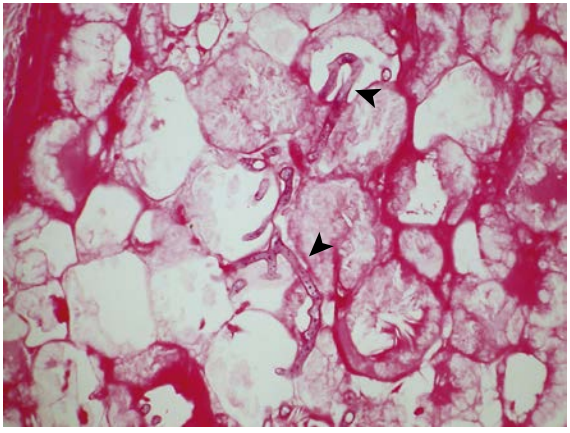


Fig. 9.8 Mucor is the largest of the pathogenic fungi and consists of broad, non-septate branching hyphae *(arrowheads)*. In this case the hyphae are present within infracted orbital fat.

Fungi

Fungal pathogens (see Chapter 8, p. 482) vary depending on the part of the eye affected. For example, *Candida* spp. is the most common cause of endogenous endophthalmitis, particularly in drug addicts. *Candida* spp. also causes corneal infections, most often post-surgery. *Aspergillus* spp. and *Fusarium* spp. cause a significant proportion of traumatic infectious keratitis and may invade the corneal and anterior segment.

Mucormycosis is the result of a blood-borne infection and occurs in patients with poorly controlled diabetes or who are immunocompromised. Several fungi of the order Mucorales can cause mucormycosis. There is a predilection for the lumen of blood vessels, which are occluded by secondary thrombosis: the organism parasitizes the ophthalmic artery and its branches, leading to necrosis of the orbital tissues, the nose and the eye (Fig. 9.8).

Some fungal infections are more prevalent in certain regions of the world, e.g. in the USA, where coccidioidomycosis, cryptococcosis, histoplasmosis and blastomycosis are all reported as ocular pathogens. Fungi use innate immune receptors such as Dectin-1 signalling through CARD 9 (see Chapter 7, p. 400) on myeloid cells to mediate a host response in which Th17 cells provide protection. Genetic defects in Dectin-1 are associated with susceptibility to debilitating overwhelming fungal infections which cause extensive tissue destruction. In contrast, deficiency in NOD2 signalling (see Chapter 7, p. 400) offers protection against *Aspergillus* infection.

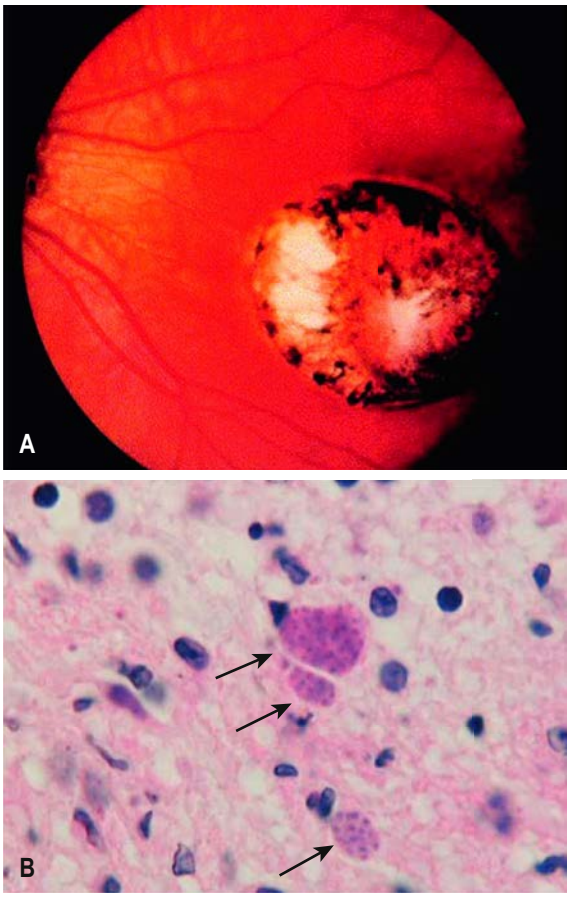


Fig. 9.9 (**A**) Quiescent toxoplasmic retinochoroidal scar in a congenitally infected patient. (**B**) *Toxoplasma* cysts *(arrows)* in a retina in which the normal architecture has been destroyed. (**A** reprinted from, Roberts, F., McLeod, R., 1999. Pathogenesis of toxoplasmic retinochoroiditis. *Paristol. Today*. 15(2), 51–57 with permission from Elsevier.)

Protozoal and metazoal infections

Several organisms from this group cause ocular pathology (see Chapter 8).

Toxoplasma gondii. This is the commonest protozoal parasite to infect the eye. Congenital infection may occur when a woman becomes infected for the first time during pregnancy. Infection causes a classic tetrad of clinical features (meningoencephalitis, hydrocephalus, intracranial calcification and retinochoroiditis). Disease severity depends on the trimester in which the infection occurs. Congenital ocular toxoplasmosis is a recurring and progressive disease because of the persistence of the parasite as bradyzoites within tissue cysts, which can reactivate. Foci of reactivation can be seen as an irregular area with associated vitreous haze at the border of a retinochoroidal scar (Fig. 9.9).

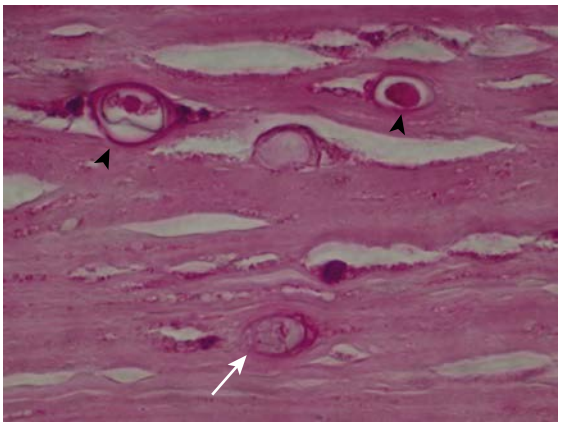


Fig. 9.10 *Acanthamoeba* cysts *(black arrowheads)* and a trophozoite *(white arrow)* lying between the corneal stromal lamellae in a soft contact lens wearer (PAS stain).

Acquired infection (see Chapter 8) is commonly asymptomatic, although some individuals will develop a flu-like illness with lymphadenopathy. Approximately one-third of the world's population has been exposed to toxoplasma infection as shown by the presence of anti-toxoplasma antibodies, the prevalence of which increases with age. Retinochoroiditis may occur as a result of acute acquired infection, but in contrast with congenital infection there is no pre-existing scar. *Toxoplasma* retinochoroiditis is less common in immunocompromised patients than cytomegalovirus retinitis but can cause extensive retinal necrosis. Individual parasites (tachyzoites) and cysts (containing bradyzoites) can be identified in paraffin sections (Fig. 9.9).

Acanthamoeba spp. The incidence of acanthamoeba keratitis initially increased with greater use of soft contact lenses, but this form of keratitis it is now less common because of increased use of disposable contact lenses. In the cornea, the acanthamoebae phagocytose remnants of dying keratocytes and polymorphs (Fig. 9.10). The organism can be difficult to identify without the use of immunohistochemistry. Acanthamoeba keratitis is painful but does not normally penetrate the cornea, remaining as a superficial keratitis; it responds only slowly to appropriate therapy.

Toxocara canis. The adult worm of *Toxocara canis*, which can be several centimetres long, lives in the intestinal tract of young dogs and the eggs are passed out with the faeces. Human infants may ingest the eggs, which release second-stage larvae in the stomach. The larvae of *Toxocara* can pass easily through the body tissues and the living organisms do not elicit an inflammatory response. The pathological consequences are usually related to death of the organisms, which causes an immune response. The most sensitive organs appear to be the liver, lung, brain and eyes where both a delayed-type and immediate-type hypersensitivity

response may occur. In the eye there are three possible outcomes:

- At the posterior pole, a slowly developing low-grade fibrous reaction within the retina can produce a tumour that resembles a retinoblastoma.
- In the mid-periphery, a more rapid active inflammatory reaction, which is characterized by the presence of numerous eosinophils, is followed by exudation into the retina and subretinal space, and then secondary retinal detachment.
- Inflammation and fibrosis occur in the vitreous base over the pars plana and can induce a form of posterior uveitis, easily mistaken for pars planitis.

Other parasites

In equatorial climates, infection by helminths such as *Wuchereria* spp., *Loa loa*, *Ascaris* and several other nematode larvae causes retinal damage and inflammation when microfilariae migrate into the retina and vitreous (see Chapter 8, p. 487). Inflammation can be relatively low-grade, resulting in a mottled pigmentary retinopathy (diffuse unilateral neuroretinitis, DUSN) caused by the worm migrating in the subretinal space.

Inflammation may occur without infection

Many forms of intra- and extraocular inflammation occur without evidence of infection due to 'immune-meditated' mechanisms.

Granulomatous inflammation

Macrophages accumulate often with multinucleate giant cells to form granulomas around a focus of tissue destruction. In addition to infectious diseases mentioned previously, it is also a feature of a number of non-infectious diseases. Immunologically, these reactions represent a delayed-type hypersensitivity response (DTH response; see Chapter 7).These granulomatous reactions include the following types.

Reactions to endogenous materials

Products of plasma, blood or cell breakdown can induce a granulomatous reaction when released into tissue. The commonest eyelid granuloma is the chalazion, which is a reaction to a blocked meibomian gland duct (Fig. 9.11). Rupture of a cyst (either an epidermal inclusion cyst or a dermoid cyst) releases keratin, which is an irritant and induces a chronic granulomatous reaction. Red cells and plasma in the extracellular matrix provoke an inflammatory reaction and the cellular response depends on the presence

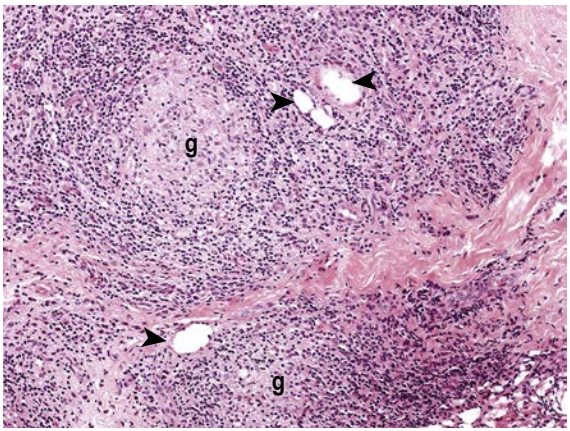


Fig. 9.11 In a chalazion, blockage of the meibomian gland results in lipid release in the form of fat globules and this stimulates the formation of a granuloma *(g)*. The lipid is dissolved during processing, leaving empty 'fat' spaces *(arrowheads)*. The granulomatous foci are surrounded by lymphocytes and plasma cells. Recurrence is common in chalazia, because the inflammatory reaction and the subsequent fibrosis interfere with drainage from the adjacent glands in the tarsal plate. Sebaceous gland carcinoma is an important entity in the clinical differential diagnosis of recurrent chalazion.

of fibrinolysins (plasmin and plasminogen activator) in the tissue fluids. In the anterior chamber, fibrin is diluted by the aqueous and is rapidly dissolved by fibrinolysins present in the aqueous. Thus, the cellular response to red cells is restricted to migration of macrophages from the iris into the anterior chamber. A similar reaction occurs in the vitreous after bleeding from a tear in the retina or from torn preretinal vessels in vasoproliferative retinopathy. Similar haemogranulomas also occur within the orbit after trauma followed by bleeding (traumatic blood cysts).

Reactions to exogenous non-biological materials

Implantation in tissue of vegetable or organic matter such as wood, for example following trauma, excites a similar cellular response to that of sutures derived from cotton or synthetic materials. Synthetic fibres or fragments of plant are seen in polarized light as birefringent particles surrounded by macrophages and lymphocytes. Metallic fragments are slowly dissolved in tissue fluids, but elements such as iron are toxic to the retina, which undergoes neuronal loss. Hence, retained metallic foreign bodies in the vitreous or retina are especially dangerous. Brass contains copper and tin and the reaction to copper ions is often pyogenic. On occasion, similar reactions may be seen to materials used in ophthalmic surgery. Remarkably, tiny foreign bodies, released during metal beating, may enter the

eye and lodge in the retina, vitreous or ciliary body unbeknownst to the patient, particularly if the vision remains unaffected. Slow release of metallic ions causes the retina degeneration and in the case of copper inflammation.

Non-granulomatous inflammation

Lymphocytes and plasma cells are found in a number of conditions in which the aetiology and pathogenesis are unknown. Many clinical forms of anterior and posterior uveitis, e.g. Behçet's disease, are characterized by diffuse and intense lymphocyte and plasma cell infiltration as well as a prominent neutrophil infiltration. Lymphocytic perivasculitis is a feature of demyelinating disease (multiple sclerosis) in the optic nerve and of retinal vasculitis. In endocrine exophthalmos there may be focal clusters of lymphocytes (lymphorrhages) within the extraocular muscles (see p. 515).

Inflammation due to autoimmune disease

The basic mechanisms of autoimmune disease pathophysiology have been dealt with elsewhere (see Chapter 7, p. 443). As indicated previously, many inflammatory diseases of the eye and adnexae occur in the absence of identifiable causative organisms. Some are associated with generalized connective tissue disease or diseases with a recognized autoimmune aetiology. Others are restricted to the eye and periocular tissues. The following section describes some of these disorders.

Sjögren's syndrome

Sjögren's syndrome is a disorder of the lacrimal gland, acinar glands of the conjunctiva (glands of Wolfring and Krause, see Chapter 1, p. 85) and salivary glands. The secretory acinar tissue of the lacrimal gland is destroyed by a lymphocytic infiltrate, with formation of lymphoepithelial lesions. Impaired secretion of saliva and tears leads to a dry mouth and dry eyes, respectively, and there is an associated loss of goblet cells in the conjunctival epithelium with squamous metaplasia of the surface epithelium. Patients with primary Sjögren's syndrome have specific antibodies (anti-Rho, anti-La).

Rheumatoid eye disease

Destruction of ocular tissue in rheumatoid arthritis (RA) is T cell-mediated but also includes an component of immune complex-mediated vasculitis. Indeed, several of the vasculitic autoimmune diseases, which include and systemic lupus erythematosus (SLE) involve both B cells and autoantibodies in addition to neutrophils, which display NETosis (see Chapter 7, p. 392) (Apel et al., 2018). Vasculitic phenomena and neutrophil activity are thought to underlie the impairment of blood flow to the anterior segment causes a necrotizing scleritis and peripheral corneal ulceration. Spontaneous central corneal ulceration (often without an inflammatory cell infiltration) occurs in rheumatoid disease (corneal melt), a phenomenon that is promoted by release of metalloproteinases. If the destruction of collagen in the sclera is severe, the tissue undergoes fibrinoid necrosis with a granulomatous reaction around the necrotic sclera. Thinning of the sclera leads to exposure of the underlying uveal tract (scleromalacia perforans), but ciliary body prolapse and perforation are uncommon. In some patients the inflammatory process is slower and is accompanied by fibrosis with massive thickening of the sclera (brawny scleritis). Posterior nodular scleritis, a localized inflammatory reaction at the posterior pole of the eye, causes macular oedema and the mass may simulate a malignant melanoma with or without an exudative retinal detachment.

Other ocular surface disease

Autoimmune disease is seen in the skin of the eyelid in bullous diseases such as pemphigus and mucous membrane pemphigoid, which are extremely debilitating diseases and very difficult to treat. In the conjunctiva, autoimmune responses against basement membrane components and the attachments of the epithelium lead to severe stromal fibrosis. Immunofluorescence studies performed on tissue submitted in Michel's transport medium show linear deposition of immunoglobulin G and sometimes C3 along the basement membrane (Fig. 9.12). Detachment of epithelium from the basement membrane leads to inadequate protection and predisposes to secondary inflammation and the exudation of fibrin. In the fornix, a fibrinous exudate provides a scaffold for fibroblastic migration; the subsequent scarring process leads to adhesions between the eyelids and the globe (symblepharon).

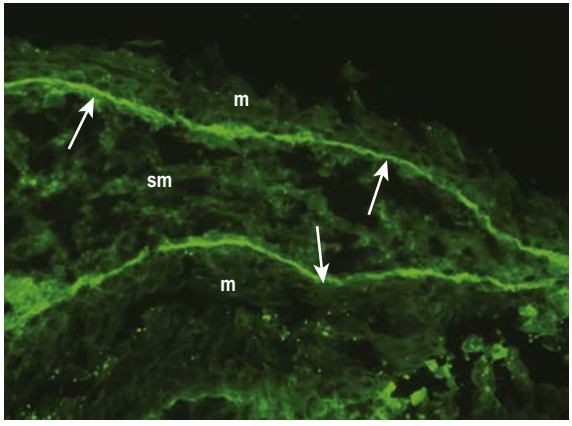


Fig. 9.12 In ocular cicatricial pemphigoid, immunofluorescence studies may show linear staining for IgG in the basement membrane *(arrows)* in the conjunctiva. There is only background staining in the mucosa *(m)* and submucosa *(sm)*.

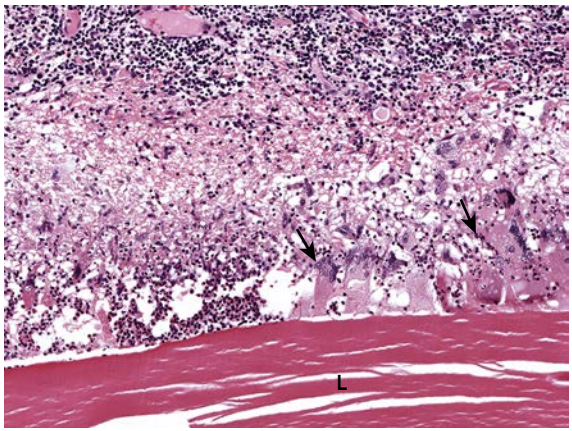


Fig. 9.13 In lens-induced uveitis, lens cortical matter *(L)* is attacked by a giant cell granulomatous reaction *(arrows)*. There is a surrounding lymphocytic infiltrate.

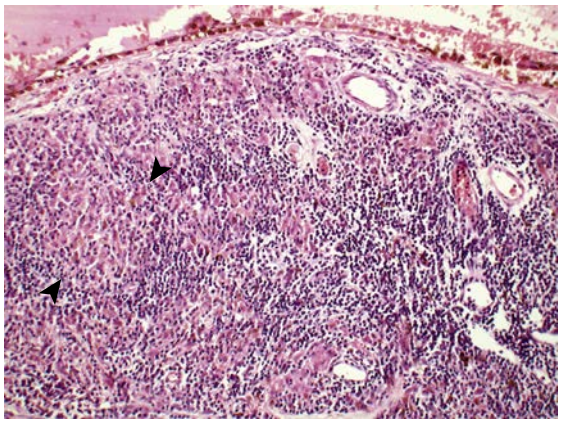


Fig. 9.14 In sympathetic ophthalmia there is expansion of the choroid by a chronic inflammatory infiltrate including granulomas *(arrowheads)*. A fine dusting of melanin pigment within macrophages is characteristic. The choriocapillaris is spared and the retinal pigment epithelium contains infiltrating macrophages.

Genetic predisposition has been shown in GWAS studies (see Chapter 3, p. 150) in patients who have polymorphisms in the galactocerebrosidase gene on chromosome 14 as well as other intergenic regions on chromosome 16.

Lens-induced uveitis

A cataractous lens contains proteins that are the breakdown products of the primary soluble crystallins and other proteins (see Chapter 4, p. 234). Leakage of lens protein into the anterior chamber, either spontaneously or as a result of trauma, may induce a massive giant cell granulomatous reaction (Fig. 9.13). Macrophages and lymphocytes enter the anterior chamber from dilated blood vessels in the iris and ciliary body and may enter the lens cortex itself directly through a rupture in the lens capsule. There are commonly significant numbers of neutrophils and eosinophils in the inflammatory infiltrate. The inflammatory cells then pass directly through the epithelium of the pars plana during the associated cyclitis.

Leakage of lens protein into the anterior chamber does not automatically induce a prominent inflammatory response. Sometimes lens protein is relatively inert and induces an uncomplicated macrophage response. This may be associated with a rise in pressure caused by outflow obstruction by engorged macrophages (phacolytic glaucoma). If the rupture is acute, the outflow system is blocked by lens matter (lens particle glaucoma).

Sympathetic ophthalmia

A bilateral granulomatous inflammation of the choroid, ciliary body and iris (panuveitis) can occur after injury to one eye; the injury usually includes uveal incarceration within the sclera. Subsequent sensitization to previously sequestered ocular antigens leads to posterior uveitis mediated by major histocompatibility complex class II-restricted $CD4^+$ T cells, which affect both the exciting eye and contralateral sympathizing eye. This immune response is directed against ocular self-antigens (see Chapter 7). The uveal tract becomes considerably thickened by an inflammatory infiltrate that includes collections of macrophages (Fig. 9.14). The inflammatory process also involves the retinal pigment epithelium with the accumulation of macrophages (essentially small granulomas) at this site (Dalen–Fuchs nodules).

Vasculitis

This is a group of conditions with inflammation primarily targeting components of the vessel wall, often accompanied by necrosis. The majority of the vasculitides are immune-mediated. These are classified according to the size of the vessel principally affected.

Large vessel vasculitis

Temporal (giant cell) arteritis. This systemic condition involves the cerebral arteries, the ophthalmic arteries, the posterior ciliary branches and the central retinal arteries. Diagnosis is important as this is a preventable cause of blindness most commonly caused by anterior ischaemic optic neuropathy. In susceptible arteries, local dendritic cells recruit and activate $CD4^+$ T cells, which direct the activity of effector inflammatory macrophages. At an early stage, $CD4^+$ T cells are polarized into Th1 and Th17 cells. The latter are most important in the early stages. The cytokines

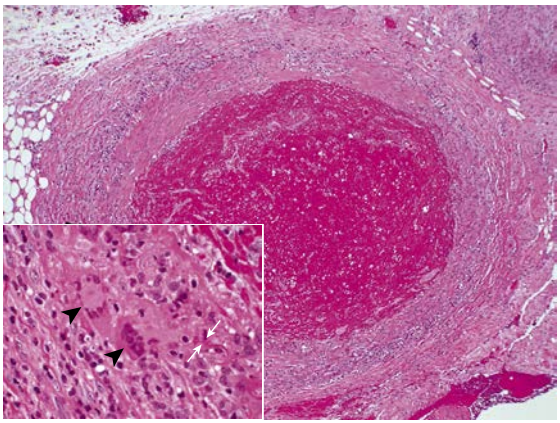


Fig. 9.15 In temporal arteritis the lumen may be occluded by a thrombus. There is intimal proliferation and inflammatory cell infiltration of the media with multinucleated giant cells *(inset, arrowheads)* in relation to the internal elastic lamina.

activate macrophages (including IL-1β and IL-6), giant cells and smooth muscle cells, leading to vascular remodelling. No specific antigen has as yet been identified in this autoimmune condition but dysregulation of PD-1:PDL-1 checkpoint inhibitor control of dendritic cell: T-cell activation has been shown to underlie the condition (Zhang et al., 2017).

Microscopic examination shows extensive inflammatory cell infiltration, with lymphocytes, plasma cells, eosinophils and macrophages in the media (Fig. 9.15). Inflammation may be patchy with skip lesions (foci of vasculitis alternating with patches of normal vessels). Multinucleated giant cells are usually located near the fragmented internal elastic lamina. The lumen may be obliterated by a fibrin thrombus, which becomes organized and recanalized. The adventitia and the small periarterial vessels are often involved in the chronic inflammatory process. Steroid therapy does not always totally suppress the inflammatory process and prevent occlusion of the arteries with subsequent blindness in the contralateral eye.

Takayasu's disease. This condition histologically resembles temporal arteritis but the major vessels involved are those arising from the arch of the aorta and its primary branches. The condition can lead to ischaemia of brain, eyes, face, arms, heart and kidneys in varying combinations. Unlike temporal arteritis, Takayasu's disease usually occurs in patients under 50 years of age.

Medium vessel vasculitis

Polyarteritis nodosa. In polyarteritis nodosa medium- and small-sized arteries are focally involved in an inflammatory process, which leads to fibrinoid necrosis of the vessel wall and thrombosis. This leads to ischaemia and focal infarction in the heart, central nervous system, kidney and muscle. Polyarteritis nodosa may unusually involve the ophthalmic artery, the central retinal artery and, in addition, the retinal and choroidal blood vessels; subretinal exudation of plasma leads to retinal detachment in this condition.

Kawasaki's disease. This is a disease in infants affecting the main aortic branches, particularly the coronary artery. Ophthalmic involvement is usually in the form of anterior uveitis or conjunctival injection.

Small vessel vasculitis

ANCA associated. Antineutrophil cytoplasmic antibodies (ANCA) are antibodies against antigens found in the cytoplasmic granules of neutrophils and monocytes. A perinuclear pattern of staining (p-ANCA) is seen in microscopic polyarteritis and about 50% of cases of eosinophilic granulomatosis with polyangiitis (formerly Churg–Strauss syndrome). Cytoplasmic staining (c-ANCA) is seen in granulomatosis with polyangiitis (GPA, formerly Wegener's granulomatosis).

On histological examination, GPA is characterized by necrosis, granulomatous inflammation and small vessel vasculitis. Ocular manifestations of GPA are more common and include scleritis, corneoscleral ulceration or an orbital mass. Ocular findings may be part of the generalized systemic disease or its limited form. Generalized disease classically presents with renal, lung, upper respiratory tract and paranasal sinus involvement. The limited form manifests upper respiratory and lung disease without kidney involvement. c-ANCA is positive in over 90% of patients with generalized Wegener's granulomatosis with the majority of the antibodies directed against proteinase 3(PR3) and a smaller number against myeloperoxidase (MPO). c-ANCA is only positive in 60% of those with the limited form.

Ocular manifestations of microscopic polyangiitis and eosinophilic granulomatosis with polyangiitis occur but are less common.

Immune complex mediated

The best known example of this is IgA vasculitis (Henoch-Schönlein purpura). The systemic manifestations of arthritis, intestinal and renal involvement. Ocular manifestations are rare and include keratitis, anterior uveitis and ocular ischaemia.

Systemic lupus erythematosus

In this multisystem autoimmune disease, antibodies to blood constituents such as leucocyte DNA are responsible for anaemia, thrombocytopaenia and leukopaenia. Phospholipid antibody levels are raised and the consequent

disturbances of coagulation explain the occlusive vasculopathy and haemorrhagic diathesis. In addition, DNA–antiDNA antibody complexes are formed in response to the release of DNA from dying cells and these initiate a type III hypersensitivity reaction. Small-vessel vasculitis may occur in association with systemic lupus erythematosus and is responsible for tissue damage in the heart, lungs, kidney, brain and skin.

Ocular disease is rare but appears as a lupus retinopathy characterized by the presence of retinal microinfarcts; in more severe forms there is occlusive disease of the central retinal artery and vein with haemorrhagic infarction. Choroidopathy is uncommon as a clinical manifestation but may be apparent on fluorescein angiography.

Metabolic disease

The eyes and extraocular tissues are involved in a number of systemic metabolic diseases and the ophthalmic complications may often be serious.

Diabetes

Diabetic retinopathy is predominantly a microvascular disease in which capillary occlusion and retinal ischaemia are the major features (see section on degenerative vascular disease, p. 512). The fundamental abnormality in the smaller vessels is multilayering of the basement membrane and degeneration of the endothelial cells and the pericytes induced by hyperglycaemia. However, microvascular complications continue to develop even with tight glycaemic control. The causes of this continued stress are considered to be multifactorial and include advanced glycation end products, mitochondrial damage, oxidative stress and epigenetic effects such as DNA methylation and histone modification (see Chapter 4).

Once the initial vascular damage occurs, the thickened, occluded capillaries in turn lead to retinal hypoxia, which becomes progressively worse with increasingly large areas of retinal tissue involved. This alters the balance of various growth factors, including an increase in vascular endothelial growth factor (VEGF) and placental growth factor (PlGF), which are both angiogenic, with reduction in pigment epithelium-derived growth factor (PEDF) which is anti-angiogenic. Untreated retinal ischaemia then leads to proliferative diabetic retinopathy (PDR), which leads to haemorrhage and retinal damage, the immediate cause of blindness.

In addition to vascular changes (see Video 9.1), diabetes may affect other tissues in the eye, with histological features that include:

- vacuolation of the iris pigment epithelium
- thickening of the basement membranes of the ciliary processes
- cataract (there are no specific features).

The macula is particularly at risk from oedema, due to a combination of retina capillary leakage and dysfunction of fluid transport across the retinal pigment epithelium. Visual loss resulting from macular oedema is more common but is slowly progressive.

Thyroid eye disease

Endocrine exophthalmos may present with unilateral (15%) or bilateral proptosis and limitation of ocular movement in the absence of other clinical signs of hyperthyroidism. In addition to the clinical signs, the diagnosis is usually made by abnormally high values of triiodothyronine (T_3) and thyroxine (T_4), and low values of thyroid-stimulating hormone. The demonstration of uniform swelling of the extraocular muscles on computed tomography or orbital ultrasonography confirms endocrine exophthalmos and is present in a high proportion of cases (up to 85%) in the absence of overt proptosis. This is an immune-mediated disease for which orbital fibroblasts appear to be the main target. In endocrine exophthalmos, the fibroblasts have increased numbers of insulin-like growth factor 1 receptors. Binding of these receptors can increase extracellular matrix production, activate T cells and macrophages and, through cytokine production, initiates and propagates an immune cascade.

This correlates with the findings in extraocular muscle of perivascular lymphocytic infiltration (lymphorrhage) with accumulation of glycosaminoglycans within and around muscle fibres (Fig. 9.16). As the disease progresses, there is replacement fibrosis between the muscle fibres. Study of the nerves at the orbital apex reveals loss of larger axons in the motor nerves; this is attributed to compression by the swollen muscle and explains the limitation of movement (ophthalmoplegia).

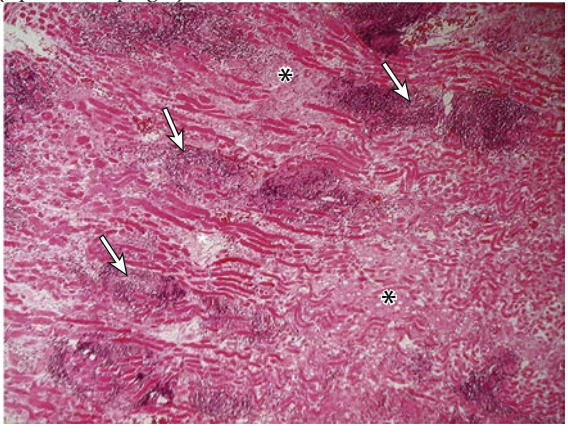


Fig. 9.16 Endocrine exophthalmos. The striated muscle fibres in the extraocular muscle are separated by lipid and mucopolysaccharide *(*)*. Throughout the muscle there are collections of lymphocytes *(arrows)*.

Disorders of amino acid metabolism

Homocystinuria

The biochemical abnormality in homocystinuria is a reduction in levels of cystathione β-synthetase. Patients suffering from this condition are at risk during surgery and anaesthesia because of a tendency to thromboembolic disease, which can be fatal. Dislocation of the lens (inferiorly and somewhat posteriorly) is the result of an acquired metabolic abnormality of the zonular fibres. Histological examination of the zonular fibres reveals deposition of a thick band of periodic acid–Schiff-positive material on the inner surface of the ciliary processes and the pars plana.

Cystinosis

The biochemical disturbance occurs in the lysosomal membrane transport of cystine, which is continuously released in the lysosomes in the degradation of protein. In cystinosis, the amino acid is trapped within the lysosome because of a defect in the transport system. The ocular manifestations, i.e. the accumulation of birefringent cystine crystals in conjunctiva, cornea, choroid, pigment epithelium and retina, have long been recognized. Alcohol fixation (100%) is necessary for histological verification of the presence of cystine crystals in a conjunctival biopsy.

Ageing, degeneration and dystrophies

Ageing can be difficult to define but in general there is overlap in the conditions to be described here since a gradual decline in tissue function (degeneration) is often accompanied by failing nutritive resource to the tissues (dystrophy) as part of the ageing process.

Ageing

The two main theories of ageing, which need not be mutually exclusive, are the programmed and error theories. The programmed theory suggests that we are genetically programmed to age and that this will differ between individuals. These intrinsic controlling factors include expression of ageing genes in mitochondria and loss of cells' ability to divide because of sequential telomeric shortening during life. Telomeres are a region of repetitive nucleotide sequence at the end of the chromosome arms that protects the chromosome from progressive shortening, which naturally occurs during cell division. Telomeres consist of approximately 8000 base pairs at birth and around 3000 base pairs as people age. Cells can normally only divide between 50 and 70 times, with the telomeres getting progressively shorter before the cells become senescent, die or sustain genetic damage that causes neoplastic transformation. In addition to genetic programming, hormonal regulation of ageing occurs through the insulin/IGF-1 signalling pathway. The immune system is also programmed to decline over time and dysregulated immune responses have been linked with diseases of ageing (e.g. Alzheimer's disease).

The error theory suggests that damage is caused by the additive effects of sublethal injury by a variety of injurious agents, e.g. diet, smoking, UV light, etc. Progressive wear and tear is caused by the production of tissue free radicals induced by damage from these extrinsic factors and cross-linked proteins that accumulate in cells. Both amyloid plaques in Alzheimer's and intracellular Tau protein aggregates are characteristic of ageing in Alzheimer's disease. Recent studies have shown that retinal changes can be predictive of these ageing changes. Ageing changes occur in all the structures of the eye and cause varying effects. The effects of ageing and the eye are summarized in Box 9.8.

Degeneration

Degeneration of tissues in pathological terms commonly involves connective tissue components such as collagen, elastin and proteoglycans, but probably represents progressive failure, dysfunction and death of parenchymal tissue cells, which are responsible for the continuous production of these elements. Degenerations are described in terms of their composition and their location.

Composition/type of degeneration

There are many types of degeneration, which vary in terms of their biochemical composition, probably as a reflection of the initial insult that triggered the process.

Degeneration can affect many tissues (systemic) or be localized to one site, for instance in relation to a previous injury or insult (the withered arm of birth injury is an example). The eye can be the preferred site for certain degenerative processes.

Hyalinization. This describes the replacement of normal cells by an acellular, almost transparent, matrix (which consists of collagens and glycoproteins) that imparts a glassy appearance to the tissues on microscopy. Hyalinization is typically seen in the eye and kidney in the walls of small blood vessels in ageing, benign hypertension and diabetes. Leakage of plasma into the vessel wall, owing to breakdown of the normal endothelial barrier, is thought to be one cause of hyalinization.

Fatty degeneration. The most innocuous form of fatty infiltration in tissue is seen in the peripheral corneal stroma as part of the normal ageing process and is described as arcus senilis. After prolonged inflammation followed by

Box 9.8 **Ageing and the eye**

Ageing changes occur in all structures of the eye with variable effects on vision.

Eyelid and lacrimal system

Skin thinning and laxity

- Redundant skin folds and wrinkles
- Ectropian or entropian
- Blepharoptosis and brow prolapse
- Orbital fat prolapse
- Tearing (resulting from punctual eversion)

Atrophy of lacrimal gland

- Dry eye

Cornea

- Changes in curvature, loss of lustre and sensitivity
- Arcus senilis
- Hassall–Henle bodies at periphery
- Krukenberg's spindle (uveal pigment in endothelium)
- Decrease in corneal endothelium

Trabecular meshwork and uvea

- Increased pigmentation of trabecular meshwork
- Increased resistance to aqueous outflow
- Decreased pupil size and reactivity
- Fibrosis of ciliary muscle (contributing to presbyopia)

Lens

- Increased absorption of blue light due to accumulation of yellow pigments
- Nuclear sclerosis (contributing to presbyopia)
- Loss of elasticity of zonules (contributing to presbyopia)

Vitreous

- Condensation of vitreous gel
- Liquefaction of vitreous gel with lacunae
- Detachment of posterior vitreous face

Retina and retinal pigment epithelium

- Neuronal cell loss and degeneration
- Thickening of basement membrane
- Atrophy of retinal pigment epithelium with decrease in melanin and increase in lipofuscin
- Age-related macular degeneration (see Box 9.12)

corneal vascularization, plasma lipids leak from the blood vessels and are deposited in the stroma (Fig. 9.17).

Deposition of fat (neutral lipids and cholesterol) in the intima of medium and large muscular arteries (atheroma) is followed by thrombosis (see previously in the section on vascular disease). Lipids within clumps of macrophages in the dermis of the eyelid (xanthelasma) are usually a feature of ageing, but hypercholesterolaemia must be excluded.

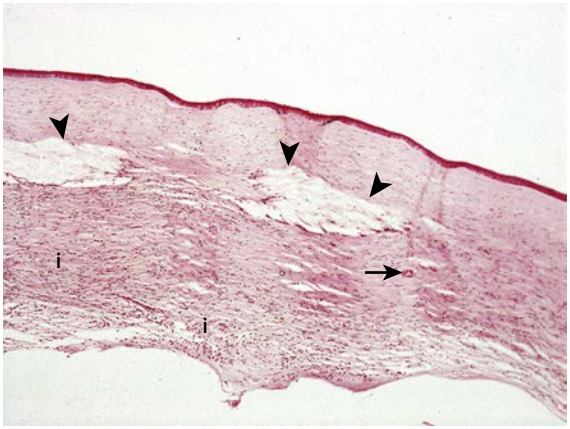


Fig. 9.17 At the end stage of corneal inflammatory disease the stroma is invaded by blood vessels *(arrow)* that leak lipid in the form of cholesterol crystals *(arrowheads)* leading to lipid keratopathy. There is an intense inflammatory infiltrate *(i)*.

Elastic fibre degeneration. Elastic fibres can be visualized with special stains (such as orcein). They appear as fine strands in tissues such as skin and blood vessels, and zonule of the lens. The constituent protein, elastin, is arranged in coils, imparting elasticity to the strand (see Chapter 4). 'Elastotic degeneration' in skin is frequently the result of chronic sun exposure, which induces defective fibroblast function and an altered elastic matrix, which has poor elasticity. Reduced skin elasticity is also seen in pseudoxanthoma elasticum. In this condition, ruptures in Bruch's membrane expose the choroid (angioid streaks).

Pinguecula and pterygium. These are types of degeneration specific to the ocular surface in which UV light damage plays a role. In the conjunctiva, deposition of elastic-like material causes thickening and formation of nodules on the bulbar conjunctiva (pinguecula). At the limbus, foci of elastotic degeneration form in the interpalpebral fissure and encroach on the cornea as a wing-shaped wedge (pterygium) (Fig. 9.18). Various changes, including dysplasia and carcinoma, may occur in the epithelium overlying the elastotic tissue. Growth hormone releasing hormone may be involved in the development of pterygia.

Calcification. Calcium is deposited in both healthy and diseased tissue as hydroxyapatite crystals ($Ca_{10}(PO_4)_6(OH)_2$). In hypercalcaemic states such as hyperparathyroidism, hypervitaminosis D and excessive bone resorption from skeletal metastases, calcium is deposited in healthy tissue such as the kidney and the conjunctiva; this process is called metastatic calcification.

In contrast, calcium can be deposited in hyalinized connective tissue (blood vessels) or necrotic tissue (such

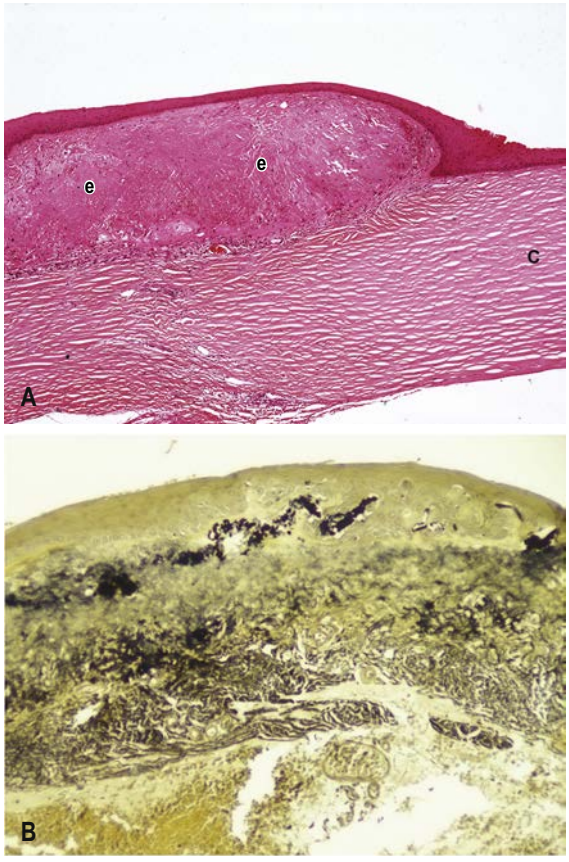


Fig. 9.18 (**A**) A pterygium encroaching on the cornea *(C)*. It consists of dense deposition of degenerating elastin *(e)*. (**B**) The degenerate elastin stains *black* using a special stain (Elastic Van Gieson).

as post-tuberculous scars in lung, atheromatous plaques, necrotic tumour tissue in a retinoblastoma) in a normocalcaemic state; this is referred to as dystrophic calcification. Calcification of ocular tissues also occurs in the end-stage phthisical eye usually within the metaplastic fibrous tissue derived from the retinal pigment epithelium. This may sometimes ossify.

Amyloid

Amyloid is insoluble protein deposited in tissues, particularly around blood vessels and in basement membranes. Amyloid is formed by aggregation of soluble extracellular protein into fibrils and is defined mainly by its tertiary and quaternary structure. Many proteins can aggregate into the general structure of amyloid. In haematoxylin and eosin-stained sections, amyloid has a homogeneous pink appearance; staining with Congo red, followed by examination in polarized light, reveals apple green birefringence (Fig. 9.19). The composition of amyloid varies and the effects depend on whether it is systemic or localized.

Systemic amyloid deposition.

- Associated with monoclonal plasma cell proliferation, e.g. myeloma, Waldenström's macroglobulinaemia. In this case, the amyloid is light-chain-derived (AL) from fragments of immunoglobulin.
- Associated with chronic inflammation, e.g. rheumatoid arthritis; genetically inherited familial Mediterranean fever. The amyloid is derived from soluble serum amyloid protein (SAA), an acute-phase reactant in many inflammatory conditions.

Localized amyloid deposition.

- Amyloid derived from polypeptide hormones may be deposited in endocrine tumours, e.g. medullary carcinoma of the thyroid.
- Amyloid derived from prealbumin may be deposited in the heart, brain and joints in the elderly. Cerebral deposits of amyloid are important in Alzheimer's disease where amyloid-β is the characteristic protein. Amyloid protein is also a component of drusen in macular degeneration.

Amyloid may be observed as a solitary nodule within the eyelid, the orbit or in the conjunctiva in the absence of systemic disease. Amyloid is deposited in the choroid and vitreous in systemic amyloidosis. In the cornea, amyloid deposition (Fig. 9.19) is the characteristic feature of lattice dystrophy.

The exfoliation syndrome (pseudoexfoliation syndrome)

Pseudoexfoliation syndrome is an ageing-related systemic disease in which granular amorphous eosinophilic material accumulates throughout the body. In the eye, this material forms fluffy white deposits on the surface of the lens, the ciliary processes, the iris surface and the inner surface of the trabecular meshwork. Involvement of the outflow system leads to secondary open-angle glaucoma. The exudates contain basement membrane proteins and various matrix enzymes but the true pathogenesis of the exfoliation syndrome is not fully understood although it is thought to have a genetic component due to increased prevalence in certain populations (e.g. Scandinavians).

Degenerative vascular disease

Degenerative vascular disease (atheroma, atherosclerosis) is a major cause of morbidity and mortality in industrialized nations and is also increased in metabolic disease such as diabetes. The degenerative process in the vessel wall can take on various forms which include:

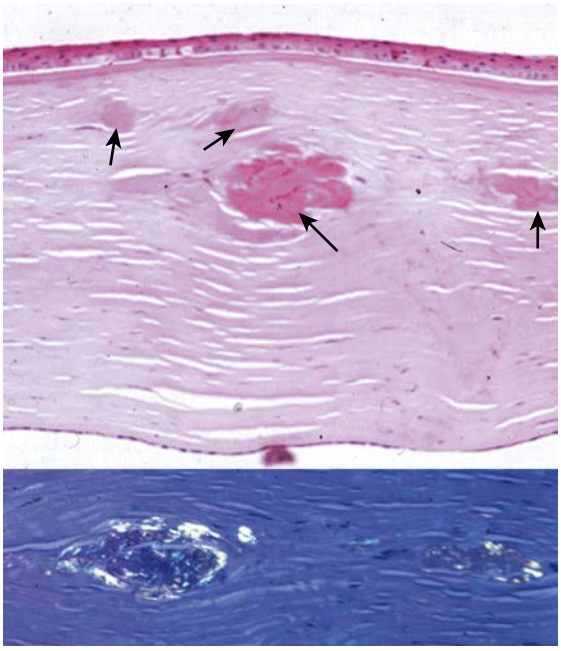


Fig. 9.19 (**A**) In lattice dystrophy of the cornea, stromal deposits *(arrows)* have the staining characteristics of amyloid and exhibit apple green birefringence when this Congo red-stained section is observed in polarized light (**B**).

Hyalinization. In ageing, hypertension and diabetes, the walls of arterioles and venules become thickened by deposition of collagen (hyalinization) by fibroblasts with a loss of the normal smooth muscle layer (Fig. 9.20). Consequently, there is a reduced capacity to respond to metabolic demand while narrowing of the lumen reduces the overall perfusion of the tissues. Hypertension can be superimposed on both diabetic and senile degenerative vasculopathy, adding a contractile (vasospastic) component to the disease.

Vaso-occlusive disease. Retinal vaso-occlusive disease may be macro- or microvascular. Macrovascular occlusions involve the central retinal artery or vein or its major branches. Microvascular occlusion occurs at the arteriolar level, causing non-perfusion of tissue and ischaemia.

Microvascular occlusion. The following abnormalities may be found in conditions in which focal capillary occlusion causes patchy interference with blood flow within the retina such as hypertension, diabetes, AIDS, radiation vasculopathy and the vasculitides:

- *Microinfarction:* fluffy white swellings (cotton-wool spots) in the retina on ophthalmoscopy are representations of the swollen ends of interrupted axons. Build-up of axoplasmic flow occurs at the edge of the area previously supplied by the occluded vessel (Fig. 9.21).

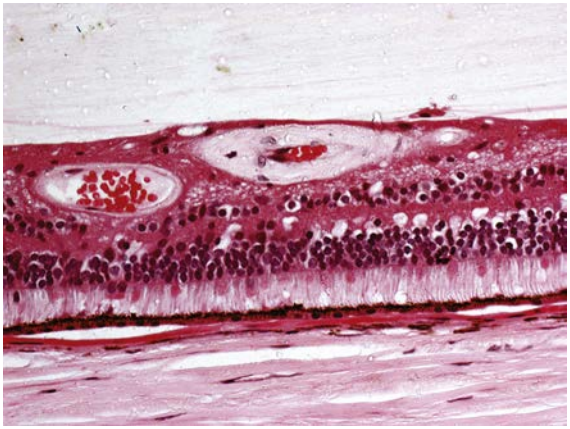


Fig. 9.20 A hyalinized retinal vessel with a thickened wall and narrowed lumen in a case of long-standing hypertension.

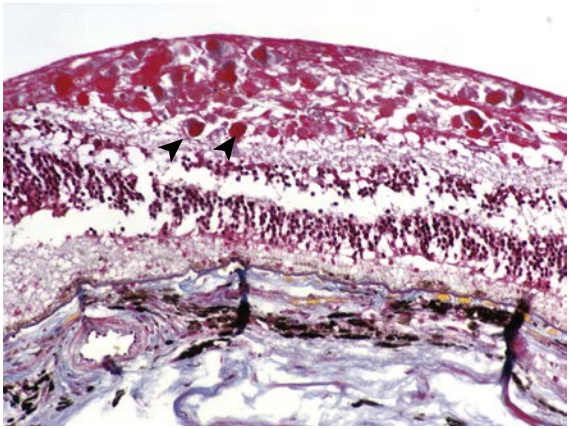


Fig. 9.21 A microinfarct in the retina is seen as a swollen sector of disrupted axons. Smudgy eosinophilic structures (cytoid bodies) represent the swollen ends of axons *(arrowheads)*. The infarct mainly involves the nerve fibre layer (Bodian stain).

- *Hard exudates:* underperfusion of the vascular bed and damage to the endothelium of the deep capillaries leads to plasma leakage into the outer plexiform layer. Clinically this exudation is yellow and well circumscribed (Fig. 9.22). Histologically, 'hard' exudates are eosinophilic masses and these contain foamy macrophages with cytoplasmic lipid inclusions (Fig. 9.22).
- *Microaneurysms:* another effect of ischaemia on the capillary is weakening of the wall by necrosis of the supporting cell (the pericyte) in diabetes and the endothelial cell in central retinal vein occlusion. The ensuing small bulges or blowouts in the capillary wall are referred to as microaneurysms. With time, microaneurysms become filled by basement membrane deposits and consequently may disappear on fluorescein angiography.

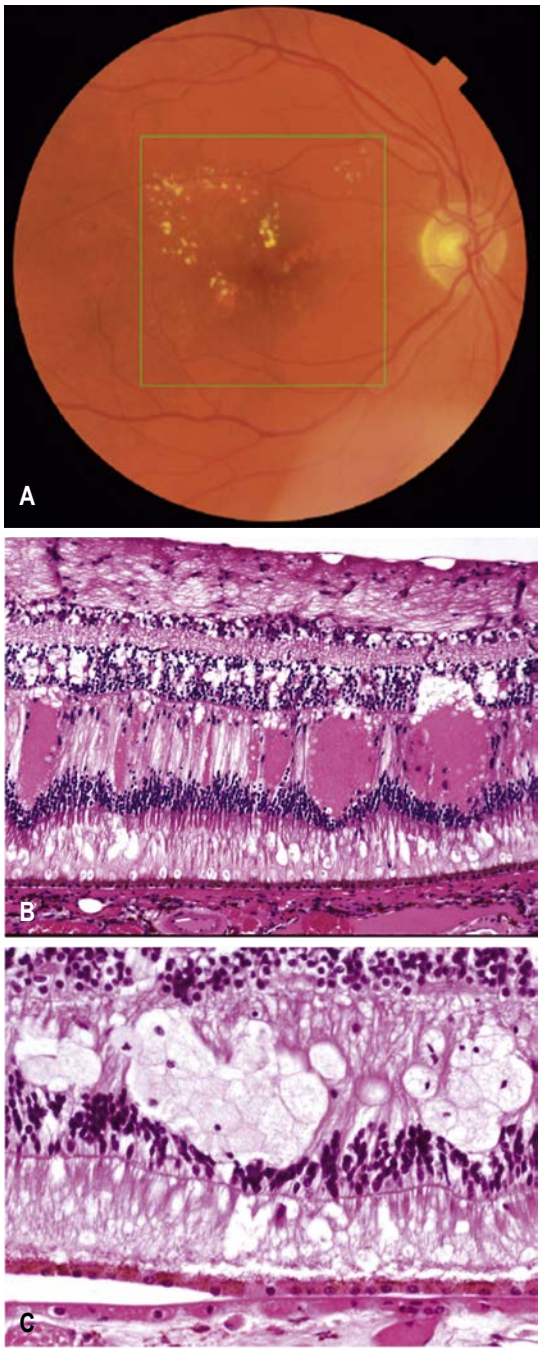


Fig. 9.22 (**A**) Fundoscopic image of diabetic macular oedema with hard exudates surrounding the macular area. (**B**) Hard exudates are the result of leakage of plasma through the capillary endothelium. The exudates occur in the outer plexiform layer and initially consist of a proteinaceous exudate. (**C**) Later, foamy macrophages are attracted to the deposit in an attempt to remove it. (**A** courtesy Dr Graeme Williams, Glasgow.)

- *Haemorrhage:* breakdown of the vessel wall leads to leakage of red cells and can take several forms in the retina:
  - flame haemorrhages follow rupture of a small arteriole so that blood tracks into the nerve fibre layer
  - dot haemorrhages follow rupture of capillaries in the outer plexiform layer; these are smaller and more circumscribed than flame haemorrhages
  - blot haemorrhages are larger than dot haemorrhages and represent bleeding from capillaries with tracking between the photoreceptors and the RPE.
- *Neovascularization:* newly formed vessels grow from the venous side of the capillary bed within an area of arteriolar non-perfusion; this change represents a response to ischaemia within the retina. These vessels leak on fluorescein angiography and they occur in an eye that will progress to vasoproliferative retinopathy. The new vessels arise in the prevenular capillaries and in the walls of hyalinized venules and proliferate within and on the surface of the retina. If the vitreous is detached, the fibrovascular tissue grows on the inner surface of the retina; the membrane contracts, leading to retinal detachment. Fibrovascular proliferation within the attached vitreous leads to haemorrhage and further formation of traction bands. The diffusion of vasoformative factors from the vitreous through the posterior and anterior chambers induces blood vessel formation on the iris surface (rubeosis iridis) and the inner surface of the trabecular meshwork, sealing off the angle and causing secondary (neovascular) glaucoma. Vasoproliferation also occurs in retinopathy of prematurity (Box 9.9), which is intricately associated with relative retinal ischaemia.

Macrovascular occlusion. Macrovascular occlusion refers to obstruction of vessels of diameters equal to or greater than a medium-sized arteriole. It includes thrombotic occlusion of the central retinal artery, in

Box 9.9 **Retinopathy of prematurity**

Normal retinal vascularization

Begins at approximately 16 weeks
Extends from optic nerve and proceeds peripherally
Reaches ora serrata 32 weeks (nasal), 36–40 weeks temporoal

Preterm birth

Retinal vasculature incomplete and development disrupted because of several factors: hypoxia, hypotension and hyperoxia.
Return to normal oxygenation leads to relative hypoxia with release of angiogenic factors
Rapid and disorganized proliferation of vessels within the retina and vitreous
Haemorrhage with secondary inflammation and fibrosis can lead to traction and retinal detachment.

association with systemic conditions such as hypertension and atherosclerosis.

- *Hypertension:* the classic textbook appearances of the retina in accelerated (malignant) hypertension, i.e. haemorrhage, exudates and papilloedema, are not normally seen with antihypertensive therapy. In mild cases of hypertensive retinopathy, the vessels are described as 'copper-wire' or 'silver-wire' as a result of hyalinization. In more advanced disease, narrowing of the blood column followed spasm of the vessels, produces ischaemic damage to the endothelium distal to the constriction. Swelling and degeneration of the endothelium is followed by leakage of fibrin into the vessel wall and further narrowing of the lumen. Fibrinoid necrosis, characteristic of hypertensive renal vasculopathy, may be found in the walls of choroidal and retinal vessels in uncontrolled hypertension. If the choriocapillaris is occluded by fibrinoid necrosis, there is exudation beneath a necrotic retinal pigment epithelium, leaving small areas of depigmentation (Elschnig's spots).
- *Central retinal artery occlusion:* the central retinal artery is an end-artery and obstruction of flow leads immediately to blindness. The normally transparent retina becomes opaque because of cessation of axoplasmic flow in the retinal ganglion cells, thus preventing transmission of the red reflex created by the choroidal vasculature except at the macula where the choriocapillaris is visible due to reduction in ganglion cell fibres at the fovea (see Chapter 1, p. 39) (the classic cherry-red spot) (Fig. 9.23). Central retinal artery occlusion may be the result of thrombosis in a degenerate central retinal artery but is more often the result of an embolus, typically from a mural thrombus on the endocardium after a myocardial infarction, an atheromatous plaque in the carotid artery, or the heart valves in subacute bacterial endocarditis. After total infarction, none of the inner retinal tissue survives and vasoformative factors are not released; thus, rubeotic (neovascular) glaucoma occurs as a complication in fewer than 5% of cases, unlike central retinal vein occlusion in which up to 50% of cases may progress to glaucoma. Emboli may originate from many sources (Box 9.10).
- *Central retinal vein occlusion:* the characteristic difference between central retinal artery and vein occlusion on fundoscopy is the presence of extensive haemorrhages within the retina in the latter. There may also be some recovery of vision in venous occlusion, unlike retinal occlusion, but in a significant number of patients, rubeotic (neovascular) glaucoma develops within 3 months (Fig. 9.24). Preretinal neovascularization and glaucoma are most frequent where there is extensive retinal ischaemia and may be prevented by peripheral retinal photoablation if the ischaemia is identified by fluorescein angiography. Frank neovascular glaucoma is usually not amenable to treatment, which is usually aimed at palliation of symptoms.



Fig. 9.23 Fundoscopic image of central retinal artery occlusion with cilioretinal sparing. The retina is pale except in the distribution of the cilioretinal artery *(inside green box)*. The cilioretinal artery is a branch of the short posterior ciliary artery (itself a separate branch of the ophthalmic artery). The macula is also evident as a red dot due to the lack of nerve fibre layer in this area. (Courtesy Dr Graeme Williams, Glasgow.)

Central retinal vein occlusion is also seen in a younger age group, where it may be extremely difficult to differentiate from retinal vasculitis (see earlier). This form of central retinal vein occlusion has a better prognosis if it is the result of inflammatory/immune mediated retinal vasculitis, and immunotherapy with steroids or cyclosporin A may prevent its progression to neovascular glaucoma. The pathogenesis is not entirely clear (Box 9.11).

Cataract

Cataract is part of the ageing process, which starts with the hardening of the lens, which prevents accommodation and near vision for reading (presbyopia, see Chapter 1, p. 32). Secondary cataract also occurs after any insult to the lens (see Chapter 4, p. 234).

Cataract associated with ageing. The lens crystallins (see Chapter 4, p. 234) break down to albuminoids, partly as an age-related process and partly in response to exposure to light, particularly of ultraviolet/blue wavelength. The amino acids (e.g. tyrosine) that are released are converted to adrenaline and melanin, so that lens pigmentation progresses from yellow to brown (brunescent cataract) to black (cataracta nigra) (see Chapter 4, p. 242).

Box 9.10 **Embolism**

An embolus is any abnormal mass of matter carried in the bloodstream and large enough to occlude a vessel. The various types of embolism are listed below:

- *Thrombotic:* thrombus formation in the leg and pelvic veins is the principal cause of pulmonary embolism and intraoperative and postoperative death. In the retinal or choroidal vessels, an embolus can originate from a thrombus on the mitral and aortic valves or from ulcerating atheromatous plaques in the aorta or carotid arteries. Another source may be a mural thrombus in the left ventricle.
- *Air embolism* occurs when negative pressure in the neck veins follows thyroid surgery or when fluid or air is forced into the venous circulation during a blood transfusion. Frothing of the blood in the right ventricle interferes with ventricular pumping and is fatal.
- *Tumour emboli* are usually small and not visible in the retinal circulation; larger metastases occur in the choroid.
- *Fat and marrow embolization* occurring after severe trauma to the limbs and trunk is accompanied by multiple fractures. Purpuric spots are seen on the upper thorax and small haemorrhages are found in the retina. A severe form may rarely occur as Purtscher's retinopathy in which florid embolization of multiple small vessels occurs.
- *Emboli from atheromatous aortic or carotid plaques* may consist of cholesterol/calcified tissue/fibrin/platelets, and can be seen migrating through the retinal circulation by ophthalmoscopy or video fluorescein angiography.
- *Septic emboli* were described in the retina by Roth in 1905, when subacute bacterial endocarditis was common. The typical Roth's spot has a white centre and red surround and is thought to be the result of vascular damage from an impacted mass of white cells and bacteria in a retinal arteriole. Roth's spots can be caused by leukaemia, diabetes, ischaemia and hypertensive retinopathy where the white centre is a result of tumour cells, platelets or fibrin.
- *Amniotic fluid embolism* is a complication of parturition, particularly when manipulation of the fetus is required. Release of amniotic fluid, vernix, hairs and fetal squames into the maternal circulation is commonly fatal.

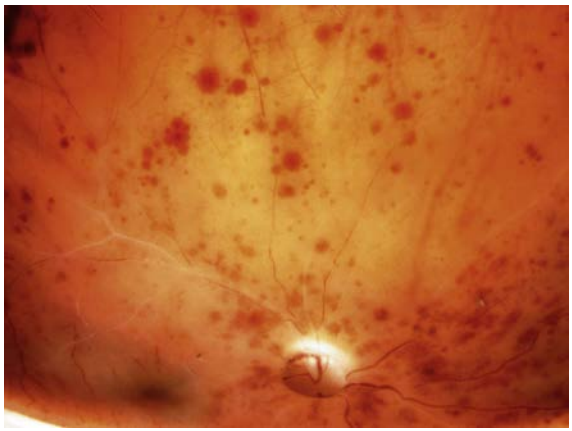


Fig. 9.24 Macroscopic photograph of the fundus in central retinal vein occlusion. There are numerous haemorrhages within the retina. The optic disc is cupped as a result of neovascular glaucoma.

Box 9.11 **Pathogenesis of central retinal vein occlusion**

If one or more factors in Virchow's triad are fulfilled, this can contribute to central retinal vein occlusion:

1. Abnormality of blood flow
   - Venous narrowing in lamina cribrosa (radius 50% of prelaminar region; required to maintain pressure in retinal capillary bed against intraocular pressure especially if raised) predisposes to turbulence.
   - Abrupt fall in systemic blood pressure and hence pressure in central retinal artery
2. Abnormality of vessel wall
   - Turbulent flow because of the anatomical arrangement of the central retinal vein predisposes to turbulence and endothelial damage.
3. Abnormality of blood constituents
   - Increased blood viscosity (e.g. hyperglobulinaemia, polycythaemia, oral contraceptives, etc.)

Secondary cataract. The biochemical requirements for the maintenance of transparency of the tissue are discussed in Chapter 4, p. 234. Any metabolic disturbance, such as diabetes or hypocalcaemia, may potentially alter this microenvironment and can lead to lens opacification. The epithelial cells in the lens are also particularly sensitive to ionizing radiation and mechanical trauma; breakdown of transport mechanisms in the membranes of the lens fibre cells and the epithelium promotes ionic imbalance and fluid inflow, causing disorganization of the lens proteins and loss of transparency.

Glaucoma

Although malformation of the outflow system can lead to congenital or juvenile glaucoma (see Chapter 2, p. 133), the most common forms of glaucoma are the result of degenerative disease and may be subclassified into primary or secondary types.

Primary open-angle glaucoma. Primary open-angle glaucoma (POAG) is a disease that increases in incidence with age and is one of the most common causes of blindness. Although mutations have been identified in certain genes (myocilin, optineurin and TANK binding kinase 1), less than 10% of cases are caused by specific gene mutations and the remaining cases are polygenic in which the exact pathogenesis remains elusive.

In POAG the raised intraocular pressure is attributable to abnormal resistance of the outflow system; however, to date, no significant morphological abnormality has been demonstrated within the outflow system. Nonetheless, in POAG, obstruction to aqueous outflow develops progressively and the intraocular pressure gradually rises from the normal value of 18–23 mmHg to 25–35 mmHg.

The slow, progressive rise in intraocular pressure may be accompanied by occlusive disease in the posterior ciliary arteries, so that ischaemic optic atrophy may contribute to visual loss. Damage to the prelaminar optic nerve fibres may therefore be compounded by pressure-induced ischaemia in the capillary bed of the optic disk or to direct mechanical pressure preventing axoplasmic flow (see Chapter 1, p. 56) in the axons passing through the lamina cribrosa, across which a pressure differential builds up. Nerve fibre bundles passing into the optic nerve head above or below the horizontal line on the temporal side of the disk are selectively damaged and the prelaminar part of the nerve becomes atrophic (Fig. 9.25). Clinically, defects occur in the visual field (arcuate scotoma), but fibres from the macula, the papillomacular bundle, are spared. As the atrophy progresses, the cup in the optic nerve head is enlarged more extensively in the vertical plane than in the horizontal.

Primary closed-angle glaucoma. Primary closed-angle glaucoma is also the result of degenerative disease. With age, the lens thickens anteroposteriorly and in constitutively small eyes, such as in long-sighted individuals, the anterior surface of the lens displaces the pupillary part of the iris anteriorly, causing the anterior chamber to become shallow; as the angles become narrower, pressure builds up behind the iris and pushes the peripheral iris towards the trabecular meshwork. This initiates a vicious circle and pressure rises in the posterior chamber to reach levels of 40–80 mmHg. Because the pressure increase is acute, the effect on the prelaminar nerve fibres is to block axoplasmic flow so that the optic disk swells (papilloedema). In Asian populations, a form of chronic angle closure glaucoma occurs which presents much less acutely and often painlessly, but with severe retinal nerve fibre layer damage and blindness.

Secondary glaucoma. Secondary glaucoma is also of two types: secondary open-angle and secondary closed-angle. In secondary open-angle glaucoma the angle is obstructed by cells in inflammation (uveitis), haemorrhage or tumour cell infiltration, melanomacrophages following tumour necrosis or by lens matter when a degenerate lens capsule ruptures, flooding the anterior chamber with cortical lens matter and macrophages that have phagocytosed the lens cell fragments (Fig. 9.26).

In secondary closed-angle glaucoma, the chamber angle may be closed mechanically by anterior displacement of the lens, for instance by an intraocular tumour (e.g. uveal melanoma or retinoblastoma). A more common form of secondary closed-angle glaucoma occurs in uveitis, in which fibrin initiates adhesion formation between the peripheral iris and the trabecular meshwork (anterior synechiae), or the pupillary iris and the lens (posterior synechiae). Both can produce a rise in pressure because of obstruction to flow at the pupil (iris bombé) or

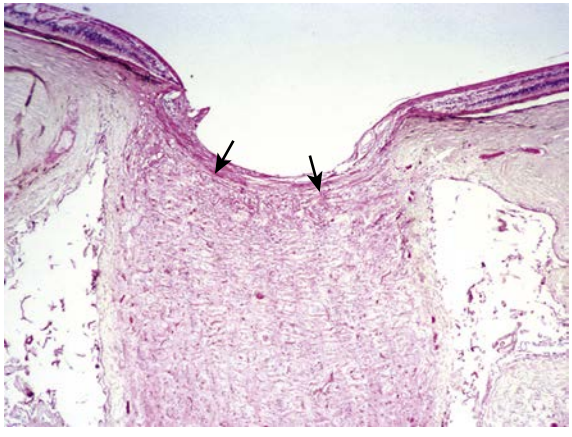


Fig. 9.25 The optic nerve is atrophic in advanced primary open-angle glaucoma and the lamina cribrosa *(arrows)* is bowed posteriorly, hence the optic disc is cupped.

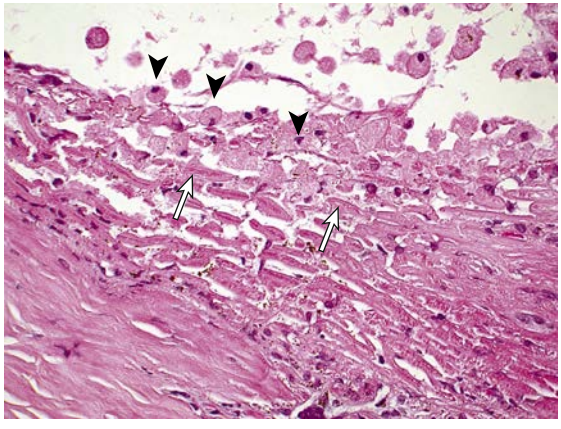


Fig. 9.26 In phacolytic glaucoma, swollen macrophages that have phagocytosed lens matter *(black arrowheads)* clog the trabecular meshwork *(white arrows)*.

in the angle. A special form of secondary glaucoma is neovascular (rubeotic) glaucoma caused by fibrovascular proliferation in the chamber angle; this is most commonly secondary to retinal ischaemia and the release of high levels of VEGF and other angiogenic factors (see earlier). Rubeotic fibrovascular proliferation produces adhesions between the iris and the trabecular meshwork, which lead to a painful high-pressure glaucoma that is particularly resistant to therapy.

Congenital glaucoma in infants and children is the result of malformation of the chamber angle and failure of development in the trabecular meshwork.

Age-related macular degeneration and disciform degeneration of the macula

Age-related macular degeneration (ARMD in Box 9.12) is a disease leading to severe visual loss and blindness in the elderly population. Its pathogenesis is poorly understood and is probably multifactorial. It appears to involve a complex interaction of metabolic, functional, genetic and environmental factors. Abnormalities are seen in the photoreceptors, retinal pigment epithelium (RPE), Bruch's membrane and choriocapillaries. The impairment of RPE cell function is an early and crucial event in the molecular

Box 9.12 Pathogenesis of age-related macular degeneration

The pathogenesis of age-related macular degeneration (AMD) is considered to be multifactorial, involving complex interactions of genetic, metabolic and environmental factors. At least four processes contribute to the disease:

Genetics

- Familial component to AMD
- Genetic variants in complement genes
  - Complement Factor H(CFH): The most common variant is a single nucleotide polymorphism resulting Y402H increases the risk of AMD (7× in homozygotes; 2–3× in heterozygotes). Other variants of this gene also increased risk of AMD. Common haplotype deletion of complement factor H related genes *CFHR1* and *CFHR3* are protective for AMD
  - Complement Factor I (CFI): A single nucleotide polymorphism G119R confers a 5 x higher risk of developing AMD.
  - Genetic variants in other complement factors including C2, C3, C9, Complement factor B and vitronectin have also been associated with the disease
- A promoter variant of the gene encoding high-temperature requirement protein A1 (HTRA1) is part of a mutant allele that causes increased HTRA1 expression and contributes to AMD.
- Other genes identified as potentially having a role include *ABCA4, ELOVL4, FIBL-6, APOE, SOD2.*

Impairment of RPE cell function

- RPE cells have an important role in retinal homeostasis supporting the blood retinal barrier, polarizing growth factor release and transporting ions, water and metabolic products from the subretinal space to the blood
- Changes that occur in retinal pigment epithelium (RPE) function include:
  - Mitochondrial DNA damage both direct and secondary to reactive oxygen species (ROS): this occurs during electron transport chain and can result in production of faulty electron transport proteins with further accumulation of ROS
  - Lipofuscin accumulates in the RPE. The RPE continuously phagocytoses and through lysosomes degrades the membranous discs of the rods and cones. Lipofuscin accumulation is a by-product of this degradation.
  - Amyloid β production is increased, which upregulates inflammatory pathway genes.
  - Enhanced tissue factor expression, which can promote inflammation and angiogenesis.
  - Lysosomal dysfunction with dysregulated autophagy (lysosomal mediated degradation). Autophagy is increased in RPE in ageing but dysfunctional in ARMD as degradation is impaired leading to accumulation of autophagosomes.

Sub-RPE deposit formation

- Sub-RPE deposits are commonly found in AMD
- Hard drusen: hyaline and well circumscribed; occur between the RPE basement membrane and Bruch's membrane
- Soft drusen: granular and vesicular and often more extensive; occur in same location as hard drusen
- Basal linear deposit occurs between the RPE basement membrane and RPE cell cytoplasm
- Local inflammation related to RPE dysfunction described previously and complement activation, may contribute to these deposits
- Photoreceptor cell death occurs overlying these deposits.

Choroidal neovascularization

- Increased production of extracellular matrix due to local inflammatory reaction may cause local ischaemia
- Inflammatory cells, including neutrophils and macrophages, may produce pro-angiogenic molecules such as VEGF, resulting in an imbalance in pro-angiogenic and anti-angiogenic molecules.

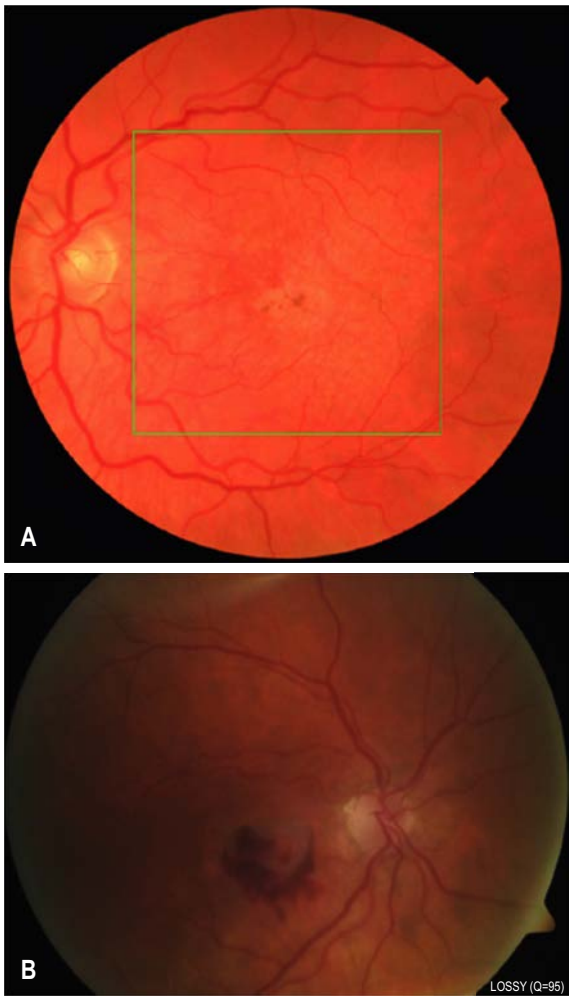


Fig. 9.27 (**A**) Fundoscopic image of dry age-related macular degeneration showing relatively well delineated area of hypopigmentation *(green box)* resulting from loss of photoreceptors and RPE. (**B**) Fundoscopic image of wet age-related macular degeneration showing area of RPE atrophy with adjacent haemorrhage caused by bleeding from fibrovascular proliferation. (Image courtesy Dr Graeme Williams, Glasgow.)

pathways leading to clinically relevant AMD changes (Fig. 9.27). As the RPE degenerates, there is irreversible degeneration of photoreceptors. The pathogenesis is described in Box 9.12 and pathological features in Fig. 9.28.

Dystrophies

In contrast to degeneration, dystrophies may occur at any age because they represent a disturbance of normal cellular functions (*dys:* altered; *trophy*: nutrition). Dystrophies encompass a range of disorders in which functional and morphological abnormalities appear in cells at various stages in life.

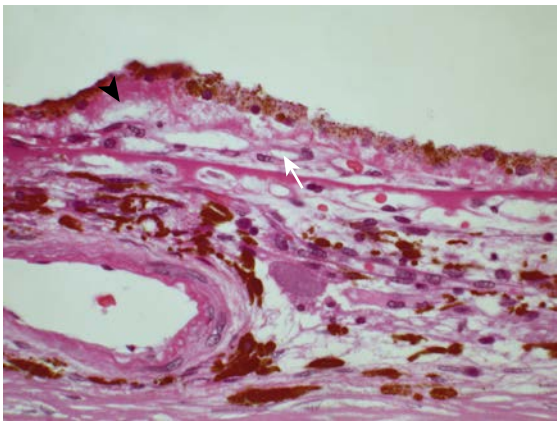


Fig. 9.28 Histology of early age-related macular degeneration showing the surviving retinal pigment epithelium overlies a basal linear deposit *(arrowhead)*, which attracts blood vessels *(arrow)*; this precedes the florid fibrovascular proliferation seen in disciform degeneration of the macula. The overlying retina is detached by artefact.

Retinal dystrophies

There are many disorders in which visual loss is the result of photoreceptor degeneration associated with patchy atrophy and proliferation in the RPE. A detailed account is outside the scope of this chapter and the reader is referred to the recommended reading.

Further reading is available online at https://expertconsult.inkling.com

A dystrophy may initially involve only the peripheral retina and progress later towards the macula or may primarily involve the macula, i.e. heredomacular degeneration. Over 100 gene mutations have been linked to one or more forms of retinal dystrophy. The same genetic defect can be associated with two or more phenotypically different forms of dystrophy, suggesting epigenetic effects on the disease manifestation (see Chapter 3, p. 149) (an updated list of the genetic abnormalities associated with inherited retinal degenerations (IRDs) can be obtained at www.ncbi.nlm.nih.gov/omim).

Three examples of retinal dystrophies are provided here to demonstrate aspects of the pathology.

Retinitis pigmentosa. This group of diseases mostly affects individuals in early adult life (for details of the genetic aspects see Chapter 3; p. 158). The first symptoms are night blindness and a progressive reduction in visual field from the periphery toward the posterior pole. At the end stage, retinal function is restricted to the central macular region

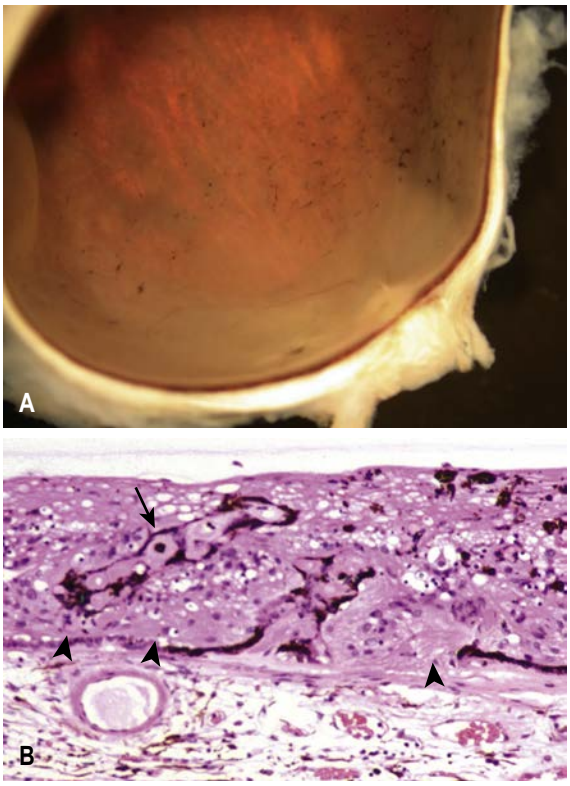


Fig. 9.29 (**A**) Macroscopic photograph of the retina in retinitis pigmentosa. The retina surrounding the macula and optic disc is thickened and pale. In the periphery there is delicate bone-spicule pigmentation. (**B**) On histological examination there is extensive atrophy of the photoreceptor layer, which is replaced by glial cells *(arrowheads)*. The retinal pigment epithelium migrates into the retina and forms clusters around hyalinized blood vessels *(arrow)*. This is the histological correlate of the bone-spicules seen in (**A**.)

(tunnel vision). Cataract is a common late complication. Fundoscopy reveals retinal atrophy, opacification and narrowing (hyalinization) of the retinal vessels and mixed fine and coarse strands of pigmentation (bone spicules); the macular region is spared until the final stages (Fig. 9.29).

On microscopy of advanced disease, the outer nuclear layer at the fovea appears as a single layer of cells with markedly stunted photoreceptors. Towards the periphery, the outer nuclear layer vanishes and is replaced by Müller cells, which fuse with the RPE. The RPE cells react by proliferation and migration into the retina to become distributed around the hyalinized vessels (Fig. 9.29), hence the bone spicules seen on fundoscopy.

Vitelliform dystrophy (Best's disease). This is an autosomal dominant heredomacular degeneration in which



Fig. 9.30 Fundus in Best's disease. There is a classic 'egg-yolk' lesion with satellite lesions. (From Hart Moss, M.D, Eyewiki. aao.org/Best_Disease.)

there is loss of central visual acuity associated with a disk of yellow tissue at the macula. Histologically there is a massive accumulation of lipofuscin in the RPE cells in association with atrophy of the photoreceptor layer of the retina. Mutations in the *BEST1* and *PRPH2* genes cause vitelliform macular dystrophy. *BEST1* encodes bestrophin, a protein channel which controls chloride ion movement. *PRPH2* encodes peripherin 2, which is essential for the normal function of photoreceptor cells in the retina (Fig. 9.30).

Stargardt's disease (fundus flavimaculatus). In this condition there is atrophy of the macula in association with the appearance of small yellow flecks. At the end stage, the outer layer of the retina is lost and the pigment epithelium is absent, and therefore the gliotic retina fuses with Bruch's membrane. At the earliest stages, the RPE is enlarged by accumulation of lipofuscin and melanin. The majority of cases of Stargardt's disease are caused by mutations in the *ABCA4* gene, which normally transports potentially toxic substances out of the photoreceptor cells (Fig. 9.31).

Corneal dystrophies

This term includes a large group of inherited conditions that cause bilateral, slowly progressive, corneal opacification and occur in the second, third and fourth decades of life. The conditions are classified as epithelial, stromal or endothelial, although the stromal diseases may involve the other layers.

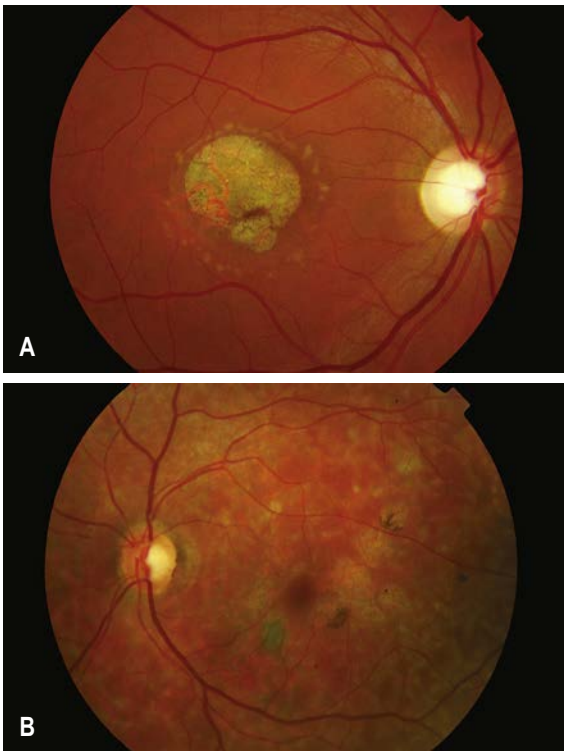


Fig. 9.31 Two different examples of the fundus Stargardt's disease (fundus flavimaculatus). There are yellow flecks in the fundus with a beaten-metal appearance at the macula owing to RPE atrophy. (Courtesy Noemi Lois.)

Epithelial and subepithelial dystrophies. These include a group of superficial dystrophies involving the epithelium and in some, the subepithelial (Bowman's) layer. The epithelium is unable to maintain normal replication and adhesion to Bowman's layer, which leads to an unstable epithelium. Various changes occur depending on the subtype. One of the more common, epithelial basement membrane dystrophy clinically shows map, dot or fingerprint-like patterns whilst on histology there may be multilamellar basal laminar material or intraepithelial cysts. The majority of cases have no documented inheritance.

Epithelial-stromal TGFβI dystrophies. The following dystrophies are associated with mutations in the transforming growth factor β–induced gene situated on chromosome 5q31. This encodes a protein that is expressed on the cell membrane of corneal epithelium and stromal keratocytes and plays a role in adhesion and wound healing. Mutations result in abnormal proteins, which may undergo transformation to deposits of non-fibrillar amyloid material. In this instance, the condition thus departs from the classic description of amyloid. Inheritance is autosomal dominant with complete penetrance.

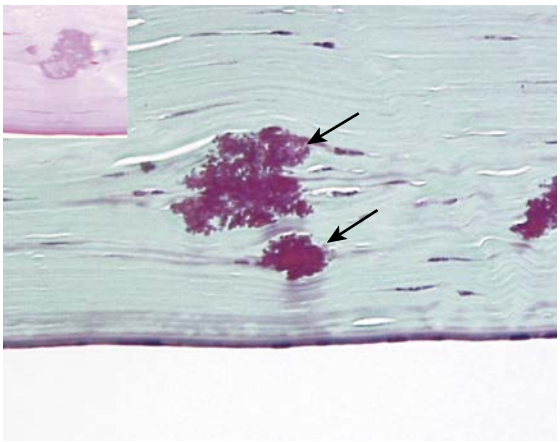


Fig. 9.32 In granular dystrophy of the cornea, the hyaline amorphous deposits *(arrows)* (keratinoid) stain strongly with the Masson stain and weakly in a haematoxylin and eosin-stained section (see *inset*).

Corneal dystrophies of Bowman's layer: Reis-Bücklers dystrophy (CDB 1) and Thiel–Behnke dystrophy (CBD 2) commence in early adult life with fine reticular opacities in the superficial cornea in CDB 1 and honeycomb opacities in CDB 2. The histological features are identical and not specific, with nodules of fibrous tissue between Bowman's layer and the epithelium. However, electron microscopy shows characteristic electron-dense rods in CDB1 and curly fibres in CDB2.

Lattice dystrophy, type 1: This dystrophy is characterized clinically by fine lines criss-crossing the stroma. Microscopy shows the deposits to consist of amyloid (Fig. 9.19). Secondary non-specific amyloid deposition is sometimes seen in the cornea at the end-stage of post-inflammatory scarring and fibrosis. Lattice dystrophy commonly recurs in a graft. Variants of lattice dystrophy have been described but are rare.

Granular corneal dystrophy, type 1: In this dystrophy there are discrete opaque granules with the stroma. Histologically, there are non-birefringent hyaline deposits (Fig. 9.32) from superficial stroma to Desçemet's membrane. The endothelium is not involved. Granular dystrophy occasionally recurs in a graft.

Granular corneal dystrophy, type 2: Formerly known as Avellino dystrophy in this dystrophy there is a combination of stromal amyloid and hyaline deposits extending the full stromal thickness.

Other stromal dystrophies. There is a group of stromal dystrophies that are not associated with TGβ1 mutations. These are uncommon and a full discussion is outside the scope of this chapter. One example is given.

Macular dystrophy. This disease is a localized form of mucopolysaccharidosis of autosomal recessive inheritance

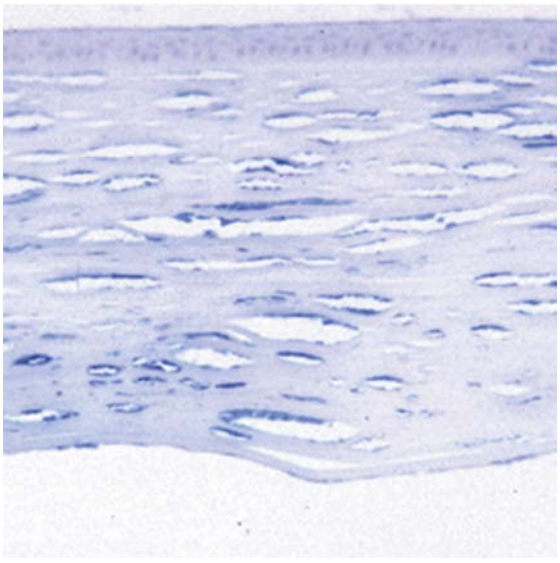


Fig. 9.33 In macular dystrophy of the cornea, mucopolysaccharide is deposited within the keratocytes and the endothelium (Alcian blue stain).

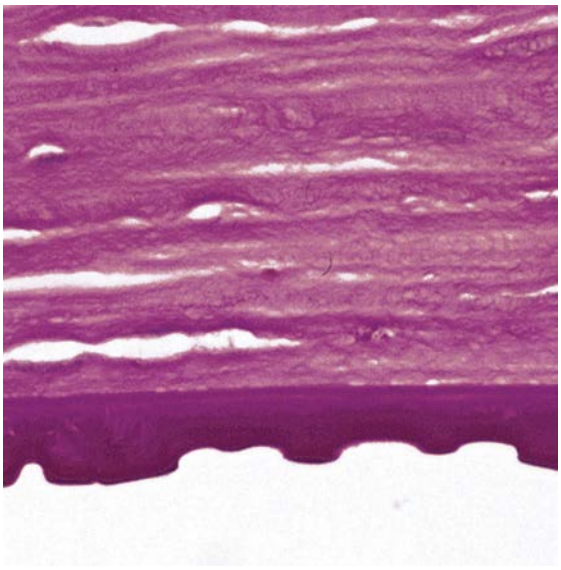


Fig. 9.34 In Fuchs' dystrophy, Desçemet's membrane is thickened and large excrescences project from the posterior surface (PAS stain).

due to a mutation in chromosome 16q22. Smudgy snowflake-like deposits cause severe visual impairment. Mucopolysaccharide (acidic glycosaminoglycans) granules accumulate in the cytoplasm of the keratocytes and in the adjacent interlamellar spaces when the cells rupture. The corneal endothelial cells are involved and the material also accumulates within Desçemet's membrane and beneath the epithelium (Fig. 9.33). Macular dystrophy rarely recurs in a graft.

Endothelial dystrophies. This is a group of disorders characterized by corneal oedema with opacification, occurring relatively early in life in the absence of pre-existing inflammation, glaucoma or identifiable systemic metabolic disorders. The clinical patterns are more important in the classification than the pathological findings, which often overlap.

Fuchs' endothelial dystrophy: This is a common dystrophy in elderly patients which affects females more than males. The genetic basis is complex and heterogeneous and cases without known inheritance are most common. Late onset Fuchs' dystrophy has been linked to mutations in the sodium bicarbonate transporter-like (SLC4A11) gene (Vilas et al., 2011) as well as other corneal dystrophies (see later) but the significance is not clear because this protein is widely distributed in many tissues. The clinical presentation includes bilateral diffuse oedema with cloudiness of the stroma. The abnormalities are restricted to epithelial oedema and thickening of Desçemet's membrane with an obvious reduction in the endothelial cell population. Large nodular excrescences are present on the posterior surface of Desçemet's membrane (Fig. 9.34) and the endothelial cells are of varying size. When excrescences (Hassall–Henle warts) are confined to the far periphery of the normal cornea, the effect on corneal transparency is minimal.

Posterior polymorphous dystrophy: This rare autosomal dominant disease, which causes bilateral non-progressive, circumscribed or diffuse opacities, is not usually severe enough to require keratoplasty until after the second decade. There are three variants and the mutated gene has been identified in two (COL8A2 and ZEB1). Morphologically, in the severe diffuse form of the disease, the posterior corneal surface is lined by stratified cells with prominent desmosomal attachments resembling corneal epithelial cells (Fig. 9.35).

Congenital hereditary endothelial dystrophy: A cloudy cornea in childhood or early adult life may occur in this dystrophy, the inheritance of which is autosomal recessive and, as for late Fuchs' dystrophy, is linked to mutations in SLC4A11. The endothelium is abnormal and may be attenuated and vacuolated, but the characteristic feature is seen in Desçemet's membrane, which shows fine lamination because of the deposit of an abnormal layer of collagen at the ultrastructural level.

X-linked endothelial corneal dystrophy: This causes a minimally progressive corneal haze in male infants with moon crater-like endothelial changes on retro-illumination and scarce endothelial cells on microscopy.

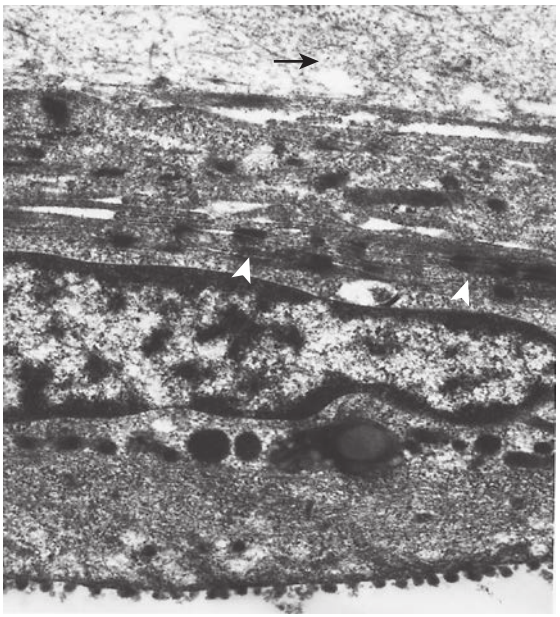


Fig. 9.35 In posterior polymorphous dystrophy, the cornea is lined on its posterior surface by cells that, by electron microscopy, exhibit all the features of epithelial cells with intracytoplasmic filaments and desmosomal attachments *(arrowheads)*. An abnormal collagenous layer *(arrow)* is deposited on the posterior surface of the original Desçemet's membrane.

Neoplasia

A neoplasm is a proliferation of cells, the growth of which is progressive, purposeless, regardless of surrounding tissue, not related to the needs of the body and persists after the stimulus that initiated it has been withdrawn.

Neoplasms may be classified clinically, as benign or malignant or according to their histological tissue of origin (Table 9.2). A benign tumour is usually well circumscribed and may be encapsulated. They grow slowly and remain localized at the site of origin. They may affect the host by producing pressure on adjacent structures (e.g. proptosis secondary to pleomorphic adenoma of the lacrimal gland). Malignant tumours have an irregular, ill-defined boundary and are non-encapsulated. They grow rapidly with local and distant spread and produce effects by destroying adjacent structures (e.g. liver metastases from uveal melanoma).

Pathogenesis of neoplasia

Premalignancy

A number of pathological conditions are associated with the development of malignancy. The main categories include malignant transformation of benign tumours, chronic inflammatory conditions and intraepithelial neoplasia.

Benign tumours may undergo malignant transformation. A good example of this is colonic cancer arising from a benign adenoma. This is thought to occur by progressive acquisition of genetic changes. Malignant transformation of benign tumours also occurs in ophthalmic pathology. For example, adenocarcinoma may arise in a long-standing pleomorphic adenoma of the lacrimal gland. Malignancy may develop in chronic inflammatory conditions, particularly if they are long-standing. For instance, in Sjögren's syndrome there is chronic lymphocytic infiltration of the lacrimal gland with acinar atrophy later leading to the clinical symptoms of dry eye. Evolution to lymphoma occurs in a significant number of patients with Sjögren's syndrome, possibly by the development of monoclonal lymphocytic populations within the lacrimal gland. This may in part be a result of dysregulated immunity, i.e. where suppression of host anti-tumour responses supervenes via immune checkpoint molecules such as PD-1 on T-regulatory cells (see Chapter 7, p. 454). Intraepithelial neoplasia represents an intermediate stage in the production of cancer. In the skin, excessive exposure to ultraviolet light may lead to development of an actinic or solar keratosis. Clinically, these appear as hyperkeratotic lesions on the face and histological examination reveals premalignant changes in the epidermis. This is seen as an increased mitotic rate, a loss of the normal polarity of maturation from basal cells to squamous cells, and a marked variation in the size and shape of nuclei (pleomorphism) within the epithelium (dysplasia). These histological changes precede invasion through the basement membrane of the epithelium into the underlying tissue and are therefore designated carcinoma *in situ*.

Carcinogenesis

Both environmental and genetic factors contribute to a cell undergoing malignant change. This should be regarded as a multi-step process. The three major environmental factors that induce tumours are chemicals, radiation and viruses. In chemical carcinogenesis, the first step, initiation, involves a short exposure of the cell to a carcinogen. This is followed by promotion, the long-term exposure to a substance that is usually not mutagenic but acts by stimulating cell proliferation (although some compounds can act as both initiators and promoters, so-called complete carcinogens). Ionizing radiation directly damages DNA, especially during cell proliferation, and can result in a range of changes from single gene mutations to major chromosome deletions. Ultraviolet radiation mainly affects the skin-forming pyrimidine dimers that can usually be excised by DNA repair mechanisms. In xeroderma pigmentosum these are deficient and multiple skin tumours occur. Viruses may contribute to the development of some human cancers. Oncogenic viruses may contribute to several conditions in the eye. Conjunctival

Table 9.2 **Histological classification of tumours**

Histological origin	Benign	Malignant
Epithelial cells		
Surface	Papilloma	Carcinoma (squamous, basal cell, etc.)
Glandular	Adenoma	Adenocarcinoma
Mesenchymal cells		
Adipose	Lipoma	Liposarcoma
Fibrous	Fibroma	Fibrosarcoma
Cartilage	Chondroma	Chondrosarcoma
Bone	Osteoma	Osteosarcoma
Smooth muscle	Leiomyoma	Leiomyosarcoma
Striated muscle	Rhabdomyoma	Rhabdomyosarcoma
Neuroectodermal cells		
Glial cells		Glioma
Nerve	Ganglioneuroma	Neuroblastoma
Retinal cells		Retinoblastoma
Melanocytes		Melanoma
Meninges	Meningioma	Malignant meningioma
Schwann cells	Neurofibroma	Malignant peripheral nerve sheath tumour
Haemopoietic/lymphoreticular		Leukaemia
		Lymphoma
Germ cells	Benign teratoma	Malignant teratoma
		Dysgerminoma
		Seminoma

Tumours may arise from any tissue in the body but for convenience these are divided into five groups. Not all malignant tumours have a benign counterpart and similarly there are some types of benign tumour for which malignant counterparts are extremely rare.

papillomas and papillomas of the lacrimal passages may be caused by human papillomavirus types 16 and 11, respectively. Epstein–Barr virus contributes to the development of orbital Burkitt's lymphoma and to intraocular diffuse large B-cell lymphoma occurring in immunosuppressed patients.

A genetic influence in cancer is now well recognized. Certain syndromes inherited in a Mendelian fashion show a high risk of cancer. Examples of these include xeroderma pigmentosum, described previously, and neurofibromatosis, an autosomal dominant trait characterized by multiple neurofibromas and increased risk of sarcoma, which is the result of a defect of the *NF1* gene on chromosome 17. In Li–Fraumeni syndrome there is a high risk of several types of cancer, such as childhood sarcomas and breast cancer in young women. This is the result of a germ-line mutation of the *p53* gene.

Oncogenes and tumour suppressor genes

Cellular proto-oncogenes are normal genes that stimulate cell division. Tumour suppressor genes are normal genes that inhibit cell division. Proto-oncogenes and tumour suppressor genes are active during somatic growth, regeneration and repair and the balance between stimulation and inhibition of cell growth is strictly controlled. This balance is permanently lost in cancer cells.

Cellular proto-oncogenes code for a number of proteins involved in cell proliferation, including growth factors, growth factor receptors, signal transducers within the cell cytoplasm and nuclear-regulating proteins. In cancer these normal genes are permanently changed to oncogenes and proliferation is uncontrolled. Proto-oncogenes may become oncogenes by mutation, resulting in the production of a functionally abnormal protein or overexpression.

Tumour suppressor genes are normal genes that switch off cell proliferation. Loss of both copies of a tumour suppressor gene is required for cancer to develop. Loss of the *Rb* gene is important in the development of retinoblastoma (see Chapter 3, p. 161).

Tumour spread and metastases

Malignant tumours spread by several routes:

- local invasion of normal tissue (e.g. basal cell carcinoma)
- lymphatic spread (e.g. squamous carcinoma of the eyelid) or haematogenous spread (e.g. malignant melanoma of the choroid)

- intraepithelial spread (e.g. pagetoid spread of sebaceous carcinoma of the eyelid)
- dissemination along natural passages (e.g. retinoblastoma extending to subarachnoid space).

The basic mechanisms of tumour cell invasion involve several processes:

- tumour cells secrete lytic enzymes to breach the basement membrane
- there is a loss of cell–cell adhesion molecules, often accompanied by an increase in cell–matrix adhesion molecules
- increased cell movement allows tumour cells to penetrate further and spread.

Hamartomas

A hamartoma is a tumour-like but non-neoplastic malformation consisting of a mixture of tissues normally found at a particular site. The commonest forms of hamartoma are those composed of blood vessels and those involving melanocytes of the skin.

Haemangiomas

Capillary haemangiomas are a proliferation of small-calibre vascular channels with a lobulated growth pattern. Cavernous haemangiomas consist of large-calibre thick-walled vascular channels with intervening fibrous septae. Both capillary and cavernous haemangiomas may occur in the eyelid, orbit or choroid. Extensive haemangiomas may occur as part of encephalo-trigeminal angiomatosis (Sturge–Weber syndrome). Capillary haemangiomas may regress spontaneously during childhood but cavernous haemangiomas show no tendency for spontaneous regression.

Naevi

A naevus is the result of abnormal migration, proliferation and maturation of neuroectodermal cells. They may occur in the uvea as well as the skin (Fig. 9.36). In the skin, the melanocytes initially form clumps at the junction between the epidermis and dermis. Clinically this appears as a brown macule, referred to as a junctional naevus. With age, the proliferating melanocytes begin to detach from the epithelium and migrate into the dermis, forming a brown papule. When proliferation is found in the dermis as well as the junctional area, the naevus is classified as compound. At a later stage the proliferation is wholly in the dermis and is classified as an intradermal naevus.

Naevi, similar to their cutaneous counterpart, occur in the conjunctiva (Fig. 9.37). In the iris and choroid, naevi are seen as static flat brown or black areas. Naevi in any site may occasionally progress to malignant melanoma.

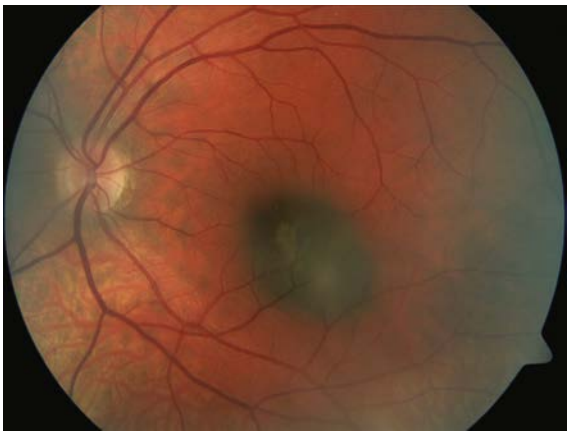


Fig. 9.36 Fundoscopic image of a choroidal naevus showing a well-circumscribed, slightly elevated, pigmented tumour. Distinguishing features from melanoma are the clearly defined margins and lack of overlying subretinal fluid and orange pigment. Over time there will be minimal or no documented growth. (Image courtesy Dr Paul Cauchi, Glasgow.)

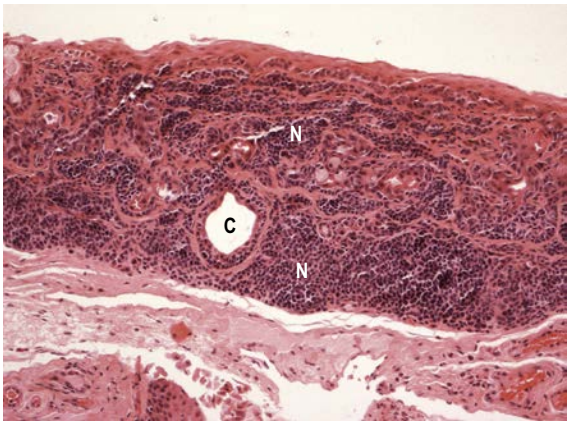


Fig. 9.37 A compound naevus of the conjunctiva showing nests of mature naevus cells *(N)*. There is entrapment of the surface epithelium, which results in characteristic small cysts *(C)* within the naevus.

Choristomas

In contrast with hamartoma, a choristoma is a tumour-like but non-neoplastic malformation consisting of a mixture of tissues not normally present at a particular site.

Dermoid

Epibulbar dermoids are relatively common choristomas. They occur as a nodule (smooth white swellings from which hairs project) on the bulbar conjunctiva in children

or at the outer angle of the bony orbit on the skin. Histological examination reveals a mixture of fat, fibrous tissue, hair follicles and sweat glands.

Teratoma

This is a tumour derived from totipotent germ cells. They can occur at any site in the mid-line where germ cells have stopped on their migration to the gonads. Orbital teratomas are rare and occur in neonates. An orbital teratoma causes proptosis and histological examination of the large cystic retroocular mass will reveal tissue derived from the three embryonic germ cell layers such as respiratory or gastrointestinal epithelium, stroma containing fat, cartilage and bone, and neuroectodermal tissues. Most orbital teratomas are benign and surgical removal is curative.

Benign epithelial tumours

Benign tumours of surface epithelium

A papilloma is a benign tumour originating from an epithelial surface. In the eyelid the commonest tumours are basal cell papilloma (seborrhoeic keratosis) and squamous cell papillomas. Benign squamous proliferations may be associated with poxvirus (molluscum contagiosum) or human papillomavirus (viral wart). Conjunctival papillomas can be pedunculated or sessile. The pedunculated papillomas are usually covered by conjunctival epithelium, whereas sessile papillomas commonly show squamous differentiation. Conjunctival papillomas are also commonly associated with human papillomavirus.

Benign tumours of adnexal glands

An adenoma is derived from the ducts and acini of glands. In the eyelid and caruncle these may be derived from sweat glands, pilosebaceous hair follicles and sebaceous glands, the largest being the meibomian gland in the tarsal plate. Sweat gland adenomas are subclassified according to the degree of differentiation towards acinar or ductular structures. Similarly, tumours of hair follicles are classified according to differentiation towards different components of the hair follicle. For example, a pilomatrixoma shows differentiation towards hair matrix. Sebaceous adenomas are proliferations of lipid-laden sebaceous cells and most commonly occur as a yellow mass at the caruncle.

Malignant epithelial tumours

Basal cell carcinoma

Basal cell carcinoma is the most common malignant tumour in clinical ophthalmology, accounting for more than 90% of malignant eyelid tumours. It usually occurs in Caucasians over 50 years of age and is associated with sunlight exposure. It may also occur in younger patients in association with the Gorlin–Goltz syndrome (basal cell naevus syndrome). Clinically, these tumours may present as nodular lesions, which later may develop a central ulcer with a rolled edge. Morpheiform types present as a scirrhous plaque. Occasionally, basal cell carcinomas are pigmented because of melanin deposition and clinically may be confused with malignant melanoma. Basal cell carcinoma is locally aggressive and adequate surgical excision is the treatment of choice to prevent recurrence and orbital invasion. Orbital invasion may necessitate exenteration.

Four main histological subtypes should be recognized:

- *Nodular/solid basal cell carcinoma* consists of well-circumscribed and relatively large islands of proliferating basal cells with a peripheral palisade of nuclei (Fig. 9.38A). Cystic degeneration may occur. Surgical excision should not present problems.
- *Superficial basal cell carcinoma* is less common than the nodular subtype and presents as a scaly plaque. Histology shows small nests of tumour cells budding from the undersurface of the epidermis only as far as the superficial dermis. There may be substantial gaps between the nests of tumour cells such that complete surgical excision may be difficult. Topical therapy may be effective.
- *Infiltrative/sclerosing subtype* corresponds to the morpheic clinical subtype and is a more aggressive form of basal cell carcinoma where the tumour cells grow in small strands rather than nests and are embedded in a dense fibrous stroma (Fig. 9.38B). A lack of distinct border makes adequate surgical excision difficult.
- *The micronodular subtype* forms small nodular aggregates of basaloid cells and, similar to the infiltrative subtype, subclinical involvement is often significant.

Squamous cell carcinoma

Compared with basal cell carcinoma, the incidence of this form of malignancy is low, representing between 1% and 5% of all eyelid cancers. Risk factors for squamous cell carcinoma include sunlight exposure and immunosuppression. Clinically, squamous cell carcinoma presents as a rapidly growing nodular ulcer or as a papillomatous growth, which in some cases has an overlying keratinous horn. Inadequate primary local excision may be followed by recurrence and orbital invasion. Lymphatic spread may occur to pre-auricular and submandibular lymph nodes according to the site of origin (upper and lower lid, respectively).

Histologically, squamous cell carcinoma may be classified as well, moderately or poorly differentiated. In a well-differentiated tumour, the cells have glassy pink cytoplasm and intercellular bridges and keratin pearls may be present (Fig. 9.39). Some of these features are lost in more poorly differentiated tumours, but intercellular bridges can usually still be identified. Rarely, squamous cell carcinoma will adopt a spindle cell morphology and this variant is more aggressive.

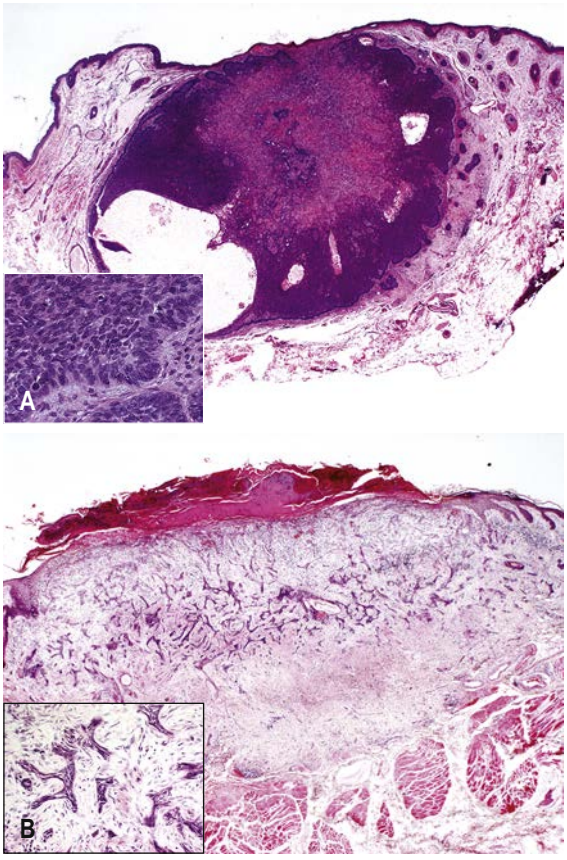


Fig. 9.38 (**A**) A nodular basal cell carcinoma has a well-defined margin. Cystic degeneration is common. The *inset* shows the characteristic peripheral palisading at the edge of tumour cell lobules. (**B**) The margins of a sclerosing basal cell carcinoma are poorly defined and small strands of tumour can invade deeply without being clinically apparent. The jagged outline of the strands of tumour cells are shown in the *inset*. Peripheral palisading of nuclei is not prominent in this type of basal cell carcinoma. The fibroblastic intervening tissue is known as a desmoplastic stroma.

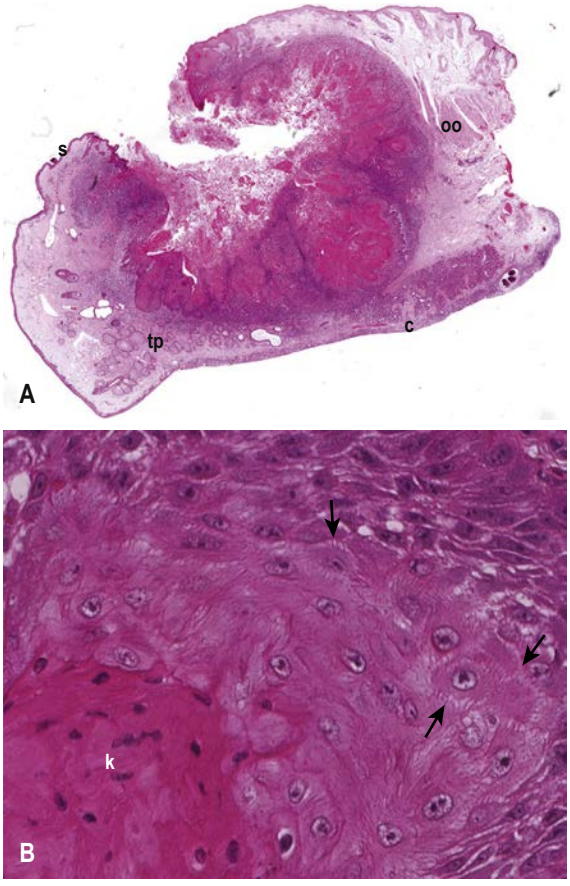


Fig. 9.39 (**A**) This squamous cell carcinoma of the eyelid infiltrates extensively from the skin surface *(s)* through orbicularis oculi *(oo)* and just into the tarsal plate *(tp)*. The conjunctival surface *(c)* is not involved. (**B**) The cells show keratinization *(k)* and intercellular bridges *(arrows)*.

Ocular surface squamous neoplasia (OSSN) describes the clinical spectrum of *in situ* and invasive squamous cell carcinoma, which may involve the conjunctiva and cornea . The morphology of these tumours is identical to that of the eyelid tumours and they are also associated with sunlight exposure and immunosuppression, particularly HIV infection.

Sebaceous gland carcinoma

Sebaceous gland carcinoma accounts for 1%–5% of all eyelid cancers. These tumours usually originate in the meibomian gland but may also arise from the gland of Zeis or sebaceous glands of the eyelid skin. Sebaceous gland carcinoma commonly occurs in elderly patients and shows a female preponderance. The clinical appearance of sebaceous gland carcinoma is variable and it may be indistinguishable from squamous cell carcinoma or basal cell carcinoma or may mimic a range of benign conditions, including chalazion and blepharoconjunctivitis.

The variable clinical appearance is related to the different histological growth patterns of this tumour. These tumours may show a nodular or diffuse pattern of growth and may be well, moderately or poorly differentiated. The nodular pattern consists of lobules of tumour cells with foamy or vacuolated cytoplasm (Fig. 9.40A). Diffuse tumours show individual tumour cells spreading within the surface epithelium (pagetoid spread) and adnexal structures (Fig. 9.40B). Stains for fat performed on frozen section are rarely done nowadays. Immunohistochemical staining, which is positive for epithelial membrane antigen, adipophilin and androgen receptors can be helpful. Sebaceous carcinomas may be seen in Muir–Torre

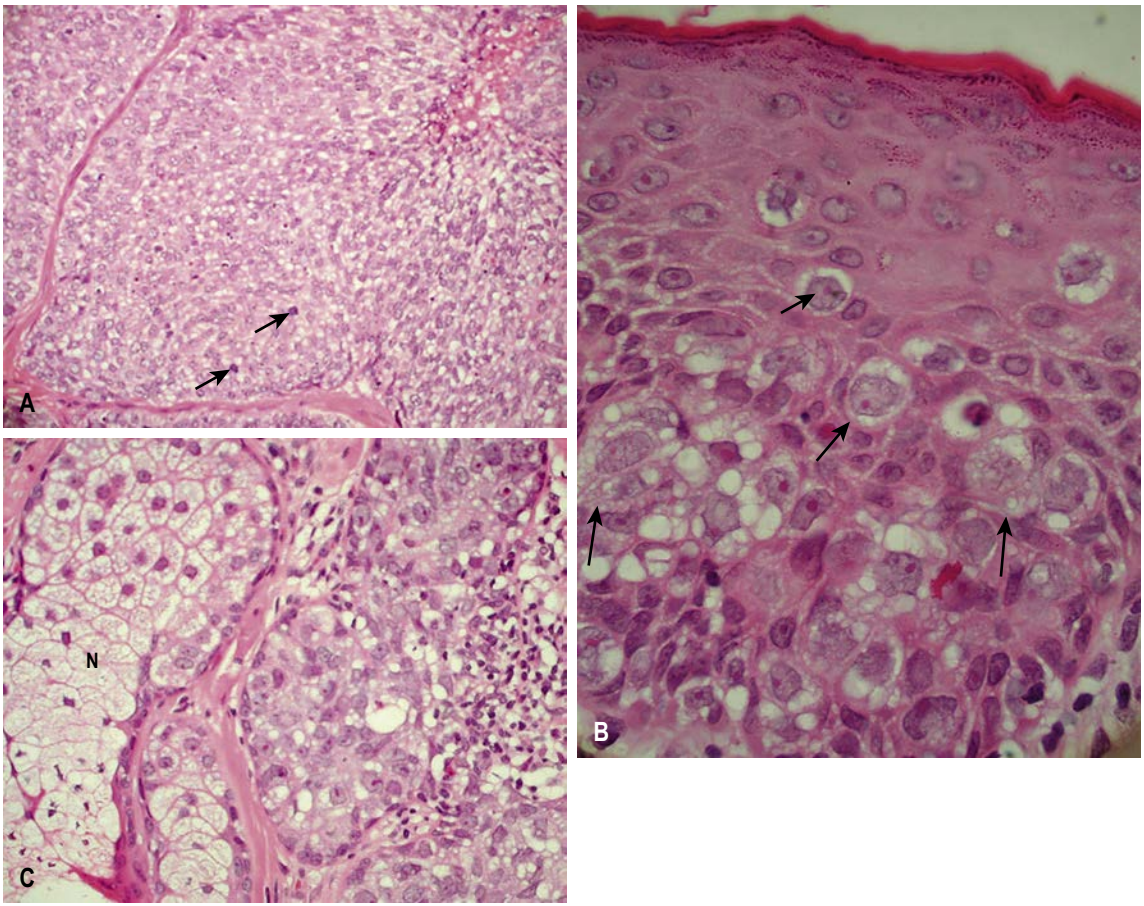


Fig. 9.40 (**A**) Sebaceous gland carcinoma consisting of lobules of tumour cells with vacuolated cytoplasm. There are frequent mitotic figures *(arrows)*. (**B**) Pagetoid spread in sebaceous gland carcinoma showing individual atypical cells with vacuolated cytoplasm *(arrows)* within the epidermis. (**C**) Sebaceous gland carcinoma colonizing a sebaceous gland. An adjacent normal sebaceous gland *(N)* is present for comparison.

syndrome in association with a visceral malignancy. In these tumours there may be germline mutation in the DNA repair *genes MSH2*, *MLH1*, *MSH6* and *PMS2*. Loss of staining can be detected with immunohistochemistry. The prognosis is poor compared with most other malignant eyelid tumours but is significantly improved with early diagnosis and surgery.

Malignant melanoma

Conjunctiva

Malignant melanoma may arise from primary acquired melanosis, a pre-existing naevus or *de novo*. Primary acquired melanosis (also known as conjunctival intraepithelial melanocytic neoplasia) appears as unilateral or bilateral, diffuse flat areas of conjunctival pigmentation in middle-aged to elderly patients. Conjunctival melanoma presents as a raised, pigmented or fleshy conjunctival lesion. Like its cutaneous counterpart (and in contrast with uveal melanoma), it has a tendency to metastasize to regional lymph nodes but may also spread to the brain and other organs. The prognosis is worse for tumours thicker than 5 mm and located in the fornix. Complete excision is the treatment of choice. Those arising on a background of primary acquired melanosis may be multifocal and topical chemotherapy with mitomycin C can be helpful.

Uveal melanoma

Malignant melanoma of the uveal tract arises from melanocytes in the iris, ciliary body and choroid, and the relative incidence is roughly in proportion to the volume of tissue in each compartment (8%, 12% and 80%, respectively). These tumours are almost always unilateral and grow initially as pigmented or non-pigmented plaque-like lesions; the macroscopic appearances are shown in Fig. 9.41 (A–D).

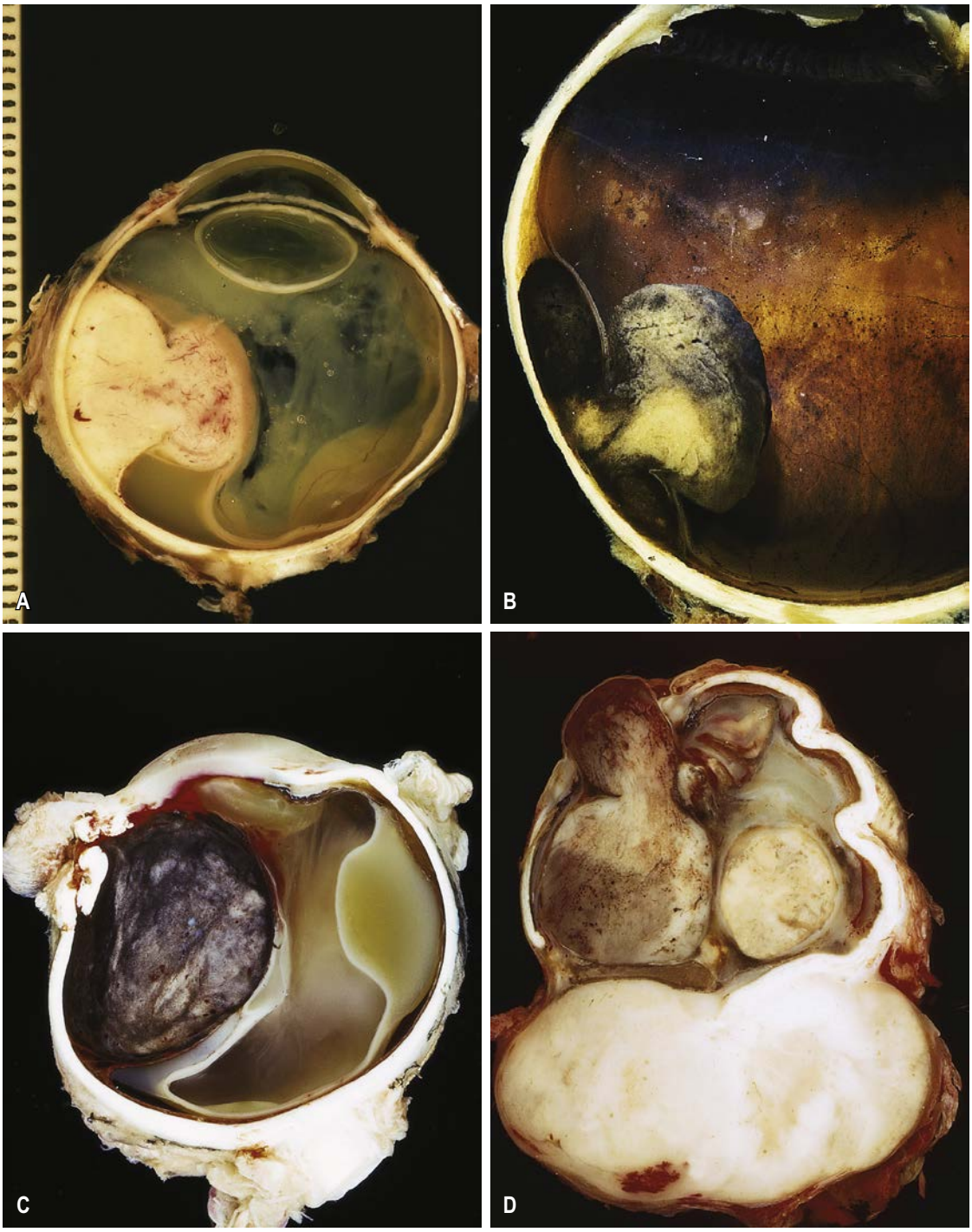


Fig. 9.41 Various macroscopic appearances of uveal melanoma. (**A**) An amelanotic tumour that has breached Bruch's membrane and spread in the subretinal space resulting in a mushroom shape mushroom shape. (**B**) A partially pigmented melanoma that has perforated the retina. (**C**) This ovoid black melanoma has leaked fluid into the subretinal space, causing an exudative retinal detachment. An attempt to remove the tumour surgically was abandoned. (**D**) An advanced melanoma perforating the anterior and posterior sclera.

Iris melanomas. Iris melanomas are usually slow-growing nodular tumours that may be present for many years. On histology, iris melanomas may consist of small, rather bland, spindle-shaped cells and the diagnosis depends on the identification of surface or stromal invasion. Recurrent iris melanomas often transform into pleomorphic epithelioid tumours. Although they often remain localized for a long period of time, iris melanomas can spread diffusely on the iris surface and around the chamber angle, resulting in secondary glaucoma as a result of infiltration of the trabecular meshwork.

Ciliary body and choroidal melanomas. Ciliary body and choroidal melanomas can grow to a large size (10–20 mm) before recognition. The macroscopic appearances can vary considerably. The tumours may be ovoid, nodular or a classical mushroom shape. This is caused by tumour spread in the subretinal space after breaching Bruch's membrane. Tumours may cause an exudative retinal detachment or acute angle closure glaucoma as a result of direct pressure effects or neovascular glaucoma because of production of vasoformative factors by the tumour. Extraocular extension may be identified in relation to collector channels (anteriorly), vortex veins (in the mid-periphery) or short ciliary vessels (posteriorly) and tumours may occasionally present as proptosis due to an orbital mass. Larger tumours may undergo spontaneous necrosis and produce symptoms of endophthalmitis. Uveal melanoma can be simulated clinically by various other entities. Some typical examples are shown in Fig. 9.42A–D.

On histology, the tumours are classified according to cell type as spindle, epithelioid and mixed (Fig. 9.43). The majority of tumours are of mixed cell type. Vascular patterns may be assessed in melanomas using a periodic acid–Schiff (PAS) stain. The presence of microscopic intrascleral spread or extraocular extension is also important for staging uveal melanoma. Metastatic spread (most commonly to the liver, *cf* conjunctival melanoma) usually occurs within 2–3 years but has also been recorded up to 40 years later. Immunohistochemistry, which is usually positive for S100, HMB45 and Melan A, can be a useful ancillary technique, particularly in metastatic melanoma.

Treatment includes enucleation, local resection and ionizing radiation in the form of plaque brachytherapy or proton beam. Secondary enucleation following surgery or irradiation may show a surgical coloboma or neovascular glaucoma, cataract and radiation retinopathy in addition to recurrent tumour (see Video 9.2).

Prognostic parameters including cytogenetics (analysis of chromosomal abnormalities) in uveal melanoma are shown in Box 9.13.

Retinoblastoma

Retinoblastoma is a malignant tumour of infancy that is lethal if untreated; the incidence is 1 in 20,000 live births. The tumour arises from embryonal retinal cells and can be unilateral or bilateral. The macroscopic appearances are of a smooth-surfaced white mass that may show endophytic growth into the vitreous or exophytic growth into the subretinal space (Fig. 9.44A–D). Yellowish areas of necrosis or bright white flecks of calcification may be evident within the tumour. On histology the tumour consists of small blue cells with scanty cytoplasm. There is usually a high mitotic rate, with prominent apoptosis and areas of necrosis within the tumour indicating high cell turnover. Differentiation may be seen in the form of:

- Homer–Wright rosettes: a multilayered circle of nuclei surrounding eosinophilic fibrillar material (Fig. 9.45A)
- Flexner–Wintersteiner rosettes: a circle of cells limited internally by a continuous membrane (Fig. 9.45B)
- Fleurettes: primitive photoreceptor bodies arranged in 'fleur de lys' shape. These structures are most commonly found in irradiated tumours.

Features of prognostic importance in retinoblastoma include tumour size, degree of differentiation, choroidal invasion and optic nerve invasion. With early diagnosis and modern treatment, including irradiation and chemotherapy, cure rates are in excess of 90%. In untreated cases, death is caused by tumour spread to the brain through the optic nerve or along the meninges; metastatic dissemination is via the bloodstream to the viscera and skeleton.

The genetics of retinoblastoma are discussed in Chapter 3, p. 161, but it is noteworthy that the abnormal gene carries the risk of a pineal tumour in childhood (trilateral retinoblastoma), soft tissue and osteogenic sarcoma in early adult life and carcinomas in later life.

The differential diagnosis of retinoblastoma includes:

- Coats' disease (Fig. 9.46A)
- Astrocytic hamartoma (Fig. 9.46B)
- Retinopathy of prematurity
- Persistent hyperplastic primary vitreous (Fig. 9.46C)
- Endophthalmitis (Fig. 9.46D)
- *Toxocara* retinitis.

Astrocytic hamartoma

Benign astrocytic tumours occur in the retina as part of the tuberous sclerosis syndrome or as an isolated feature. They consist of astrocytes, which form a matrix conducive to the deposition of calcospherites. The presence of calcification may lead to an erroneous diagnosis of retinoblastoma.

Glioma

Juvenile and adult forms of optic nerve glioma are recognized: the former carry a good prognosis and are commonly associated with neurofibromatosis type 1 (NF1); the latter are very rare and invariably lethal, being associated with extensive intracranial extension.

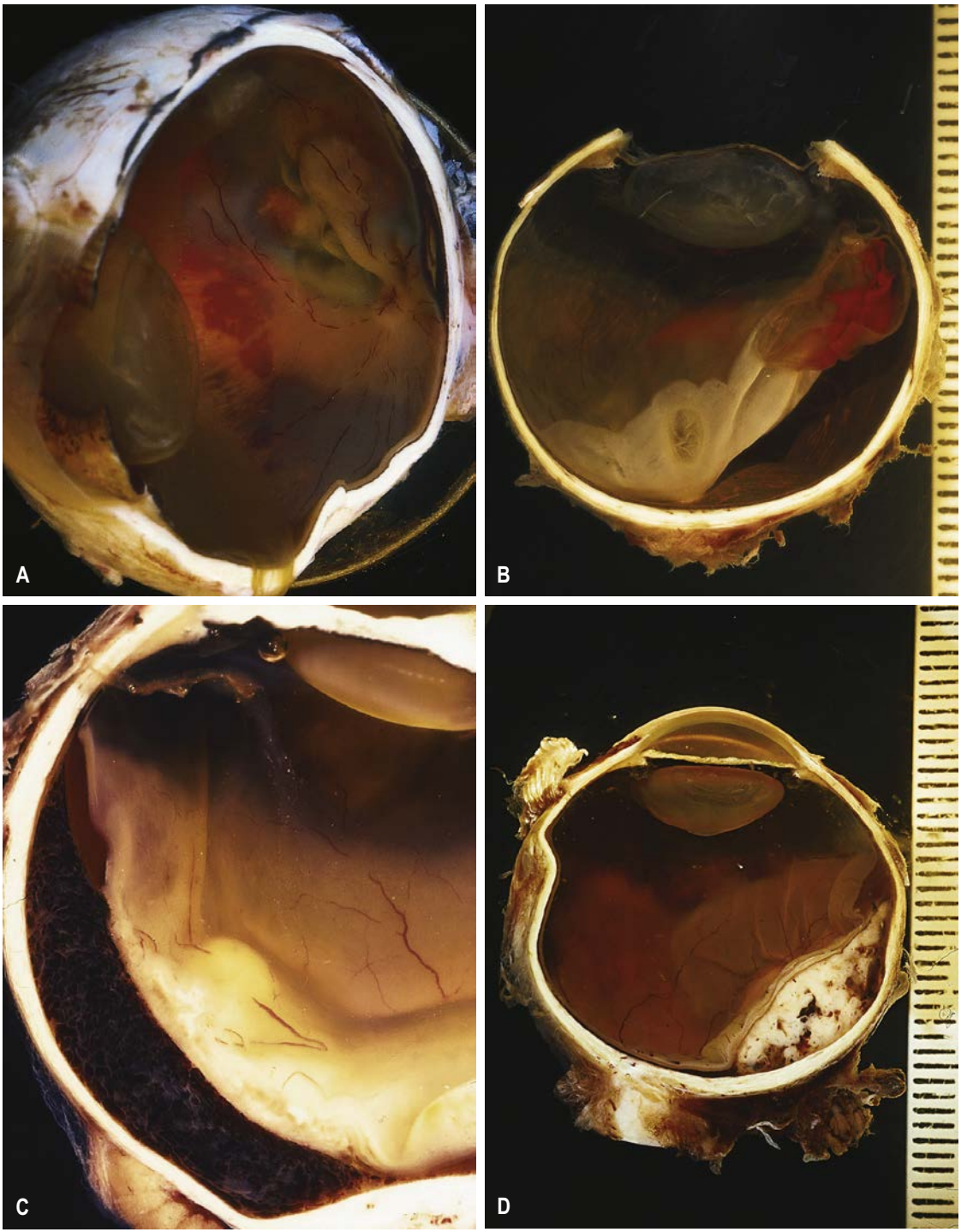


Fig. 9.42 Diseases simulating uveal melanomas. (**A**) Blood arising from disciform degeneration of the macula. (**B**) Bleeding into a macrocyst arising from a peripheral microcystoid degeneration of the retina. (**C**) Angioma of the choroid. (**D**) Metastatic tumours from breast or lung.

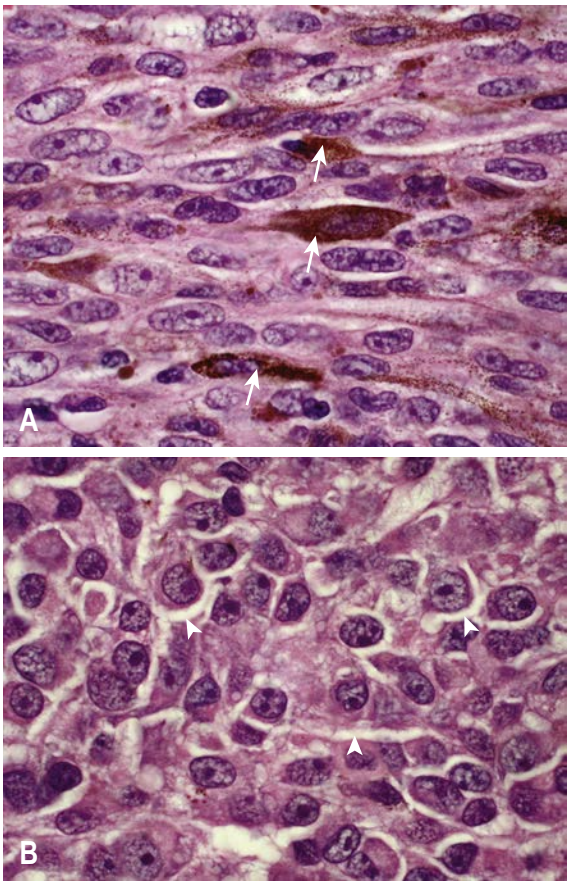


Fig. 9.43 (**A**) In this malignant uveal melanoma of spindle B type the cytoplasm of the cells contains melanin granules *(arrows)*. (**B**) Epithelioid melanoma cells are larger than spindle B cells and are separated from each other by prominent intercellular spaces *(arrowheads)*.

Around 50% of gliomas involve the orbital portion of the nerve but the intracranial or chiasmal portions may also be involved. In the orbital portion, the tumour may cause proptosis in addition to optic disc swelling and visual loss. Computed tomography or magnetic resonance imaging may be helpful in delineating the location, configuration and extent of the tumour. In over 50% of patients the tumour does not grow but in the remaining cases the tumour does grow and may require surgical intervention. The affected region of the nerve may be excised or if there is extensive tumour with secondary complications, such as exposure keratitis, then the eye may be removed along with the affected segment of the nerve. Optic nerve gliomas have an excellent prognosis following complete surgical excision although vision is usually sacrificed.

Box 9.13 **Factors associated with poorer prognosis in uveal melanoma**

Clinical features

- Older age
- Large tumour size (>16mm BD or >10mm in height)
- Tumour thickness
- Ciliary body location
- Diffuse tumour (height <20% of BD)
- Extraocular extension

Histopathology features

- Epithelioid cell type
- Closed vascular loops
- Extrascleral spread

Cytogenetic features

- Chromosome 3 loss (monosomy 3)
- Chromosome 8q gain or 8p loss
- Chromosome 1p loss
- Chromosome 6q loss

Excised tumours show a fusiform swelling of the nerve and the residual nerve may be barely visible within the tumour mass. The histology of these tumours is identical to that of intracranial astrocytomas and the majority are pilocytic, often containing areas of myxoid degeneration and Rosenthal fibres (Fig. 9.47). A potential diagnostic pitfall is that these tumours can induce proliferation of the overlying arachnoid. This hyperplastic tissue may be misdiagnosed as meningioma if the biopsy contains only perineural tissue.

Meningioma

Meningioma of the optic nerve may be primary, arising from the meninges of the optic nerve, or secondary, as the result of extension of an intracranial meningioma. In adults, meningiomas of the optic nerve characteristically show indolent growth but in children they may be more aggressive. The tumour ensheaths the optic nerve, which may become atrophic. The histology is similar to intracranial meningioma, with a transitional pattern, sometimes with psammoma bodies, predominating.

Tumours of peripheral nerves

Schwannomas and neurofibromas rarely occur in the orbit or periorbital tissues. Neurofibromas, particularly the plexiform variant, may be associated with NF1.

Tumours derived from muscle

These rare tumours may affect the eye or the orbit.

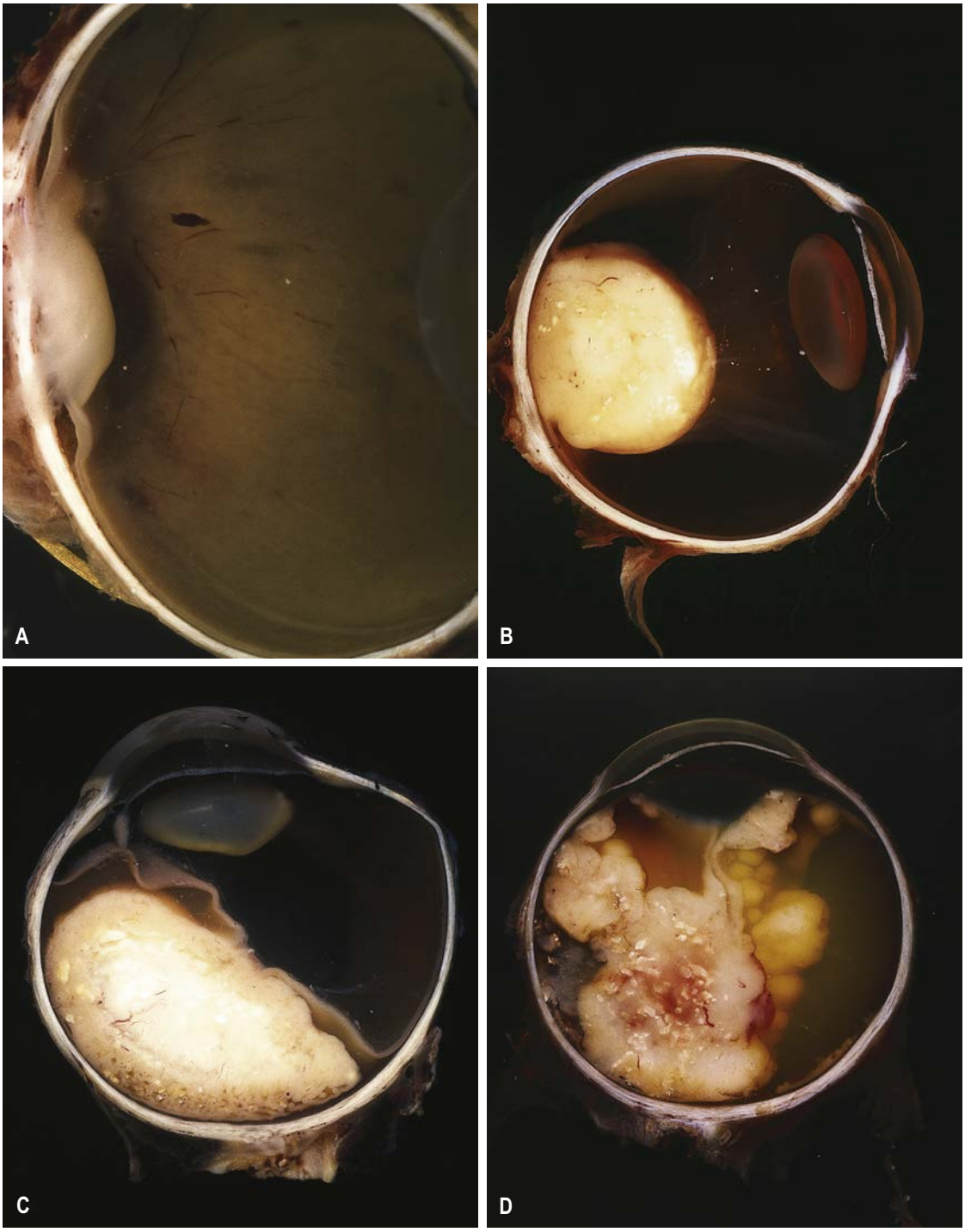


Fig. 9.44 Macroscopic appearances of retinoblastoma. (**A**) A small retinoblastoma overlying the optic nerve head contains a few flecks of calcium. (**B**) A large retinoblastoma with seedlings in the vitreous. (**C**) An exophytic retinoblastoma detaching the retina . (**D**) A large exophytic retinoblastoma with prominent calcification and funnel-shaped retinal detachment.

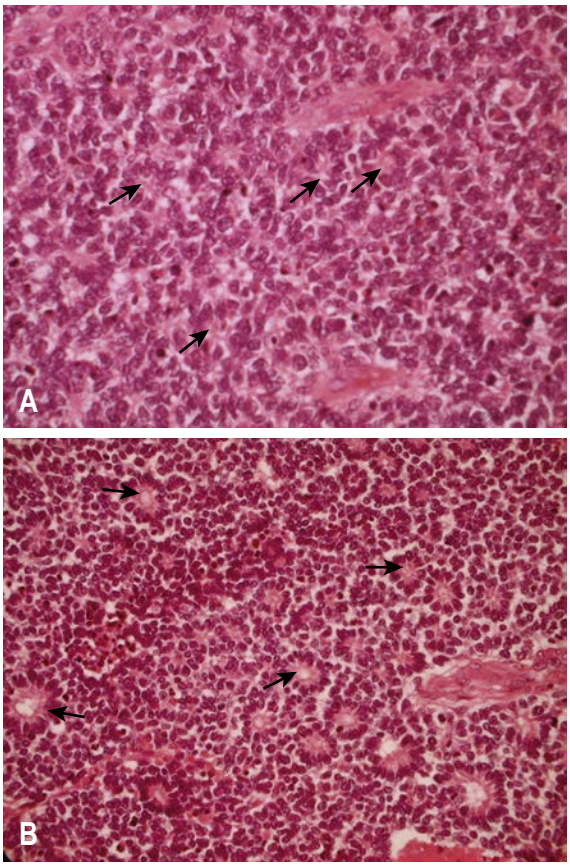


Fig. 9.45 Differentiation in retinoblastoma is represented by various forms of rosette. (**A**) Homer–Wright rosettes *(arrows)* consist of a central fibrillary tangle surrounded by cell nuclei that may overlap. (**B**) Flexner–Wintersteiner rosettes are lined internally by a membrane similar to the outer limiting membrane of the normal retina *(arrows)* and surrounded by a single layer of cells.

Leiomyoma/leiomyosarcoma

A leiomyoma occasionally arises from the smooth muscle of the iris and ciliary body. The malignant counterpart, leiomyosarcoma, is extremely rare.

Rhabdomyoma/rhabdomyosarcoma

Benign tumours of striated muscle are virtually unknown in the eyelid and orbit.

Rhabdomyosarcoma is the most common orbital malignancy of childhood. It generally occurs in the first two decades of life and usually presents with rapidly progressive proptosis and displacement of the eye. A prompt biopsy should be performed to confirm the diagnosis and the patient should be treated with a combination of chemotherapy and radiotherapy. With this regimen the survival of children with rhabdomyosarcoma has dramatically improved. On macroscopic examination, these tumours consist of tan-coloured fleshy tissue. On histopathological examination, rhabdomyosarcoma can be divided into three subtypes (embryonal, alveolar and pleomorphic). The embryonal subtype is the most common type in the orbit, consisting of sheets of small ovoid to spindle-shaped cells (Fig. 9.48). Cytoplasmic cross-striations can be seen with difficulty in a small number of cases. Immunohistochemistry for *MyoD1*, a muscle regulatory gene, may be helpful in confirming the diagnosis. Alveolar rhabdomyosarcoma is more common in older children; pleomorphic rhabdomyosarcoma is rare in the orbit and usually occurs in adults.

Vascular tumours

These tumours are also rare but occur at all age groups.

Haemangiomas

These are described under hamartomas.

Epithelioid haemangioma

Epithelioid haemangioma, previously named angiolymphoid hyperplasia with eosinophilia, is a benign vascular lesion that may occur on the eyelid skin or occasionally in the orbit. It consists of blood vessels with prominent endothelial cells and accompanying inflammatory cells, including lymphoid follicles and prominent eosinophils.

Kaposi's sarcoma

This is a tumour of endothelial cells that may occur as a rapidly growing tumour on the eyelid and conjunctiva. It most commonly occurs in immunocompromised patients, especially those with AIDS, and is caused by infection with herpes virus type 8. Histologically, these tumours consist of malignant spindle cells lining a network of sieve-like spaces containing extravasated red cells.

Other connective tissue tumours

Tumours can be derived from any of the cellular constituents of connective tissue, including adipose tissue, cartilage and bone. Orbital lipomas are considered to be relatively rare but this may reflect underdiagnosis because of the difficulty of distinguishing a lipoma from excised orbital fat. Primary liposarcoma rarely involves the orbit. Similarly, cartilaginous neoplasms of the orbit are extremely rare. Orbital osteosarcoma is also rare but well recognized as a second primary neoplasm following successfully treated retinoblastoma.

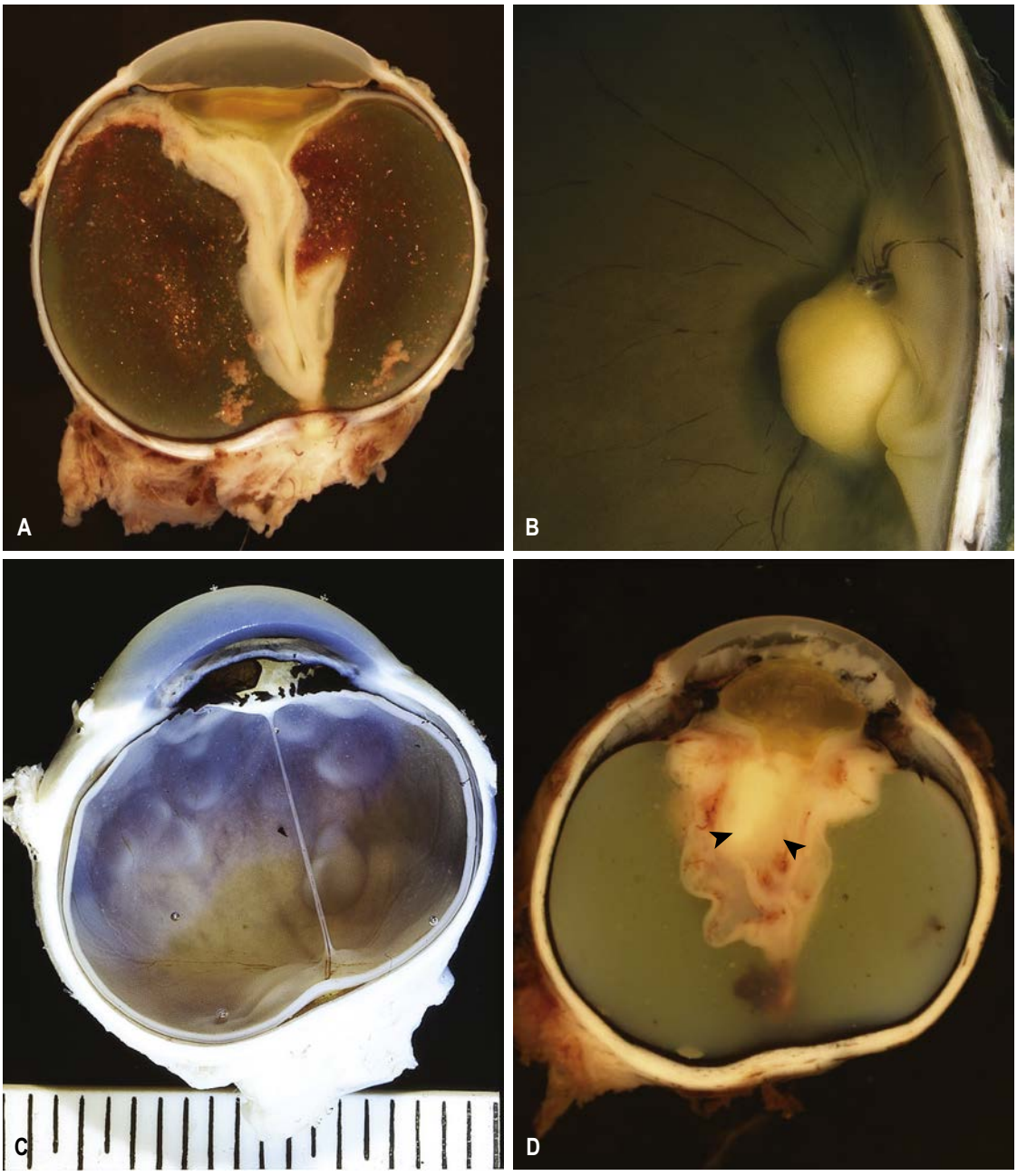


Fig. 9.46 Diseases simulating retinoblastoma. (**A**) In Coats' disease, the abnormal vasculature leaks lipid-rich plasma and cholesterol crystals are present in the subretinal exudate. (**B**) Astrocytic hamartoma appears as a static round nodule projecting from the retina into the vitreous. (**C**) Persistent hyperplastic vitreous forms a white mass behind the lens and the persistent hyaloid artery passes back to the optic nerve head. (**D**) In metastatic endophthalmitis an abscess fills the vitreous cavity, forming a white mass *(arrowheads)* and the retina is detached.

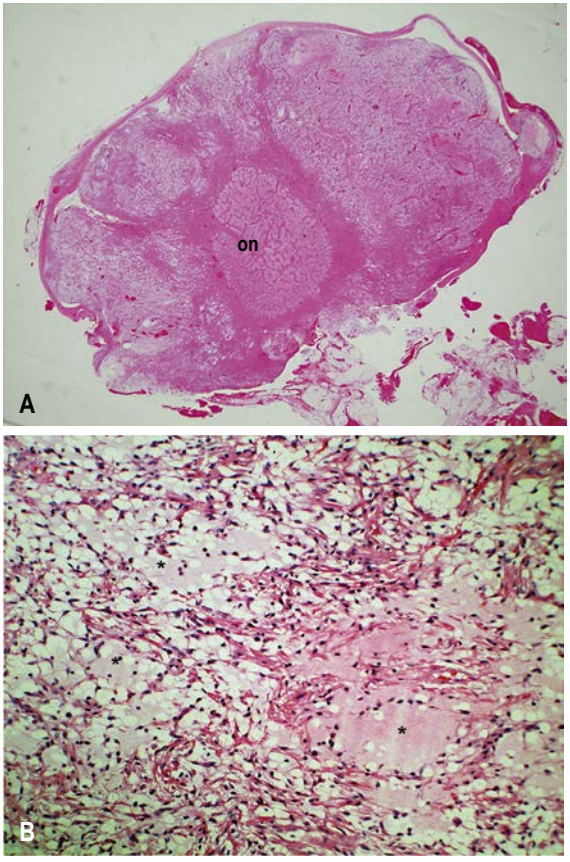


Fig. 9.47 (**A**) Low-power view of an optic glioma that has infiltrated the nerve columns but has preserved some recognizable architecture in the optic nerve *(on)*. (**B**) At higher magnification the tumour consists of irregularly arranged proliferating astrocytes with numerous small cystic spaces *(*)*.

Solitary fibrous tumour (SFT) is relatively common in the orbit and can show a range of cellularity and contain fat or stromal giant cells. Histologically, it consists of spindle cells with a thin-walled branching vascular pattern. All SFTs are positive for STAT6 on immunohistochemistry due to a chromosomal translocation (inv(12)(q13q13)), which results in a NAB2-STAT6 fusion product and overexpression of STAT6. The majority of these tumours follow a benign course but the behaviour is unpredictable. Larger and more cellular tumours are more likely to recur and metastasize.

Lymphoid tumours

Lymphomas of the ocular adnexa include lesions of the conjunctiva, eyelids, lacrimal gland and orbit. Those situated in the conjunctiva are associated with a lower incidence of systemic disease (20%) compared with those of the orbit (35%), lacrimal gland (40%) or eyelid (67%). Ocular lymphomas may be the first manifestation of disseminated disease and it is essential to undertake a full systemic and haematological examination of all patients presenting with ocular lymphoma. Immunohistochemistry (Fig. 9.49 and Box 9.14) and molecular techniques (see Boxes 9.6, 9.15 and 9.16) are important for identifying monoclonal populations and subclassifying lymphoma. Flow cytometry may be used to isolate and to phenotype different cells within a mixed population. Cells in suspension are labelled with fluorescent markers, excited by a laser and counted electronically by passing them through a flow cytometer: this may be undertaken on vitreous samples.

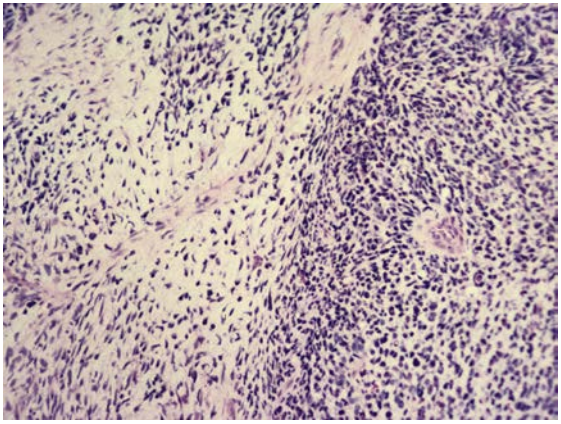


Fig. 9.48 In embryonal rhabdomyosarcoma, the cells are ovoid to spindle-shaped and there are alternating cellular and myxoid areas.

With these techniques, the vast majority of lymphoid proliferations can be classified as benign (reactive lymphoid hyperplasia) or malignant (lymphoma) and terms such as 'pseudolymphoma' should be avoided.

The most common ocular lymphoproliferative lesions include:

- Benign lymphoid hyperplasia, a similar process to reactive follicular hyperplasia of lymph nodes may form a tumour mass within the conjunctiva or orbit.
- Extranodal marginal zone lymphoma (EMZL) is the most common type of ocular lymphoma and is a low-grade B-cell lymphoma derived from the mucosal associated lymphoid tissue. EMZL usually follows an indolent course, but it may recur at other extranodal sites or rarely undergo transformation to high-grade lymphoma.
- Follicular lymphoma is identical to its nodal counterpart and in the majority of cases represents part of a systemic disease (see Video 9.3).

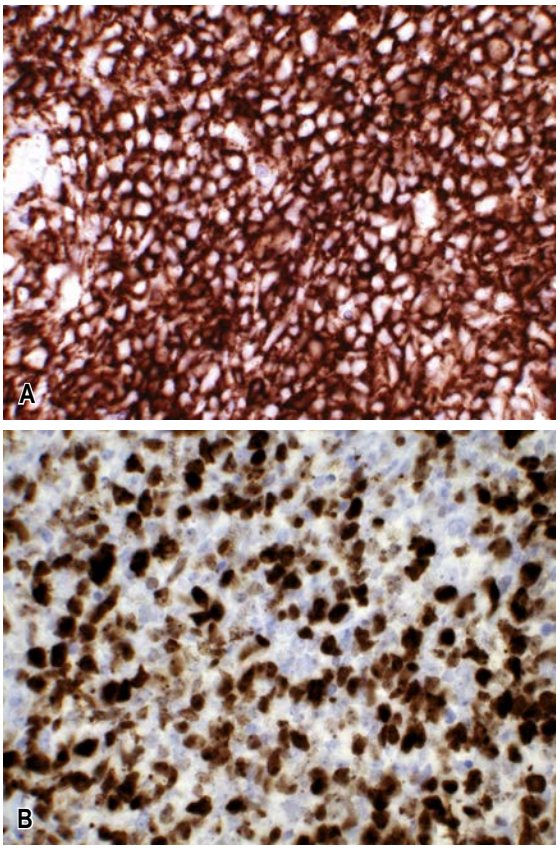


Fig. 9.49 (**A**) Immunohistochemistry for CD20 shows membranous staining of the cells in this B-cell lymphoma. (**B**) Immunohistochemistry for MIB-1 (proliferating cell nuclear antigen) shows widespread nuclear staining indicating a high proliferation fraction in this high-grade B-cell lymphoma (diffuse large B-cell lymphoma).

- Diffuse large B-cell lymphoma is less common; around 40% of cases are associated with systemic disease. These lymphomas tend to pursue an aggressive clinical course.
- Vitreoretinal (previously known as primary intraocular) lymphoma is a diffuse, large B-cell lymphoma that involves the retina, subretinal space, vitreous and optic nerve (Fig. 9.50). It can occur in conjunction with or independent of primary central nervous system lymphoma. These are rare lymphomas but there has been a dramatic increase in incidence in recent years. They usually occur in elderly patients but can also be associated with AIDS.

Many other lymphomas may uncommonly involve the ocular region, including mantle cell lymphoma, B-cell chronic lymphocytic leukaemia, Burkitt's lymphoma, peripheral T-cell lymphoma and natural killer cell lymphoma. Ocular Hodgkin's lymphoma is very rare. Leukaemic infiltration of the eye or orbit may also occur. In

Box 9.14 **Immunohistochemistry**

Immunohistochemistry is a method of detecting the presence of specific proteins in cells or tissues. It consists of the following steps:

- Antigen retrieval either by proteolytic digestion or microwaving the tissue sections.
- A primary antibody binds to a specific antigen.
- A secondary enzyme-conjugated antibody is then bound to this primary antibody–antigen complex.
- An appropriate substrate and chromogen are added and the enzyme catalyses the formation of a coloured deposit at the sites of antibody–antigen binding.
- Immunohistochemistry is important in:
  - *Tumour classification*: lymphomas are positive for common leucocyte antigen (CLA), carcinomas are positive for cytokeratins (CK), melanoma is positive for S100 and Melan A.
  - *Specific tumour subtyping*: lymphomas may be subtyped into B cell (CD20 positive) and T cell (CD3 positive); carcinoma of the lung is usually TTF-1 positive, carcinoma of the gastrointestinal tract is usually CDX-2 positive.
  - *Providing prognostic or therapeutic information*: lymphomas that have a high proliferation index with MIB-1 (see Fig. 9.49) are aggressive. Breast cancers that are oestrogen-receptor positive may be treated with hormonal therapy.
- *Diagnosis of infectious agents*: antibodies are available to various infectious agents such as herpes simplex virus and cytomegalovirus.

Box 9.15 **Tumour cytogenetics**

Numerical and structural chromosomal abnormalities occur in tumour cells compared with normal cells. Chromosomal abnormalities within a tumour cell can fall into one of three categories:

- Primary abnor*mality*: essential for establishing tumorigenesis, e.g. *Rb1* gene in retinoblastoma
- *Secondary abnormality*: a manifestation of tumour progression and clonal evolution, e.g. loss of *p53* in colonic adenoma–carcinoma sequence
- *Cytogenetic noise*: this is caused by genetic instability in any tumour.

For example, in lymphoma:

- in extranodal marginal zone lymphoma t(11;18) and t(1;14) are helpful in diagnosis
- in follicular lymphoma t(14;18) is helpful in diagnosis
- in diffuse large B-cell lymphoma, tumours with chromosomal translocations that result in rearrangements of *C-MYC* and *BCL2* have a poor prognosis.

Box 9.16 ***In situ* hybridization**

In this technique, single-stranded complementary nucleic acid sequences can join with specific DNA or RNA sequences in cells or tissues and these hybridization sites can be identified by the addition of a fluorescent *in situ* hybridization (FISH) or enzyme-labelled probe.

- FISH is used to detect chromosomal gains, losses and translocations in neoplasms (particularly lymphoma).
- *In situ* hybridization may be used for assessing clonality in lymphoid proliferations using probes for Kappa and Lambda light chains and detecting virus nucleic acids in infections with Epstein–Barr virus or cytomegalovirus.

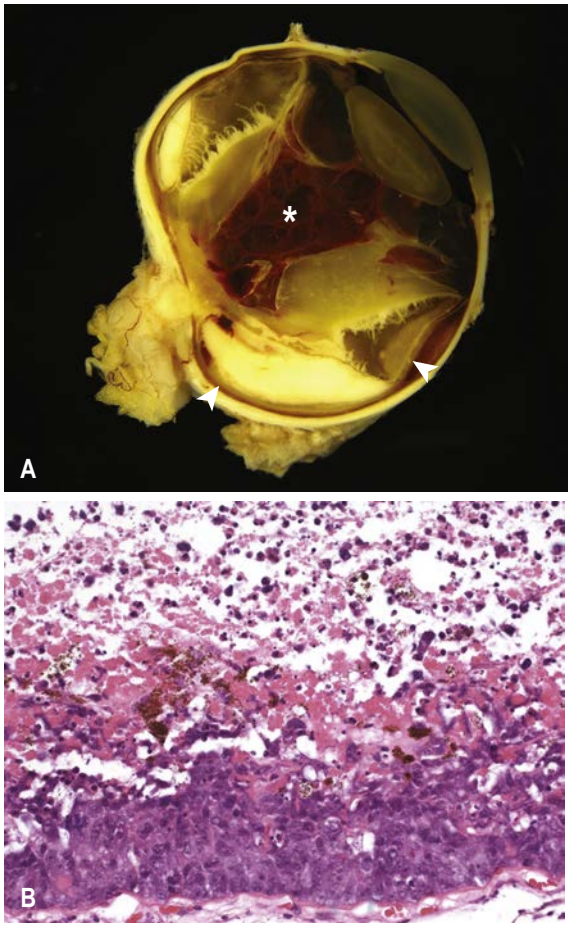


Fig. 9.50 (**A**) An eye removed for vitreoretinal lymphoma. There is haemorrhage and exudate in the vitreous *(*)*. There is creamy white tumour overlying the retina *(arrowheads)*. (**B**) On histology the retina is replaced by pleomorphic large lymphoma cells, with necrosis and apoptosis extending into the vitreous.

particular, granulocytic sarcoma can present as an isolated orbital mass in an otherwise healthy child.

Lacrimal gland tumours

Pleomorphic adenoma (benign mixed tumour)

Pleomorphic adenoma is the most common epithelial tumour of the lacrimal gland. It usually occurs in late to middle age but may occur at any age. This tumour is slow growing and pseudoencapsulated with surface bosselations. Histologically, it consists of a mixture of epithelial and mesenchymal elements, including myxoid tissue, cartilage, fat and, rarely, bone (Fig. 9.51A). It is important to excise these tumours completely because if they are 'shelled out' they can recur and there is a small but significant risk of surviving residual tumour undergoing malignant change to a pleomorphic carcinoma (malignant mixed tumour).

Adenoid cystic carcinoma

This is the second most common epithelial neoplasm of the lacrimal gland after pleomorphic adenoma. Although it is usually diagnosed in middle-aged or older patients, it frequently occurs in younger patients as well. The history is shorter than for pleomorphic adenoma and the patient may present with proptosis, numbness, pain and diplopia because invasion of nerves and extraocular muscles may occur early in tumour development. Histologically, these tumours can assume a range of patterns, the most common of which is a cribriform or 'Swiss-cheese' pattern (Fig. 9.51B). The pain and numbness may be explained by the tumour's propensity for perineural infiltration. These are aggressive neoplasms that require radical surgery.

Other malignant epithelial tumours

A small number of epithelial tumours of the lacrimal gland are adenocarcinomas arising *de novo* with no evidence of a pre-existing benign mixed tumour. Mucoepidermoid carcinoma is another rare form of carcinoma that may arise in the lacrimal gland.

Lacrimal sac tumours

Tumours of the lacrimal sac are uncommon and are usually of epithelial origin. Papillomas may show an exophytic, inverted or mixed growth pattern and the epithelium may be of squamous or transitional cell type. Carcinoma of the lacrimal sac may develop within a papilloma or arise *de novo*. These are locally aggressive tumours and if neglected can invade surrounding structures.

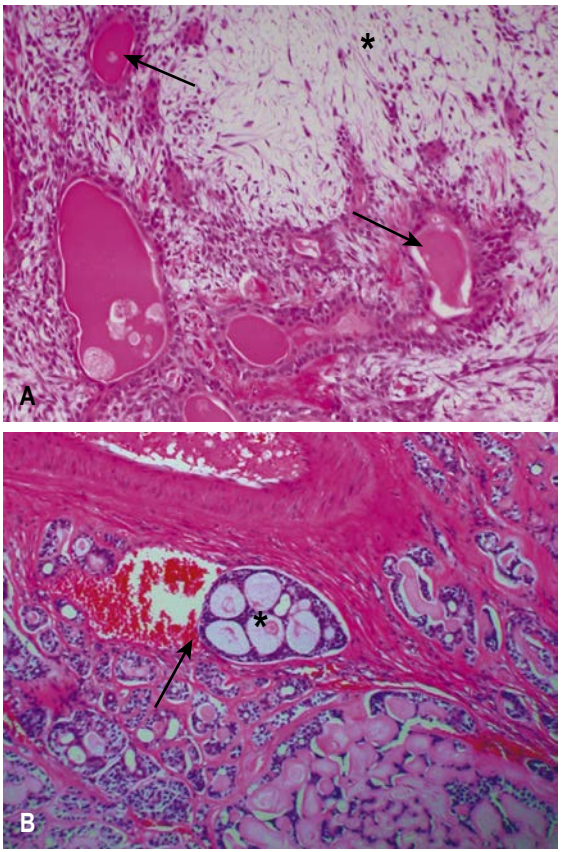


Fig. 9.51 (**A**) Pleomorphic adenoma consists of benign branching glands *(arrows)* lying within a myxoid stroma *(*)*. (**B**) Adenoid cystic carcinoma has a Swiss-cheese pattern *(*)* and may show vascular *(arrow)* and perineural invasion.

Metastatic tumours

In adults, metastatic tumours most commonly involve the uveal tract (Fig. 9.42D). Orbital involvement occurs about one-tenth as often and metastases to the eyelid and conjunctiva are rare. Most metastatic tumours are carcinomas and the most common primary sites are usually the lung, breast, prostate or gastrointestinal tract, but metastases from a wide range of primary carcinomas have been described. Immunohistochemistry (Box 9.14) and novel molecular techniques (see Box 9.17) can be useful in identifying the primary site. In children, metastatic disease is usually orbital and uveal involvement is rare. Orbital involvement by neuroblastoma, Ewing's sarcoma, Wilms' tumour and rhabdomyosarcoma may occur.

Disorders misdiagnosed as neoplasms

There are several 'lumps and bumps' around the eye and orbit that can be mistaken for tumours.

Box 9.17 Next generation sequencing

Next generation sequencing (NGS) describes a group of sequencing technologies that enable fast, accurate and cost-effective sequencing. Techniques under the term NGS include:

Gene panels
- A discrete number of genes of interest are sequenced.
- Useful in hereditary breast and ovarian cancer
- Useful for identifying drug targetable mutations in certain cancers

Whole exome sequencing
- Sequencing of the protein coding region of genes in the genome
- Allows identification of genetic variants that alter protein sequence
- Useful for gene mapping in rare disorders

Whole genome sequencing
- Sequencing of the entire DNA of the genome at a single time.
- Characterizing inherited disorders
- Characterizing driver mutations in cancer
- Tracking disease outbreaks

In metastatic carcinoma where the primary is unknown, various NGS techniques can be applied to identify primary site (because of site-specific mutations) and to identify mutations that will allow patients to receive genotype-matched targeted therapy.

Cysts

Eyelid. Simple cysts are common in the eyelid:
- *Sudoriferous cysts* or sweat gland cysts (hidrocystomas) are derived from the ducts of the glands of Moll. The cysts are thin-walled and appear as translucent or bluish swellings at the lid margin. Histologically, they are lined by a double layer of epithelium with an inner layer of cuboidal cells and an outer layer of myoepithelial cells.
- *Epidermoid cysts* may occur secondary to obstruction of the duct of a pilosebaceous follicle or as the result of epithelial inclusion following trauma or surgery. Epidermoid cysts are lined by keratinizing squamous epithelium and are filled with keratin. A foreign body giant cell reaction may be seen in relation to cyst rupture.
- *Dermoid cysts* occur in children as a result of the incarceration of ectoderm between the frontal and maxillary process during embryogenesis. Dermoid cysts contain hairs and pilosebaceous follicles.

Conjunctiva. Cystic lesions are also common in the conjunctiva. Epithelial inclusion cysts are the result of previous trauma with incarceration of conjunctival epithelial nests in the stroma. Lymphatic cysts occur when ectatic lymphatics coalesce. This lymphangiectasia may

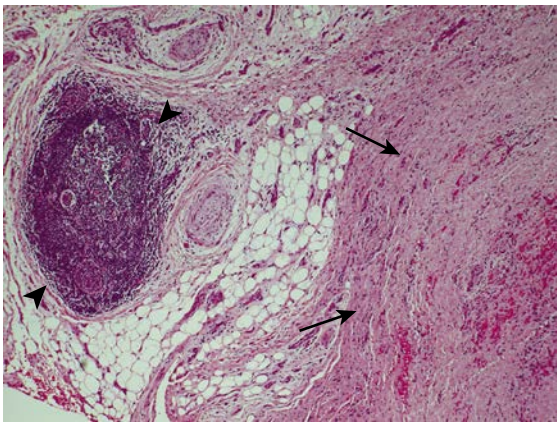


Fig. 9.52 In idiopathic orbital inflammation the orbital fatty tissue is replaced by fibrous tissue *(arrows)*. There is an accompanying inflammatory cell infiltrate that may include lymphoid follicles *(arrowheads)*.

occur as a hamartomatous malformation or secondary to inflammation.

Orbit. Cysts may also occur in the orbit and computed tomography scans may be helpful in identifying some lesions. Dermoid cysts may also occur in the orbit and should be excised intact because leaked contents may induce a granulomatous inflammatory reaction. A mucocoele is an expansion of the paranasal space secondary to drainage obstruction from chronic sinusitis. It consists of a cystic cavity lined with epithelium, which sometimes contains goblet cells. Haematic cyst is an organizing haematoma, which can occur spontaneously, or following blunt trauma.

Idiopathic orbital inflammation

Idiopathic orbital inflammation (formerly inflammatory pseudotumour) is a non-granulomatous inflammatory process within the orbit for which there is no recognized local cause or any underlying systemic disease. The disease presents as a unilateral or bilateral mass, which clinically may be mistaken for tumour. In biopsy specimens, an early lesion shows oedema of orbital tissues and an inflammatory infiltrate composed predominantly of lymphocytes and plasma cells and lymphoid follicles may be present. As the disease progresses, collagen is laid down and the collections of inflammatory cells may be separated by fibrous tissue (Fig. 9.52). Most cases show a dramatic response to corticosteroid therapy unless the lesion has extensive fibrosis. Other immunosuppressive agents, such as azathioprine, or low-dose radiotherapy may be used in patients who fail to respond to steroids. Idiopathic sclerosing inflammation is a distinct form of orbital inflammation characterized by a slow and relentless fibrosing process, with progressive involvement of orbital structures. It may be part of a multisystem disease with progressive fibrosis at other sites. In both sclerosing and non-sclerosing inflammation, raised serum IgG4 and increased IgG4 positive plasma cells may be found in the tissues in some cases (see Chapter 7, p. 406). This IgG4-related disease may be bilateral and affect other sites but is usually responsive to steroids.

Further reading is available online at https://expertconsult.inkling.com

Further reading

Apel, F., Zychlinsky, A., Kenny, E.F., 2018. The role of neutrophil extracellular traps in rheumatic diseases. Nat. Rev. Rheumatol. 14, 467–475. https://doi.org/10.1038/s41584-018-0039-z.

Forrester, J.V., McMenamin, P.G., Dando, S.J., 2018. CNS infection and immune privilege. Nat. Rev. Neurosci. 19, 655–671. https://doi.org/10.1038/s41583-018-0070-8.

Moutray, T., et al., 2018. Different lasers and techniques for proliferative diabetic retinopathy. Cochrane Database Syst. Rev. 3. https://doi.org/10.1002/14651858.CD012314.pub2. CD012314.

Roberts, F., MacDuff, E., 2018. Pathology Illustrated. Elsevier Churchill Livingstone, Edinburgh.

Roberts, F., Thum, C.K., 2014. Lee's Ophthalmic Histopathology. Springer, London.

Sehu, K.W., Lee, W.R., 2005. Ophthalmic Pathology: An Illustrated Guide for Clinicians. Blackwell Publishing, Oxford.

Vilas, G.L., Morgan, P.E., Loganathan, S.K., Quon, A., Casey, J.R., 2011. A biochemical framework for SLC4A11, the plasma membrane protein defective in corneal dystrophies. Biochemistry. 50, 2157–2169. https://doi.org/10.1021/bi101887z.

Yanoff, M., Fine, B.S., 2008. Ocular Pathology, fifth ed. Elsevier Mosby-Wolfe, London.

Zhang, H., et al., 2017. Immunoinhibitory checkpoint deficiency in medium and large vessel vasculitis. Proc. Natl. Acad. Sci. U. S. A. 114, E970–E979. https://doi.org/10.1073/pnas.1616848114. 2017.

Addendum

COVID-19 and the eye

Severe acute respiratory syndrome-coronavirus-2 (SARS-CoV-2) is a beta-coronavirus that emerged in China as the etiological agent of the potentially fatal viral pneumonia termed coronavirus disease 2019 (COVID-19). In March 2020, the WHO classified COVID-19 as a pandemic that by July 2020 had infected >14 million individuals worldwide, resulting in >600,000 deaths. Although infection primarily occurs through the respiratory tract, there are also reports that the virus can infect conjunctival epithelial cells and reach the respiratory through the nasopharynx. Conjunctival epithelial cells express the ACE2 receptor for the SARS-CoV-2 spike (S) protein, and the serine protease TMPRSS2 that is required for S-protein modification to enable viral infection. Conjunctival secretions and tears of COVID-19 patients were also found to test positive by the reverse transcriptase-PCR SARS-CoV-2-RNA assay. Patients can also present with conjunctivitis, including chemosis and excessive tear formation (and rarely with keratoconjunctivitis); those patients with conjunctivitis tend to also have more serious pulmonary disease. The inflammatory response in the conjunctiva is likely a consequence of the cellular infiltrate and cytokine storm that is associated with severe pulmonary disease. In addition to the conjunctiva, retinal changes were detected in 12 adults (six men and six women aged 25–69 years) that were examined 11–33 days after onset of COVID-19 symptoms. Of these, four patients presented with subtle cotton wool spots and micro-hemorrhages (see Figure). The underlying cause of these abnormalities has yet to be identified but may be related to breakdown of the blood-brain barrier following the cytokine storm.

At the time this book went to press in mid-2020, there were relatively few papers describing ocular abnormalities in infected individuals, and the evidence to date indicates that the ocular surface is not a major site for initial COVID-19 infection. Further, ocular manifestations are not a major outcome of SARS-CoV-2 infection, with only 0.04% of >20,000 COVID-19 patients in the United Kingdom reporting conjunctivitis among the symptoms. However, as this pandemic continues, there will be further reports of ocular involvement, and these conclusions may change.

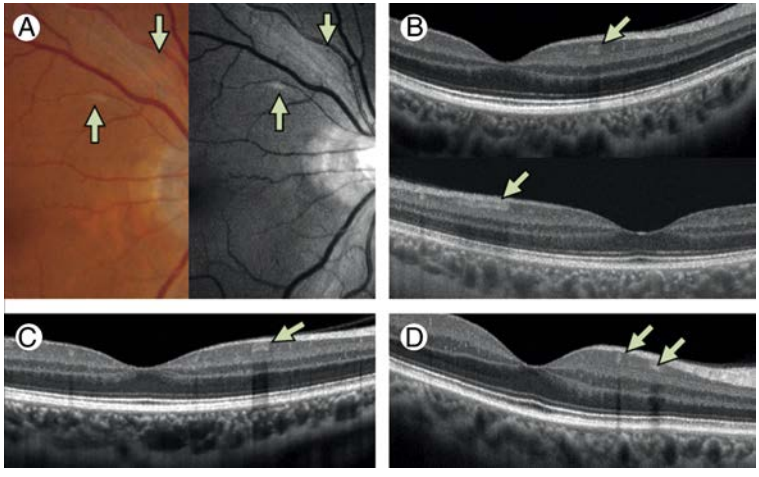


Figure Retinal Findings in four patients with COVID-19 (Panels A–D represent four different patients. (A) Colour fundus photography and red-free imaging show a cotton wood spot at the superior retinal arcade with subtle microhaemorrhage. (B–D) Cross-sectional B-scan OCT in three patients shows the hyper-reflective lesions at the level of the inner plexiform and ganglion cell layers, a feature observed in all patients. *COVID-19*, Coronavirus disease 2019; *OCT*, optical coherence tomography.)

Further reading

1. Gupta, A., Madhavan, M.V., Sehgal, K., Nair, N., Mahajan, S., Sehrawat, T.S., et al., 2020. Extrapulmonary manifestations of COVID-19. Nat. Med. 26, 1017–1032.
2. Ma, D., Chen, C.B., Jhanji, V., Xu, C., Yuan, X.L., Liang, J.J., et al., 2020. Expression of SARS-CoV-2 receptor ACE2 and TMPRSS2 in human primary conjunctival and pterygium cell lines and in mouse cornea. Eye (Lond) 34, 1212–1219.
3. Zhou, L., Xu, Z., Castiglione, G.M., Soiberman, U.S., Eberhart, C.G., Duh, E.J., 2020. ACE2 and TMPRSS2 are expressed on the human ocular surface, suggesting susceptibility to SARS-CoV-2 infection. Ocul. Surf.
4. Marinho, P.M., Marcos, A.A.A., Romano, A.C., Nascimento, H., Belfort Jr., R., 2020. Retinal findings in patients with COVID-19. Lancet 395, 1610.
5. Docherty, A.B., Harrison, E.M., Green, C.A., Hardwick, H.E., Pius, R., Norman, L., I. C. investigators, et al., 2020. Features of 20 133 UK patients in hospital with covid-19 using the ISARIC WHO Clinical Characterisation Protocol: prospective observational cohort study. BMJ 369, m1985.
6. Willcox, M.D., Walsh, K., Nichols, J.J., Morgan, P.B., Jones, L.W., 2020. The ocular surface, coronaviruses and COVID-19. Clin. Exp. Optom. 103, 418–424.