Methods in Molecular Biology 2571

Springer Protocols

Raúl González-Domínguez *Editor*

Mass Spectrometry for Metabolomics

Humana Press

METHODS IN MOLECULAR BIOLOGY

Series Editor
John M. Walker
School of Life and Medical Sciences
University of Hertfordshire
Hatfield, Hertfordshire, UK

For further volumes:
http://www.springer.com/series/7651

For over 35 years, biological scientists have come to rely on the research protocols and methodologies in the critically acclaimed *Methods in Molecular Biology* series. The series was the first to introduce the step-by-step protocols approach that has become the standard in all biomedical protocol publishing. Each protocol is provided in readily-reproducible step-by-step fashion, opening with an introductory overview, a list of the materials and reagents needed to complete the experiment, and followed by a detailed procedure that is supported with a helpful notes section offering tips and tricks of the trade as well as troubleshooting advice. These hallmark features were introduced by series editor Dr. John Walker and constitute the key ingredient in each and every volume of the *Methods in Molecular Biology* series. Tested and trusted, comprehensive and reliable, all protocols from the series are indexed in PubMed.

Mass Spectrometry for Metabolomics

Edited by

Raúl González-Domínguez

Instituto de Investigación e Innovación Biomédica de Cádiz, Hospital Universitario Puerta del Mar, Cádiz, Spain

Humana Press

Editor
Raúl González-Domínguez
Instituto de Investigación
e Innovación Biomédica de Cádiz
Hospital Universitario Puerta del Mar
Cádiz, Spain

ISSN 1064-3745 ISSN 1940-6029 (electronic)
Methods in Molecular Biology
ISBN 978-1-0716-2698-6 ISBN 978-1-0716-2699-3 (eBook)
https://doi.org/10.1007/978-1-0716-2699-3

© The Editor(s) (if applicable) and The Author(s), under exclusive license to Springer Science+Business Media, LLC, part of Springer Nature 2023
This work is subject to copyright. All rights are solely and exclusively licensed by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
The publisher, the authors, and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Humana imprint is published by the registered company Springer Science+Business Media, LLC, part of Springer Nature.
The registered company address is: 1 New York Plaza, New York, NY 10004, U.S.A.

Preface

Mass spectrometry (MS) has become the most widely employed analytical technique in metabolomics because of its sensitivity, selectivity, and broad coverage thanks to the availability of multiple instrumental configurations. This book provides a comprehensive overview of state-of-the-art metabolomics methods based on MS, and their application in food, nutrition, and biomedical research. In this volume of the Methods in Molecular Biology series, 22 chapters are assembled covering hot topics related to sample preparation, chromatographic and electrophoretic separation, MS-based analysis, as well as data processing and analysis.

The first part of this book is comprised of innovative metabolomics approaches in food science and nutrition research. Chapter 1 presents a novel method based on reversed-phase liquid chromatography (RPLC) coupled with diode array detector and MS for high-throughput, comprehensive, and quantitative fingerprinting of a broad range of phenolic compounds and related metabolites in different food products. In Chaps. 2 and 3, the authors explain how to analyze volatile metabolites in vinegar samples by applying solid-phase microextraction (SPME) and stir bar sorptive extraction (SBSE) techniques in combination with gas chromatography–mass spectrometry (GC-MS). Chapter 4 describes the application of untargeted metabolomics based on liquid chromatography coupled to high resolution mass spectrometry (LC-HRMS) for the discovery of food intake biomarkers in blood and urine samples. Finally, Chap. 5 illustrates the potential of metabolomics to investigate the anti-proliferative capacity of bioactive compounds from fruit by-products against colon cancer cells.

In the second part, we gather 14 chapters focused on the application of metabolomics in biomedical research. Chapters 6, 7 and 8 describe complementary MS-based hyphenated platforms for untargeted metabolomics of common biological fluids (e.g., blood, urine), namely reversed-phase liquid chromatography (RPLC, Chap. 6), hydrophilic interaction liquid chromatography (HILIC, Chap. 7), and capillary electrophoresis (CE, Chap. 8). In relation to CE-based metabolomics, Chaps. 9 and 10 present novel analytical developments for direct analysis of highly saline samples using in-capillary preconcentration and for enhancing the analysis of acidic metabolites using chemical derivatization, respectively. Chapter 11 outlines a metabolomics multi-platform, based on the combination of RPLC and HILIC, to capture the plasma and erythrocyte metabolomes behind childhood obesity and insulin resistance. In this vein, Chap. 12 details how to characterize the plasmatic and erythroid multi-elemental biodistribution in childhood obesity using a method based on protein precipitation under non-denaturing conditions and further analysis by inductively coupled plasma mass spectrometry. The last chapters of this part of the book deal with the analysis of other tissues and biological samples. Chapter 13 presents an untargeted metabolomics workflow for brain tissue analysis based on two LC-MS methods combining reversed-phase and HILIC chromatography, whereas Chap. 14 focuses on the targeted determination of cholesterol in liver by GC-MS. The next three chapters revolve around the potential of MS to characterize ocular samples. In Chap. 15, the authors describe a simple method for the preparation and metabolomics analysis of aqueous humor samples using RPLC-MS and HILIC-MS. Chapter 16 tackles the analysis of serotonin and L-DOPA in ocular tissues by imaging mass spectrometry. On the other hand, the method described

under Chap. 17 explains how to conduct isobaric incorporation of C13-histidine for the assessment of remyelination in the optic nerve. Chapter 18 details an innovative workflow for the isolation and lipidomics analysis of extracellular vesicles from human breast milk. Finally, Chap. 19 introduces a robust LC-HRMS-based metabolomics and lipidomics platform for the investigation of cultured hepatic cell lines.

To conclude, this book also includes three chapters addressing the importance of data processing and statistical analysis in MS-based metabolomics. Chapter 20 reviews the available tools and resources for conducting each of the steps usually required in metabolomics data processing, namely data preprocessing, data analysis, annotation, and functional analysis. In Chap. 21, a typical data processing procedure is provided for the analysis and quality control of targeted metabolomics LC-MS-based data. The final chapter presents an instrument-agnostizing methodology for LC-MS aimed to provide standardized and comparable chromatographic fingerprints regardless of the instrument.

In summary, this volume in the Methods in Molecular Biology series represents a collection of the expertise of leading scientists in the field of metabolomics and mass spectrometry research. Considering its multidisciplinary scope, I hope a broad target audience will benefit of this timely book, including chemists, biochemists, biologists, nutritionists, clinicians, and other experts working in the growing and exciting field of metabolomics.

Cádiz, Spain *Raúl González-Domínguez*

Contents

Preface . . . . . *v*
Contributors . . . . . *xi*

1 Comprehensive and High-Throughput Method for Quantitative Fingerprinting of Phenolic Compounds 1
Raúl González-Domínguez, Ana Sayago, María Santos-Martín, and Ángeles Fernández-Recamales

2 Determination of Volatile Metabolites in Vinegar by Solid-Phase Microextraction–Gas Chromatography–Mass Spectrometry (SPME–GC–MS) 13
Remedios Castro and Enrique Durán-Guerrero

3 Determination of Volatile Metabolites in Vinegar by Stir Bar Sorptive Extraction–Gas Chromatography–Mass Spectrometry (SBSE–GC–MS) 21
Enrique Durán-Guerrero and Remedios Castro

4 Discovery of Food Intake Biomarkers Using Metabolomics 33
Leticia Lacalle-Bergeron, David Izquierdo-Sandoval, Juan V. Sancho, and Tania Portolés

5 Metabolomic Characterization of the Antiproliferative Activity of Bioactive Compounds from Fruit By-Products Against Colon Cancer Cells 45
Gerardo Alvarez-Rivera, Alberto Valdés, and Alejandro Cifuentes

6 Untargeted Metabolomics by Liquid Chromatography–Mass Spectrometry in Biomedical Research 57
Caridad Díaz and Carmen González-Olmedo

7 Untargeted Metabolomics for Disease-Specific Signatures 71
Constantina Chalikiopoulou, José Carlos Gómez-Tamayo, and Theodora Katsila

8 Metabolomics Analysis of Blood, Urine, and Saliva Samples Based on Capillary Electrophoresis–Mass Spectrometry 83
Masahiro Sugimoto and Yumi Aizawa

9 Capillary Electrophoresis–Mass Spectrometry for the Direct Analysis of Metabolites in Highly Saline Samples Using In-Capillary Preconcentration 95
Marlien van Mever and Rawi Ramautar

10 Chemical Derivatization to Enhance Capillary Electrophoresis Mass Spectrometric Analysis of Acidic Metabolites in Mammalian Cells 105
Marlien van Mever and Rawi Ramautar

11 Untargeted Metabolomics Based on Liquid Chromatography–Mass Spectrometry for the Analysis of Plasma and Erythrocyte Samples in Childhood Obesity 115
Álvaro González-Domínguez, Marina Armeni, Otto Savolainen, Alfonso María Lechuga-Sancho, Rikard Landberg, and Raúl González-Domínguez

12 Characterization of the Plasmatic and Erythroid Multielemental Biodistribution in Childhood Obesity Using a High-Throughput Method for Size Fractionation of Metal Species. 123
Álvaro González-Domínguez, María Millán-Martínez, Daniel Sánchez-Rodas, Alfonso María Lechuga-Sancho, and Raúl González-Domínguez

13 Analysis of Brain Metabolites Using Two Complementary Ultrahigh-Performance Liquid Chromatography–Mass Spectrometry Methods 133
Alexa M. Jauregui, Sofia E. Parellada, Emily Neag, and Sanjoy K. Bhattacharya

14 Analysis of Cholesterol from the Liver Using Gas Chromatography–Mass Spectrometry 143
Meher A. Saleem, Betsy Benitez, Charles Yaros, Gabrielle Yamar, and Sanjoy K. Bhattacharya

15 Metabolomics Analysis of Aqueous Humor Based on Liquid Chromatography–Mass Spectrometry 149
Karolina Pietrowska, Diana Anna Dmuchowska, Adam Kretowski, and Michal Ciborowski

16 Analyses and Localization of Serotonin and L-DOPA in Ocular Tissues by Imaging Mass Spectrometry 157
Varun Krishnan, Sean Meehan, Colin Hayter, and Sanjoy K. Bhattacharya

17 Isobaric Incorporation of C13-Histidine for the Assessment of Remyelination 169
Faith Christine Harvey, Anddre Osmar Valdivia, Colin Hayter, and Sanjoy K. Bhattacharya

18 Isolation and Lipidomic Screening of Human Milk Extracellular Vesicles 177
Victoria Ramos-Garcia, Isabel Ten-Doménech, Abel Albiach-Delgado, Marta Gómez-Ferrer, Pilar Sepúlveda, Anna Parra-Llorca, Laura Campos-Berga, Alba Moreno-Giménez, Guillermo Quintás, and Julia Kuligowski

19 Mass Spectrometry-Based Untargeted Metabolomics and Lipidomics Platforms to Analyze Cell Culture Extracts. 189
Elias Iturrospe, Katyeny Manuela da Silva, Maria van de Lavoir, Rani Robeyns, Matthias Cuykx, Tamara Vanhaecke, Alexander L. N. van Nuijs, and Adrian Covaci

20 Data Processing and Analysis in Mass Spectrometry-Based Metabolomics 207
Ángela Peralbo-Molina, Pol Solà-Santos, Alexandre Perera-Lluna, and Eduardo Chicano-Gálvez

21 Data Processing and Analysis in Liquid Chromatography–Mass Spectrometry-Based Targeted Metabolomics 241
Masahiro Sugimoto, Yumi Aizawa, and Atsumi Tomita

22 Instrument-Agnostizing Methodology for Liquid Chromatography–Mass Spectrometry Systems 257
Rosalía López-Ruíz, Sandra Martín-Torres, Ana M. Jiménez-Carvelo, Roberto Romero-González, and Luis Cuadros-Rodríguez

Index *271*

Contributors

Yumi Aizawa • *Institute of Medical Science, Tokyo Medical University, Tokyo, Japan*

Abel Albiach-Delgado • *Neonatal Research Group, Health Research Institute La Fe, Valencia, Spain*

Gerardo Alvarez-Rivera • *Laboratory of Foodomics, Institute of Food Science Research, CIAL, CSIC, Madrid, Spain*

Marina Armeni • *Division of Food and Nutrition Science, Department of Biology and Biological Engineering, Chalmers University of Technology, Gothenburg, Sweden; Department of Biology and Biological Engineering, Chalmers Mass Spectrometry Infrastructure, Chalmers University of Technology, Gothenburg, Sweden*

Betsy Benitez • *Bascom Palmer Eye Institute, University of Miami Miller School of Medicine, Miami, FL, USA; Miami Integrative Metabolomics Research Center, Miami, FL, USA*

Sanjoy K. Bhattacharya • *Department of Ophthalmology, Bascom Palmer Eye Institute, Miller School of Medicine at University of Miami, Miami, FL, USA; Miami Integrative Metabolomics Research Center, Miami, FL, USA; University of Miami Miller School of Medicine, Miami, FL, USA*

Laura Campos-Berga • *Neonatal Research Group, Health Research Institute La Fe, Valencia, Spain*

Remedios Castro • *Analytical Chemistry Department, Faculty of Sciences-IVAGRO, University of Cádiz, Agrifood Campus of International Excellence (ceiA3), Puerto Real, Cádiz, Spain*

Constantina Chalikiopoulou • *Institute of Chemical Biology, National Hellenic Research Foundation, Athens, Greece*

Eduardo Chicano-Gálvez • *IMIBIC Mass Spectrometry and Molecular Imaging Unit, Maimonides, Biomedical Research Institute of Cordoba (IMIBIC), Reina Sofia University Hospital, University of Cordoba (UCO), Córdoba, Spain*

Michal Ciborowski • *Metabolomics Laboratory, Clinical Research Centre, Medical University of Bialystok, Bialystok, Poland*

Alejandro Cifuentes • *Laboratory of Foodomics, Institute of Food Science Research, CIAL, CSIC, Madrid, Spain*

Adrian Covaci • *Toxicological Centre, University of Antwerp, Antwerp, Belgium*

Luis Cuadros-Rodríguez • *Department of Analytical Chemistry, University of Granada, Granada, Spain*

Matthias Cuykx • *Laboratory of Clinical Medicine, Antwerp University Hospital, Edegem, Belgium*

Katyeny Manuela da Silva • *Toxicological Centre, University of Antwerp, Antwerp, Belgium*

Caridad Díaz • *Fundación MEDINA, Centro de Excelencia en Investigación de Medicamentos Innovadores en Andalucía, Granada, Spain*

Diana Anna Dmuchowska • *Department of Ophthalmology, Medical University of Bialystok, Bialystok, Poland*

Enrique Durán-Guerrero • *Analytical Chemistry Department, Faculty of Sciences-IVAGRO, University of Cádiz, Agrifood Campus of International Excellence (ceiA3), Puerto Real, Cádiz, Spain*

Ángeles Fernández-Recamales • *Agrifood Laboratory, Faculty of Experimental Sciences, University of Huelva, Huelva, Spain; International Campus of Excellence CeiA3, University of Huelva, Huelva, Spain*

Marta Gómez-Ferrer • *Regenerative Medicine and Heart Transplantation Unit, Health Research Institute La Fe, Valencia, Spain*

José Carlos Gómez-Tamayo • *GRIB (IMIM/UPF), PRBB, Barcelona, Spain*

Álvaro González-Domínguez • *Instituto de Investigación e Innovación Biomédica de Cádiz (INiBICA), Hospital Universitario Puerta del Mar, Universidad de Cádiz, Cádiz, Spain*

Raúl González-Domínguez • *Instituto de Investigación e Innovación Biomédica de Cádiz (INiBICA), Hospital Universitario Puerta del Mar, Universidad de Cádiz, Cádiz, Spain*

Carmen González-Olmedo • *Medical Oncology Unit, University Hospital of Jaén, Jaén, Spain*

Faith Christine Harvey • *Department of Ophthalmology, Bascom Palmer Eye Institute, Miller School of Medicine at University of Miami, Miami, FL, USA; Miami Integrative Metabolomics Research Center, Miami, FL, USA*

Colin Hayter • *Bascom Palmer Eye Institute, Miller School of Medicine, University of Miami, Miami, FL, USA; Miami Integrative Metabolomics Research Center, Miami, FL, USA; University of Miami Miller School of Medicine, Miami, FL, USA*

Colin Hayter • *University of Miami, Miami, FL, USA*

Elias Iturrospe • *Toxicological Centre, University of Antwerp, Antwerp, Belgium; Department of In Vitro Toxicology and Dermato-Cosmetology, Vrije Universiteit Brussel, Brussels, Belgium*

David Izquierdo-Sandoval • *Enviromental and Public Health Analytical Chemistry, Research Institute for Pesticides and Water (IUPA), Universitat Jaume I, Castellón de la Plana, Spain*

Alexa M. Jauregui • *Bascom Palmer Eye Institute, Miller School of Medicine at University of Miami, Miami, FL, USA; Miami Integrative Metabolomics Research Center, Miami, FL, USA; University of Miami Miller School of Medicine, Miami, FL, USA*

Ana M. Jiménez-Carvelo • *Department of Analytical Chemistry, University of Granada, Granada, Spain*

Theodora Katsila • *Institute of Chemical Biology, National Hellenic Research Foundation, Athens, Greece*

Adam Kretowski • *Metabolomics Laboratory, Clinical Research Centre, Medical University of Bialystok, Bialystok, Poland*

Varun Krishnan • *Bascom Palmer Eye Institute, Miller School of Medicine, University of Miami, Miami, FL, USA; Miami Integrative Metabolomics Research Center, Miami, FL, USA; University of Miami Miller School of Medicine, Miami, FL, USA*

Julia Kuligowski • *Neonatal Research Group, Health Research Institute La Fe, Valencia, Spain*

Leticia Lacalle-Bergeron • *Enviromental and Public Health Analytical Chemistry, Research Institute for Pesticides and Water (IUPA), Universitat Jaume I, Castellón de la Plana, Spain*

Rikard Landberg • *Division of Food and Nutrition Science, Department of Biology and Biological Engineering, Chalmers University of Technology, Gothenburg, Sweden*

Alfonso María Lechuga-Sancho • *Instituto de Investigación e Innovación Biomédica de Cádiz (INiBICA), Hospital Universitario Puerta del Mar, Universidad de Cádiz, Cádiz, Spain; Departamento Materno Infantil y Radiología, Facultad de Medicina, Universidad de Cádiz, Cádiz, Spain; Unidad de Endocrinología Pediátrica y Diabetes, Servicio de Pediatría, Hospital Universitario Puerta del Mar, Cádiz, Spain*

Rosalía López-Ruíz • *Department of Chemistry and Physics, Research Centre for Mediterranean Intensive Agrosystems and Agri-Food Biotechnology (CIAIMBITAL), University of Almeria, Agri-food International Campus of Excellence, CeiA3, Almeria, Spain*

Sandra Martín-Torres • *Department of Analytical Chemistry, University of Granada, Granada, Spain*

Sean Meehan • *Bascom Palmer Eye Institute, Miller School of Medicine, University of Miami, Miami, FL, USA; Miami Integrative Metabolomics Research Center, Miami, FL, USA; University of Miami Miller School of Medicine, Miami, FL, USA*

María Millán-Martínez • *Associate Unit CSIC-University of Huelva "Atmospheric Pollution", Center for Research in Sustainable Chemistry – CIQSO, University of Huelva, Huelva, Spain; Department of Chemistry, Faculty of Experimental Sciences, University of Huelva, Huelva, Spain*

Alba Moreno-Giménez • *Neonatal Research Group, Health Research Institute La Fe, Valencia,, Spain*

Emily Neag • *Bascom Palmer Eye Institute, Miller School of Medicine at University of Miami, Miami, FL, USA; Miami Integrative Metabolomics Research Center, Miami, FL, USA; University of Miami Miller School of Medicine, Miami, FL, USA; College of Osteopathic Medicine, Michigan State University, East Lansing, MI, USA*

Sofia E. Parellada • *Bascom Palmer Eye Institute, Miller School of Medicine at University of Miami, Miami, FL, USA; Miami Integrative Metabolomics Research Center, Miami, FL, USA; University of Miami Miller School of Medicine, Miami, FL, USA*

Anna Parra-Llorca • *Neonatal Research Group, Health Research Institute La Fe, Valencia, Spain*

Ángela Peralbo-Molina • *IMIBIC Mass Spectrometry and Molecular Imaging Unit, Maimonides, Biomedical Research Institute of Cordoba (IMIBIC), Reina Sofia University Hospital, University of Cordoba (UCO), Córdoba, Spain*

Alexandre Perera-Lluna • *B2SLab, Departament d'Enginyeria de Sistemes, Automàtica i Informàtica Industrial, Universitat Politècnica de Catalunya, Barcelona, Spain; Networking Biomedical Research Centre in the Subject Area of Bioengineering, Biomaterials and Nanomedicine (CIBER-BBN), Madrid, Spain; Institut de Recerca Sant Joan de Déu, Barcelona, Spain*

Karolina Pietrowska • *Metabolomics Laboratory, Clinical Research Centre, Medical University of Bialystok, Bialystok, Poland*

Tania Portolés • *Enviromental and Public Health Analytical Chemistry, Research Institute for Pesticides and Water (IUPA), Universitat Jaume I, Castellón de la Plana, Spain*

Guillermo Quintás • *Health and Biomedicine, Leitat Technological Center, Terrassa, Spain; Analytical Unit, Health Research Institute La Fe, Valencia, Spain*
Rawi Ramautar • *Biomedical Microscale Analytics, Leiden Academic Centre for Drug Research, Leiden University, Leiden, The Netherlands*
Victoria Ramos-Garcia • *Neonatal Research Group, Health Research Institute La Fe, Valencia, Spain*
Rani Robeyns • *Toxicological Centre, University of Antwerp, Antwerp, Belgium*
Roberto Romero-González • *Department of Chemistry and Physics, Research Centre for Mediterranean Intensive Agrosystems and Agri-Food Biotechnology (CIAIMBITAL), University of Almeria, Agri-food International Campus of Excellence, CeiA3, Almeria, Spain*
Meher A. Saleem • *Bascom Palmer Eye Institute, University of Miami Miller School of Medicine, Miami, FL, USA; Miami Integrative Metabolomics Research Center, Miami, FL, USA; Georgetown University School of Medicine, Washington, DC, USA*
Daniel Sánchez-Rodas • *Associate Unit CSIC-University of Huelva "Atmospheric Pollution", Center for Research in Sustainable Chemistry – CIQSO, University of Huelva, Huelva, Spain; Department of Chemistry, Faculty of Experimental Sciences, University of Huelva, Huelva, Spain*
Juan V. Sancho • *Enviromental and Public Health Analytical Chemistry, Research Institute for Pesticides and Water (IUPA), Universitat Jaume I, Castellón de la Plana, Spain*
María Santos-Martín • *Agrifood Laboratory, Faculty of Experimental Sciences, University of Huelva, Huelva, Spain; International Campus of Excellence CeiA3, University of Huelva, Huelva, Spain*
Otto Savolainen • *Department of Biology and Biological Engineering, Chalmers Mass Spectrometry Infrastructure, Chalmers University of Technology, Gothenburg, Sweden*
Ana Sayago • *Agrifood Laboratory, Faculty of Experimental Sciences, University of Huelva, Huelva, Spain; International Campus of Excellence CeiA3, University of Huelva, Huelva, Spain*
Pilar Sepúlveda • *Regenerative Medicine and Heart Transplantation Unit, Health Research Institute La Fe, Valencia, Spain*
Pol Solà-Santos • *B2SLab, Departament d'Enginyeria de Sistemes, Automàtica i Informàtica Industrial, Universitat Politècnica de Catalunya, Barcelona, Spain; Networking Biomedical Research Centre in the Subject Area of Bioengineering, Biomaterials and Nanomedicine (CIBER-BBN), Madrid, Spain; Institut de Recerca Sant Joan de Déu, Barcelona, Spain*
Masahiro Sugimoto • *Institute of Medical Science, Tokyo Medical University, Tokyo, Japan; Institute for Advanced Biosciences, Keio University, Tsuruoka, Yamagata, Japan*
Isabel Ten-Doménech • *Neonatal Research Group, Health Research Institute La Fe, Valencia, Spain*
Atsumi Tomita • *Institute of Medical Science, Tokyo Medical University, Tokyo, Japan*
Alberto Valdés • *Laboratory of Foodomics, Institute of Food Science Research, CIAL, CSIC, Madrid, Spain*
Anddre Osmar Valdivia • *Department of Ophthalmology, Bascom Palmer Eye Institute, Miller School of Medicine at University of Miami, Miami, FL, USA; Miami Integrative Metabolomics Research Center, Miami, FL, USA*
Maria van de Lavoir • *Toxicological Centre, University of Antwerp, Antwerp, Belgium*

Tamara Vanhaecke • *Department of In Vitro Toxicology and Dermato-Cosmetology, Vrije Universiteit Brussel, Brussels, Belgium*

Marlien van Meyer • *Biomedical Microscale Analytics, Leiden Academic Centre for Drug Research, Leiden University, Leiden, The Netherlands*

Alexander L. N. van Nuijs • *Toxicological Centre, University of Antwerp, Antwerp, Belgium*

Gabrielle Yamar • *Bascom Palmer Eye Institute, University of Miami Miller School of Medicine, Miami, FL, USA; Miami Integrative Metabolomics Research Center, Miami, FL, USA*

Charles Yaros • *Bascom Palmer Eye Institute, University of Miami Miller School of Medicine, Miami, FL, USA; Miami Integrative Metabolomics Research Center, Miami, FL, USA*

Chapter 1

Comprehensive and High-Throughput Method for Quantitative Fingerprinting of Phenolic Compounds

Raúl González-Domínguez, Ana Sayago, María Santos-Martín, and Ángeles Fernández-Recamales

Abstract

Accurate, robust, and wide-coverage analytical tools are needed in polyphenol research to deal with the high physicochemical complexity of the secondary plant metabolome. In this chapter, a novel method based on reversed-phase ultrahigh-performance liquid chromatography coupled with a diode array detector and mass spectrometry is presented, which enables high-throughput, comprehensive, and quantitative fingerprinting of a broad spectrum of phenolic compounds and related metabolites in different food products. The simplicity, low-cost, and excellent analytical performance of this method would facilitate its implementation in food science for quality control and authenticity purposes.

Key words Phenolic compounds, Polyphenols, Liquid chromatography, Spectroscopy, Metabolomics, Food analysis

1 Introduction

Phenolic compounds are secondary metabolites widely distributed in the plant kingdom, which play crucial roles in the sensory, nutritional, and nutraceutical properties of plant-origin foods. For instance, many of these phytochemicals are closely related to the color, flavor, bitterness, and astringency of plant foods and beverages [1]. Furthermore, the bioactive potential of phenolic compounds is nowadays beyond any doubt (e.g., antioxidant, anti-inflammatory, anti-hyperlipidemic, and prebiotic activities), and numerous studies have demonstrated that their consumption may prevent multiple chronic diseases, such as metabolic disorders (e.g., obesity, diabetes), cancer, and cardiovascular and neurodegenerative diseases [2]. In this vein, the development of novel analytical methods capable of characterizing the phenolic content of agrifood products is a very current topic of great interest for quality control and authentication purposes. However, it should be noted here that

Raúl González-Domínguez (ed.), *Mass Spectrometry for Metabolomics*,
Methods in Molecular Biology, vol. 2571, https://doi.org/10.1007/978-1-0716-2699-3_1,
© The Author(s), under exclusive license to Springer Science+Business Media, LLC, part of Springer Nature 2023

the secondary plant metabolome comprises thousands of metabolites with highly diverse physicochemical properties (e.g., molecular weight, polarity, volatility) and concentration ranges.

The chemical structure of phenolic compounds is based on one or more aromatic rings, to which at least one hydroxyl substituent is attached [2]. In particular, plant phenolics that are composed of more than one aromatic ring are generally referred to as polyphenols, but both terms are often used interchangeably in the literature. Although polyphenols can be categorized according to different criteria, the most common classification considers the structure of the aromatic backbone to differentiate between flavonoid and non-flavonoid compounds. All flavonoids share a phenylbenzopyran skeleton, but variations in the hydroxylation pattern and oxidation state of the central pyran ring can result in multiple flavonoid sub-classes, including flavan-3-ols, flavanones, isoflavones, flavones, flavonols, and anthocyanins. On the other hand, non-flavonoid phenolics comprise a miscellaneous group of compounds, such as phenolic acids and phenolic alcohols, characterized by a single benzene ring linked to a carboxyl or alcohol group, respectively; stilbenes, consisting of two benzene rings linked by a double bond; or lignans, which are derived from the combination of two phenylpropanoid units [2]. In turn, phenolic compounds can occur in foods as aglycones or as conjugated forms with sugars and/or organic acids. Taking all this into account, accurate, robust, and comprehensive analytical methods are needed in polyphenol research to manage this physicochemical complexity.

In recent years, our research group has made great efforts to develop powerful tools to investigate the influence of variety, geographical origin, and agronomic practices on the chemical composition of different agrifood products, with a special focus on the main crops that are grown in the province of Huelva (southwest Spain). In this respect, we have previously demonstrated the potential of a myriad of chromatographic and spectroscopic methods for the authentication and traceability of strawberry [3–6], olive oil [7–9], Iberian ham [10, 11], and wine-derived products [12, 13]. Herein, we describe a novel high-throughput and quantitative metabolomics method for wide-coverage phenolic fingerprinting of diverse food products based on reversed-phase ultrahigh-performance liquid chromatography (RP-UHPLC) coupled with a diode array detector (DAD) and mass spectrometry (MS). The use of spectroscopic detection, which is often cheaper, simpler, and more commonly available than MS instruments, facilitates the implementation of this method in any analytical laboratory, either in the food industry or the research field. To our knowledge, this is the first LC-DAD-based platform capable of accomplishing high-throughput and wide-coverage phenolic fingerprinting analysis [14], especially when compared with previously published methods that often need longer analysis times (usually

ranging from 30 to 120 min) for the determination of fewer polyphenol species (less than 20–30 metabolites).

2 Materials

2.1 Metabolite Solutions

1. Individual stock solutions: 10,000 mg L^{-1} for each target phenolic compound and internal standard (except for ellagic acid and anthocyanins, which are prepared at 5000 and 1000 mg L^{-1}, respectively), dissolved in the solvents that are specified in Table 1.
2. Multi-metabolite working solutions: 100 mg L^{-1} for each target phenolic compound in 50% water:acetonitrile (v:v). Prepare three different multi-metabolite working solutions (i.e., solutions A, B, and C) by dilution of the individual stock solutions, each one containing the metabolites that are specified in Table 1.
3. Internal standard (IS) solution: 200 mg L^{-1} 2,6-dimethoxybenzoic acid and bisphenol A in water.
4. Multi-metabolite calibration curves: 0.01, 0.05, 0.1, 0.5, 1, 5, 10, 25, 50, 75, and 100 mg L^{-1} for each target phenolic compound in water, containing 20 mg L^{-1} 2,6-dimethoxybenzoic acid and bisphenol A as the internal standards. Prepare three different calibration curves by serial dilution of the three multi-metabolite working solutions (*see* **Note 1**).

2.2 Sample Preparation and Analysis

1. Extraction solvent for strawberry samples: 1% formic acid in methanol (v:v).
2. Extraction solvent for olive oil samples: 80:20 methanol:water (v:v).
3. Mobile phase A for anthocyanin compounds: 5% formic acid in water (v:v).
4. Mobile phase B for anthocyanin compounds: 5% formic acid in acetonitrile (v:v).
5. Mobile phase A for non-anthocyanin compounds: 0.1% formic acid in water (v:v).
6. Mobile phase B for non-anthocyanin compounds: acetonitrile.

2.3 Instrumentation

1. Agilent 1260 ultrahigh-performance liquid chromatography system equipped with a binary pump, autosampler, and diode array detector (Agilent Technologies, Santa Clara, CA).
2. Agilent 6410 triple quadrupole mass spectrometer equipped with an electrospray ionization source (ESI) (Agilent Technologies, Santa Clara, CA).

Table 1
Details about standard solution preparation (solvents for individual stock solutions, composition of the multi-metabolite mixtures) and analytical parameters (retention time, maximum absorbance, m/z) for the target phenolic compounds that are measured using the RP-UHPLC-DAD-MS platform

	Solution preparation		Analytical parameters		
Metabolite	**Solvent**	**Mix**	**RT (min)**	**λ (nm)**	**m/z (Da)**
Phenolic acids—hydroxybenzoic acids					
Benzoic acid	MeOH	A	9.83	280	121 $[M-H]^-$
4-Hydroxybenzoic acid	MeOH	A	3.67	260	137 $[M-H]^-$
3,4-Dihydroxybenzoic acid	MeOH	A	2.10	260	153 $[M-H]^-$
Vanillic acid	MeOH	A	7.01	260	167 $[M-H]^-$
Gallic acid	MeOH	A	1.18	280	169 $[M-H]^-$
Methylgallate	MeOH	B	5.09	280	183 $[M-H]^-$
Ethylgallate	MeOH	B	10.67	280	197 $[M-H]^-$
4-*O*-Methylgallic acid	MeOH	A	16.03	260	183 $[M-H]^-$
Syringic acid	MeOH	A	10.14	280	197 $[M-H]^-$
Phenolic acids—hydroxyphenylacetic acids					
4-Hydroxyphenylacetic acid	MeOH	A	4.82	280	151 $[M-H]^-$
3,4-Dihydroxyphenylacetic acid	MeOH	A	2.59	280	167 $[M-H]^-$
Phenolic acids—hydroxycinnamic acids					
trans-Cinnamic acid	MeOH	A	18.54	280	147 $[M-H]^-$
o-Coumaric acid	MeOH	A	14.40	280	163 $[M-H]^-$
m-Coumaric acid	MeOH	A	12.76	280	163 $[M-H]^-$
p-Coumaric acid	MeOH	A	11.05	320	163 $[M-H]^-$
Caffeic acid	MeOH	A	7.90	320	179 $[M-H]^-$
Ferulic acid	MeOH	A	13.98	320	193 $[M-H]^-$
3-Caffeoylquinic acid	MeOH	A	10.63	320	353 $[M-H]^-$
Sinapic acid	MeOH	A	15.73	320	223 $[M-H]^-$
Phenolic acids—hydroxyphenylpropionic acids					
3-(4-Hydroxyphenyl)propionic acid	MeOH	A	8.72	280	165 $[M-H]^-$
Hydroxybenzaldehydes					
4-Hydroxybenzaldehyde	MeOH	A	5.06	280	121 $[M-H]^-$
3,4-Dihydroxybenzaldehyde	MeOH	A	3.15	280	137 $[M-H]^-$
Vanillin	MeOH	A	8.93	280	151 $[M-H]^-$

(continued)

Table 1
(continued)

Metabolite	Solution preparation		Analytical parameters		
	Solvent	Mix	RT (min)	λ (nm)	m/z (Da)
Syringaldehyde	MeOH	A	11.99	280	181 $[M-H]^-$
Phenolic alcohols					
4-Methylcatechol	MeOH	B	4.89	280	123 $[M-H]^-$
4-Ethylphenol	MeOH	A	17.54	280	121 $[M-H]^-$
4-Vinylphenol	MeOH	A	14.83	260	119 $[M-H]^-$
Eugenol	MeOH	A	22.89	280	163 $[M-H]^-$
Methoxyeugenol	MeOH	A	23.83	280	193 $[M-H]^-$
Furan derivatives					
Furfural	MeOH	A	1.79	280	97 $[M+H]^+$
5-(Hydroxymethyl)furfural	MeOH	A	1.51	280	127 $[M+H]^+$
2,5-dimethyl-4-hydroxy-furanone	MeOH	A	2.33	280	129 $[M+H]^+$
2,5-dimethyl-4-methoxy-furanone	MeOH	A	6.68	280	143 $[M+H]^+$
Phenylethanoids					
Tyrosol	MeOH	A	3.87	280	137 $[M-H]^-$
Hydroxytyrosol	MeOH	A	2.14	280	153 $[M-H]^-$
Oleuropein	MeOH	A	21.46	280	539 $[M-H]^-$
Tannins					
Ellagic acid	1 M NaOH	A	18.90	260	301 $[M-H]^-$
Stilbenes					
trans-Resveratrol	MeOH: DMSO 75:25 (v:v)	B	20.23	320	227 $[M-H]^-$
trans-Resveratrol 3-*O*-glucoside	MeOH: DMSO 75:25 (v:v)	B	17.28	320	389 $[M-H]^-$
Flavonoids—flavan-3-ols					
Catechin	MeOH	A	8.39	280	289 $[M-H]^-$
Epicatechin	MeOH	A	12.44	280	289 $[M-H]^-$
Epigallocatechin gallate	MeOH	A	13.70	280	457 $[M-H]^-$
Epicatechin gallate	MeOH	A	17.35	280	441 $[M-H]^-$

(continued)

Table 1
(continued)

Metabolite	Solution preparation		Analytical parameters		
	Solvent	Mix	RT (min)	λ (nm)	m/z (Da)
Flavonoids—flavanones					
Naringenin	MeOH: DMSO 75:25 (v:v)	B	24.89	280	271 $[M-H]^-$
Naringenin 7-*O*-neohesperidoside	MeOH: DMSO 75:25 (v:v)	B	19.95	280	579 $[M-H]^-$
Hesperetin	MeOH: DMSO 75:25 (v:v)	B	27.21	280	301 $[M-H]^-$
Hesperetin 7-*O*-rutinoside	MeOH: DMSO 75:25 (v:v)	B	20.45	280	609 $[M-H]^-$
Flavonoids—flavonols					
Quercetin	MeOH: DMSO 75:25 (v:v)	B	23.98	360	301 $[M-H]^-$
Quercetin 3-*O*-galactoside	MeOH: DMSO 75:25 (v:v)	B	19.10	360	463 $[M-H]^-$
Quercetin 3-*O*-glucoside	MeOH: DMSO 75:25 (v:v)	B	19.33	360	463 $[M-H]^-$
Quercetin 3-*O*-rutinoside	MeOH: DMSO 75:25 (v:v)	B	19.35	360	609 $[M-H]^-$
Quercetin 3-*O*-rhamnoside	MeOH: DMSO 75:25 (v:v)	B	20.61	360	447 $[M-H]^-$
Kaempferol	MeOH: DMSO 75:25 (v:v)	B	28.33	360	285 $[M-H]^-$
Kaempferol 3-glucoside	MeOH: DMSO 75:25 (v:v)	B	20.55	360	447 $[M-H]^-$
Isorhamnetin	MeOH: DMSO 75:25 (v:v)	B	29.89	360	315 $[M-H]^-$

(continued)

Table 1
(continued)

Metabolite	Solution preparation		Analytical parameters		
	Solvent	Mix	RT (min)	λ (nm)	m/z (Da)
Isorhamnetin 3-*O*-glucoside	MeOH: DMSO 75:25 (v:v)	B	21.13	360	477 $[M-H]^-$
Morin	MeOH: DMSO 75:25 (v:v)	B	21.49	360	301 $[M-H]^-$
Flavonoids—flavones					
Luteolin	MeOH: DMSO 75:25 (v:v)	B	25.14	360	285 $[M-H]^-$
Apigenin	MeOH: DMSO 75:25 (v:v)	B	28.64	360	269 $[M-H]^-$
Flavonoids—anthocyanins					
Cyanidin	MeOH	C	9.64	520	287 $[M+H]^+$
Cyanidin 3-*O*-glucoside	MeOH	C	7.46	520	449 $[M+H]^+$
Malvidin	MeOH	C	12.52	520	331 $[M+H]^+$
Malvidin 3-*O*-glucoside	MeOH	C	9.81	520	493 $[M+H]^+$
Pelargonidin	MeOH	C	10.83	520	271 $[M+H]^+$
Pelargonidin 3-*O*-glucoside	MeOH	C	8.18	520	433 $[M+H]^+$
Peonidin	MeOH	C	11.85	520	301 $[M+H]^+$
Peonidin 3-*O*-glucoside	MeOH	C	9.08	520	463 $[M+H]^+$
Delphinidin 3-*O*-glucoside	MeOH	C	6.68	520	465 $[M+H]^+$
Petunidin 3-*O*-glucoside	MeOH	C	8.39	520	479 $[M+H]^+$
Internal standards					
2,6-Dimethoxybenzoic acid	MeOH		9.44[a]/5.10[b]	280	181 $[M-H]^-$
Bisphenol A	MeOH		30.74[a]/13.72[b]	280	227 $[M-H]^-$

[a]Retention times for the internal standards when using the RP-UHPLC-DAD method optimized for non-anthocyanin compounds
[b]Retention times for the internal standards when using the RP-UHPLC-DAD method optimized for anthocyanin compounds

3. Kinetex EVO C18 column (100 mm × 2.1 mm, 2.6 μm) equipped with a SecurityGuard™ ULTRA Cartridge UHPLC C18 (Phenomenex, Torrance, CA).

3 Methods

3.1 Sample Preparation

3.1.1 Extraction of Strawberry Samples

1. Homogenize the strawberries using a kitchen mixer.
2. Weight 0.2 g of the strawberry homogenate in an Eppendorf tube.
3. Add 1 mL of the extraction solvent.
4. Vigorously vortex the mixture for 1 min.
5. Sonicate for 15 min using an ultrasonic bath.
6. Centrifuge the sample at 10,000 × *g* for 10 min.
7. Filter the supernatant using 0.22-μm PTFE filters.
8. Transfer 180 μL of the supernatant to an LC injection vial and add 20 μL of the internal standard solution (*see* **Note 2**).
9. Keep the sample extracts at −20 °C until analysis.

3.1.2 Extraction of Olive Oil Samples

1. Weigh 0.5 g of the olive oil sample in an Eppendorf tube.
2. Add 1 mL of the extraction solvent.
3. Vigorously vortex the mixture for 1 min.
4. Centrifuge the sample at 10,000 × *g* for 10 min.
5. Transfer the supernatant to a new tube and add 0.5 mL of hexane.
6. Vortex for 1 min.
7. Centrifuge at 10,000 × *g* for 10 min.
8. Remove the hexane layer and repeat **steps 5–7** for a second clean-up.
9. Filter the extract using 0.22-μm PTFE filters.
10. Transfer 180 μL of the extract to an LC injection vial and add 20 μL of the internal standard solution (*see* **Note 2**).
11. Keep the sample extracts at −20 °C until analysis.

3.1.3 Extraction of Red Wine Samples

1. Filter the red wine sample using 0.22-μm PTFE filters.
2. Transfer 180 μL of the filtrate to an LC injection vial and add 20 μL of the internal standard solution.
3. Keep the sample extracts at −20 °C until analysis.

3.2 Analysis of Phenolic Compounds by RP-UHPLC-DAD-MS

1. Set the injection volume at 5 μL.
2. Maintain the column compartment at 40 °C.
3. Deliver mobile phases at 0.5 mL min^{-1}, using the following gradient programs.
   (a) For anthocyanin compounds: 0–10 min, 0–15% B; 10–14 min, 15–100% B; 14–18 min, 100% B.

(b) For non-anthocyanin compounds: 0–3 min, 0% B; 3–16 min, 0–12% B; 16–16.5 min, 12–16% B; 16.5–21 min, 16% B, 21–25 min, 16–20% B; 25–30 min, 20% B; 30–31 min, 20–100% B; 31–34 min, 100% B.

4. Equilibrate the LC system between injections by flowing 100% A for 5 min (*see* **Note 3**).
5. Monitor five different wavelengths using the diode array detector (DAD), as follows (*see* Table 1 for further details):
   (a) 260 nm for ellagic acid, 4-hydroxybenzoic acid, 3,4-dihydroxybenzoic acid, vanillic acid, 4-*O*-methylgallic acid, and 4-vinylphenol.
   (b) 280 nm for the rest of phenolic acids (except some hydroxycinnamic acids), phenolic alcohols, benzaldehydes, furan derivatives, phenylethanoids, flavan-3-ols, flavanones, and internal standards.
   (c) 320 nm for hydroxycinnamic acids (except trans-cinnamic acid, o-coumaric acid, and m-coumaric acid) and stilbenes.
   (d) 360 nm for flavonols and flavones.
   (e) 520 nm for anthocyanins.
6. Acquire MS spectra under positive and negative ionization modes to monitor the target phenolic compounds at the m/z values that are listed in Table 1 (*see* **Note 4**), using the following MS conditions:
   (a) Capillary voltage: 4000 V.
   (b) Gas flow rate: 10 L min^{-1}.
   (c) Gas temperature: 300 °C.
   (d) Nebulizer pressure: 35 psi.

3.3 Data Processing

1. Integrate the peak area of target phenolic compounds and internal standards at each concentration point of the calibration curves (Fig. 1).
2. Compute the area ratio for each target phenolic compound (i.e., ratio between the area of the analyte and the area of the internal standard) and obtain the linear regression equation by plotting the area ratio against the concentration.
3. Integrate the peak areas and compute area ratios in the samples (*see* **Note 5**).
4. Interpolate area ratios into the calibration curves to calculate the corresponding concentrations.
5. Apply dilution factors to calculate the concentrations of the phenolic compounds in the original sample.

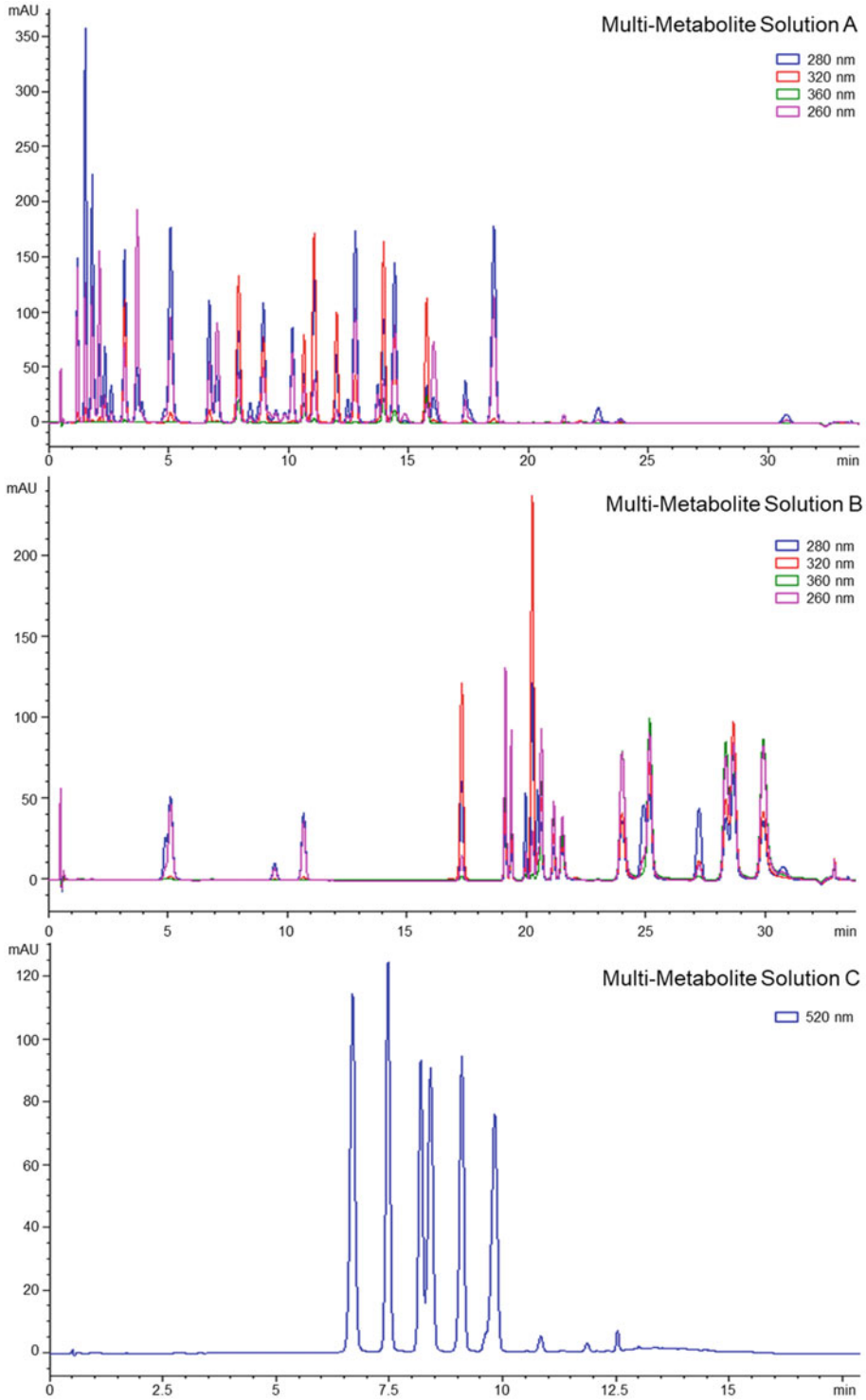

Fig. 1 Representative chromatograms for the three multi-metabolite standard solutions (mixtures A and B acquired with the RP-UHPLC method developed for non-anthocyanin compounds, mixture C acquired with the RP-UHPLC method developed for anthocyanin compounds)

4 Notes

1. Store all the metabolite solutions (i.e., stock solutions, muti-metabolite working solutions and calibration curves, internal standard solution) at −20 °C or −80 °C. Prepare low-volume aliquots to avoid thaw-freeze cycles, which may cause the degradation of phenolic compounds.
2. To increase method sensitivity, sample extracts can be pre-concentrated before analysis by taking them to dryness and further reconstitution in the volume desired. For this purpose, a nitrogen stream or a vacuum concentrator can be used, controlling the temperature during the entire process to avoid the degradation of thermally labile phenolic compounds. However, it should be noted this pre-concentration step may provoke the loss of volatile metabolites.
3. This equilibration step is crucial to minimize inter-sample variability in terms of retention times and peak areas.
4. To detect phenolic compounds, the MS instrument can be operated in different acquisition modalities, including full scan, single ion monitoring (SIM), or multiple reaction monitoring (MRM).
5. To facilitate the identification of the target compounds in the samples, a spectral library containing retention times and UV/Vis spectra can be used.

References

1. Delfanian M, Sahari MA (2020) Improving functionality, bioavailability, nutraceutical and sensory attributes of fortified foods using phenolics-loaded nanocarriers as natural ingredients. Food Res Int 137:109555
2. Durazzo A, Lucarini M, Souto EB, Cicala C, Caiazzo E, Izzo AA, Novellino E, Santini A (2019) Polyphenols: a concise overview on the chemistry, occurrence, and human health. Phytother Res 33:2221–2243
3. González-Domínguez R, Sayago A, Akhatou I, Fernández-Recamales Á (2020) Volatile profiling of strawberry fruits cultivated in a soilless system to investigate cultivar-dependent chemical descriptors. Foods 9:768
4. González-Domínguez R, Sayago A, Akhatou I, Fernández-Recamales Á (2020) Multi-chemical profiling of strawberry as a traceability tool to investigate the effect of cultivar and cultivation conditions. Foods 9:96
5. Akhatou I, Sayago A, González-Domínguez R, Fernández-Recamales Á (2017) Application of targeted metabolomics to investigate optimum growing conditions to enhance bioactive content of strawberry. J Agric Food Chem 65: 9559–9567
6. Akhatou I, González-Domínguez R, Fernández-Recamales Á (2016) Investigation of the effect of genotype and agronomic conditions on metabolomic profiles of selected strawberry cultivars with different sensitivity to environmental stress. Plant Physiol Biochem 101: 14–22
7. Sayago A, González-Domínguez R, Urbano J, Fernández-Recamales Á (2019) Combination of vintage and new-fashioned analytical approaches for varietal and geographical traceability of olive oils. LWT 111:99–104
8. González-Domínguez R, Sayago A, Morales MT, Fernández-Recamales Á (2019) Assessment of virgin olive oil adulteration by a rapid luminescent method. Foods 8:287
9. Sayago A, González-Domínguez R, Beltrán R, Fernández-Recamales Á (2018) Combination of complementary data mining methods for geographical characterization of extra virgin

olive oils based on mineral composition. Food Chem 261:42–50

10. González-Domínguez R, Sayago A, Fernández-Recamales Á (2020) Fatty acid profiling for the authentication of iberian hams according to the feeding regime. Foods 9:149
11. González-Dominguez R, García-Barrera T, Gómez-Ariza JL (2012) Iberian ham typification by direct infusion electrospray and photospray ionization mass spectrometry fingerprinting. Rapid Commun Mass Spectrom 26:835–844
12. Recamales AF, Gallo V, Hernanz D, González-Miret ML, Heredia FJ (2011) Effect of time and storage conditions on major volatile compounds of zalema white wine. J Food Qual 34:100–110
13. González-Domínguez R, Sayago A, Fernández-Recamales Á (2021) Potential of ultraviolet-visible spectroscopy for the differentiation of spanish vinegars according to the geographical origin and the prediction of their functional properties. Foods 10:1830
14. González-Domínguez R, Sayago A, Santos-Martín M, Fernández-Recamales Á (2022) High-throughput method for widecoverage and quantitative phenolic fingerprinting in plant-origin foods and urine samples. J Agric Food Chem 70:7796–7804

Chapter 2

Determination of Volatile Metabolites in Vinegar by Solid-Phase Microextraction–Gas Chromatography–Mass Spectrometry (SPME–GC–MS)

Remedios Castro and Enrique Durán-Guerrero

Abstract

Solid-phase microextraction (SPME) is an easy, sensitive, and environmentally friendly technique that has been employed, coupled to gas chromatography or liquid chromatography, to determine a huge amount of analytes with different volatilities. The present work describes the procedure to follow in order to determine volatile compounds in vinegar by SPME–GC–MS.

Key words SPME, Vinegar, Volatile compounds, Gas chromatography, Mass spectrometry, Extraction

1 Introduction

Vinegar is a solution of acetic acid used not only for flavoring but also for preserving a wide range of foods. It is produced by a double-fermentation process (alcoholic and acetic) from different raw materials such as white and red wine, cider, malted barley, honey, and pure alcohol, among others.

Its composition and organoleptic characteristics are conditioned by its production process and raw materials. From this perspective, the presence of certain volatile low-molecular metabolites in vinegar depends on the specific production conditions under which it has been elaborated [1]. Nowadays, its increasing demand has created the need to develop methodologies in order to characterize it, all of it to avoid frauds in its commercialization [2].

In order to determine volatile compounds in foods, some previous extraction techniques should be employed because, in the majority of the cases, they are present together with other compounds in very low amounts and in a very complex matrix

Raúl González-Domínguez (ed.), *Mass Spectrometry for Metabolomics*,
Methods in Molecular Biology, vol. 2571, https://doi.org/10.1007/978-1-0716-2699-3_2,
© The Author(s), under exclusive license to Springer Science+Business Media, LLC, part of Springer Nature 2023

[3–5]. Most of these extraction techniques have some limitations such as high costs and time, not being environmentally friendly, etc.

SPME appeared as a fast and inexpensive sample preparation technique in the 1990s [6]. It has been used coupled to GC-MS for the analysis, totally automatic or manual, of a wide variety of analytes and matrices [7–10]. In it, a polymer-coated silica fiber is employed to extract the analytes and then they are transferred into the injector of a GC system. Two modalities can be found: the direct immersion of the fiber into the sample (DI-SPME) or the exposure of the fiber to the headspace above a liquid or solid sample (HS-SPME).

Although it is a well-established technique, in these last years it is being improved toward the development of new applications and fibers [11–13].

2 Materials

All solutions should be prepared using ultrapure water and analytical grade reagents and be stored at 4 °C. Waste disposal regulations should be followed if waste materials are produced.

2.1 Solid-Phase Extraction

1. Carboxen–polydimethylsiloxane (CAR–PDMS, 75 μm) fiber (*see* **Note 1**).
2. SPME fiber kit (*see* **Note 1**).
3. 25-mL pipette.
4. 50-mL glass vial.
5. Sodium chloride.
6. Internal standard solution: 2.27 g/L of 4-methyl-2-pentanol (4M2P) in a solution containing 80 g/L of HPLC-grade acetic acid. Weigh 0.227 g of 4M2P with precision (*see* **Note 2**). Dissolve it with 10 mL of ultrapure water and immediately pass it to a volumetric flask. Add 7.6 mL of acetic acid (*see* **Note 3**). Make up to 100 mL with ultrapure water (*see* **Note 4**). Agitate the solution. Put it in a glass container and store at 4 °C.
7. Micropipette 10–100 μL.
8. Aluminum capsule.
9. PTFE-faced silicone septum (*see* **Note 5**).
10. Crimper (*see* **Note 6**).
11. Magnetic stirrer and stirring bar.
12. Thermostatic bath.
13. Thermostatted glass block.
14. Plastic pipes (*see* **Note 7**).

2.2 GC-MS Analysis

1. Gas chromatograph coupled to mass spectrometry detection (*see* **Note 8**).
2. Wax-type capillary column with 60 m length x 0.25 mm ID and with a 0.25 μm coating.
3. Helium.
4. Calibration curves: Prepare individual stock standard solutions of each aroma compound by weight in ethanol. Prepare a global stock standard solution containing all the analytes in a synthetic vinegar solution (2 g/L of tartaric acid, 80 g/L of acetic acid, 1 g/L of ethyl acetate, and 10 mL/L of ethanol, in ultrapure water). Prepare calibration solutions by diluting different amounts of the global standard solution in a synthetic vinegar solution. Five levels of concentration were tested in triplicate (*see* **Notes 9** and **10**).

3 Methods

3.1 Manual Solid-Phase Microextraction

1. Weigh 6.14 g of NaCl in the 50-mL glass vial.
2. Pipette 15 mL of vinegar into the glass vial.
3. Add 50 μL of the internal standard solution.
4. Add the stirring bar to the glass vial.
5. Adjust the aluminum capsule with the PTFE-faced silicone septum using the crimper.
6. Switch on the thermostatic bath and adjust the temperature at 70 °C (*see* **Note 11**).
7. Install the glass thermostatted block onto the stirrer.
8. Install the fiber into the manual syringe.
9. Slide the locking screws onto the syringe from the plunger side: the large locking screw onto the silver body and the small locking screw onto the wider part of the black plunger (*see* **Note 12**).
10. Raise the syringe plunger totally and insert the syringe and the large diameter locking screw into the upper position of the extraction guide (*see* **Note 13**).
11. Lock the syringe by rotating it until the locking screw is in the notch.
12. Adjust the syringe. The SPME fiber must be protruded about 1 cm beyond the inner base of the extraction guide. Now the lower locking screw must be tightening.
13. Place the extraction guide with the syringe on the glass vial and loosen the upper locking screw.
14. Adjust the SPME fiber by moving the black plunger.

15. Place the sample vial with the syringe body into the thermostatted glass block.
16. Slide the upper locking screw until it is flush against the top of the syringe body.
17. Stir the sample for 60 min at 70 °C (*see* **Note 14**).
18. Slide the upper locking screw up retracting the fiber into the protective needle.
19. Remove the syringe from the sample (*see* **Note 15**).

3.2 Desorption Procedure

1. Insert the syringe into the injection guide and lock into place by rotating (*see* **Notes 16** and **17**).
2. Place the adapter cup onto the GC injection port.
3. Put the syringe with the injection guide on the adapter cup and push the plunger down.
4. After 2 min, retract the fiber into the needle, and remove it from the GC port. Now it is prepared for the next extraction (*see* **Note 18**).

3.3 GC-MS Analysis

The GC-MS analysis should employ the following parameters:

1. The carrier gas is helium with a flow of 1 mL/min.
2. The transfer line temperature is 300 °C.
3. The initial temperature of the oven is 35 °C.
4. Hold this temperature for 10 min.
5. Increase the temperature at 5 °C/min up to 100 °C.
6. Increase the temperature at 3 °C/min up to 210 °C.
7. Hold this temperature for 40 min.
8. The mass detector is operated in EI+ mode at 70 eV in a range of 30–400 amu.
9. Full scan mode can be employed (*see* **Note 19**).
10. Injection mode: splitless mode for 2 min.

3.4 Obtention of Quantitative Information

1. The integration is carried out using the base peak of each volatile compound to be quantified.
2. Relative areas (area of each volatile compound related to the area of the internal standard) should be used. In this way, chromatograms similar to Fig. 1 should be obtained.
3. The relative area of each volatile compound is transformed into concentration values by means of a calibration curve previously constructed for each compound. Some calibration curves obtained in a previous study [14] are shown in Table 1.

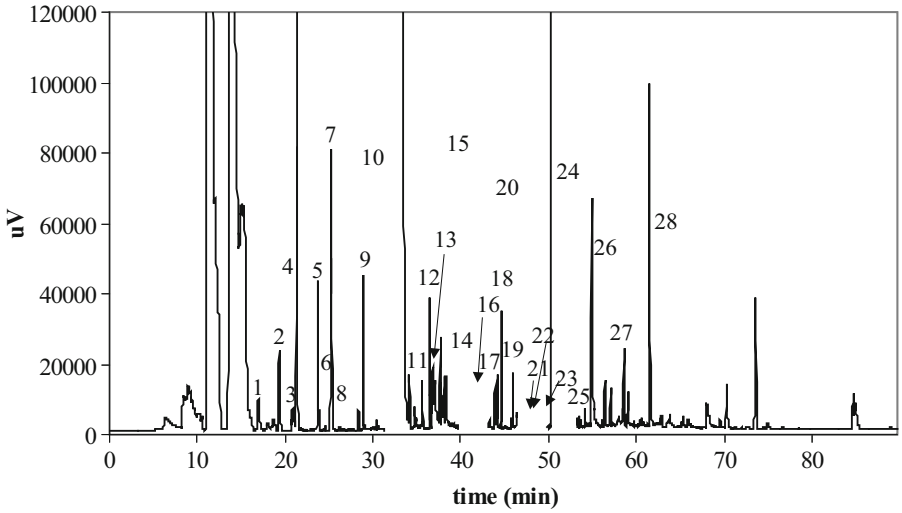

Fig. 1 Chromatogram of a Sherry wine vinegar aged in wood. 1, n-butyl acetate; 2, ethyl pentanoate; 3, 2-methyl-1-propanol; 4, isoamyl acetate; 5, 4-methyl-2-pentanol (I.S.); 6, ethyl hexanoate; 7, 2-methyl-1-butanol; 8, isoamyl alcohol; 9, 3-hydroxy-2-butanone; 10, acetic acid; 11, 2-furancarboxaldehyde; 12, benzaldehyde; 13, 2,3-butanediol; 14, ethyl decanoate; 15, isopentanoic acid; 16, diethyl succinate; 17, benzyl acetate; 18, 1,2-dihydro-1,1,6-trimethyl naphthalene; 19, ethyl-2-phenyl acetate; 20, phenylethyl acetate; 21, hexanoic acid; 22, α-ionone; 23, benzyl alcohol; 24, 2-phenylethanol; 25, 4-ethylguaiacol; 26, octanoic acid; 27, 4-ethylphenol; 28, decanoic acid

4 Notes

1. From Supelco (Bellefonte, PA, USA).
2. The standard is liquid at room temperature. It should be weighed using a Pasteur pipette. The mentioned concentration is indicative. This concentration should be recalculated in case the weighed amount was different from the mentioned one.
3. Taking into account that acetic acid is liquid at room temperature, it is simpler to measure a volume of acetic acid than to weigh it. Its density value is 1.05 g/mL (25 °C).
4. In order not to suffer losses by evaporation, this step should be done as fast as possible.
5. It should be used with its PTFE face toward the sample.
6. The crimper´s diameter should be the same that the aluminum capsule's one.
7. The thermostatic bath together with the double-sided glass thermostatted block and the plastic pipes allows controlling the temperature by means of a water close circuit.
8. The appropriate GC-specific adaptor cup on the end of the injection guide should be employed.

Table 1
Characteristics of some calibration curves obtained by SPME–GC–MS

Compound	Linear range (mg/L)	Regression coefficient	Linearity (LOL, %)	Slope ± S.D.	Intercept ± S.D.
n-Butyl acetate	0.00924–1.85	0.9973	98.80	0.2403 ± 0.0029	0.0236 ± 0.0023
Ethyl pentanoate	0.00564–1.00	0.9962	98.59	0.8881 ± 0.0125	0.0241 ± 0.0053
Isoamyl acetate	0.0417–11.0	0.9989	99.05	0.4930 ± 0.0047	0.0133 ± 0.0221
Ethyl hexanoate	0.00196–0.0980	0.9919	97.49	1.54 ± 0.0388	0.0170 ± 0.0190
Isoamyl alcohol	0.0367–50.0	0.9993	99.14	0.0731 ± 0.0060	0.0670 ± 0.0022
3-Hydroxy-2-butanone	22.0–1000	0.9984	98.75	0.0016 ± 0.0000	-0.0217 ± 0.0059
2-Furancarboxaldehyde	0.0241–3.60	0.9991	99.23	0.1563 ± 0.0012	0.0126 ± 0.0020
Benzaldehyde	0.00726–0.605	0.9967	99.41	4.40 ± 0.0700	0.1232 ± 0.0208
Ethyl decanoate	0.00261–0.0653	0.9972	98.31	43.73 ± 0.7385	0.2470 ± 0.0253
Isopentanoic acid	1.0–95.0	0.9976	98.64	0.0370 ± 0.0005	-0.0128 ± 0.0159
Diethyl succinate	0.0447–3.00	0.9979	98.85	0.1916 ± 0.0022	0.0237 ± 0.0023
Benzyl acetate	0.0108–0.239	0.9958	97.30	1.32 ± 0.0389	0.0413 ± 0.0045
Phenylethyl acetate	0.0185–3.31	0.9968	98.23	3.72 ± 0.0557	0.3865 ± 0.0496
α-Ionone	0.00261–0.0652	0.9953	97.82	18.39 ± 0.0145	0.0957 ± 0.0145
Benzyl alcohol	0.0583–1.50	0.9976	98.44	0.1497 ± 0.0023	0.0214 ± 0.0011
2-Phenylethanol	0.853–67.5	0.9944	97.35	0.1051 ± 0.0020	0.2403 ± 0.0505
Octanoic acid	0.0268–1.49	0.9986	98.96	1.46 ± 0.0151	0.0790 ± 0.0113
4-Ethylphenol	0.00858–0.319	0.9931	97.36	2.57 ± 0.0679	0.0126 ± 0.0027

9. The concentrations cover the concentration ranges expected for the various volatile metabolites in vinegar.
10. All the solutions must be stored at 4 °C.
11. Make sure that the water recirculation system has no leaks.
12. Do not overtighten the screws; it will be done later.
13. The screw must not be tightened against the black plunger.
14. Make sure that the sample is being correctly agitated.
15. Be careful with the fiber.
16. The distance from the tip of the SPME fiber to the groove inside the adapter cup should be not more than 67 mm.
17. Twist the locking ring down until it locks on the body of the injection guide.
18. In the case of the extraction of very complex samples, the fiber should be conditioned according to the manufacturer's instruction between samples.
19. Full Scan mode is usually valid to determine volatile compounds up to ppb concentration levels. If only specific compounds are going to be measured, to increase the sensitivity, single ion monitoring (SIM) could be used.

References

1. Solieri L, Giudici P (2009) Vinegars of the world. In: Solieri L, Giudici P (eds) Vinegars of the world. Springer, Milan
2. Natera R, Castro R, García-Moreno MV, Hernández MJ, García-Barroso CG (2003) Chemometric studies of vinegars from different raw materials and processes of production. J Agric Food Chem 51:3345–3351
3. Šikuten I, Štambuk P, Kontić JK, Maletić E, Tomaz I, Preiner D (2021) Optimization of SPME-arrow-GC/MS method for determination of free and bound volatile organic compounds from grape skins. Molecules 26:7409
4. Castro R, Natera R, Durán E, García-Barroso C (2008) Application of solid phase extraction techniques to analyse volatile compounds in wines and other enological products. Eur Food Res Technol 228:1–18
5. Ferrer-Valverde MA, Sánchez-Palomo E, Osorio-Alises M, Chaya C, González-Viñas MA (2021) Volatile and sensory characterization of La Mancha Trujillo melons over three consecutive harvests. Foods 10:1683
6. Arthur CL, Pawliszyn J (1990) Solid phase microextraction with thermal desorption using fused silica optical fibers. Anal Chem 62:2145–2148
7. Louch D, Motlagh S, Pawliszyn J (1992) Dynamics of organic compound extraction from water using liquid-coated fused silica fibers. Anal Chem 64:1187–1199
8. De la Calle D, Reichenbacher M, Danzer KJ, Hurlbeck C, Bartzsch C, Feller K (1998) Analysis of wine bouquet components using headspace solid-phase microextraction-capillary gas chromatography. J High Resolut Chromatogr 21:373–377
9. Yang X, Peppard T (1994) Solid-phase microextraction for flavor analysis. J Agric Food Chem 42:1925–1930
10. Jia M, Zhang QH, Min DB (1998) Optimization of solid-phase microextraction analysis for headspace flavor compounds of orange juice. J Agric Food Chem 46:2744–2747
11. Resende dos Santos R, Orlando RM, De Lourdes Z, Menezes HC (2021) Assessment of polycyclic aromatic hydrocarbons and derivatives in beer using a new cold fiber-solid phase microextraction system. Food Control 126:108104
12. Ji X (2021) Comparative investigation of volatile components and bioactive compounds in beers by multivariate analysis. Flavour Fragr J 36:374–383

13. Zhou C, Zhou Y, Hu Y, Li B, Zhang R, Zheng K, Liu J, Wang J, Zuo M, Liu S (2021) Integrated analysis of metabolome and volatile profiles of germinated brown rice from the Japonica and Indica Subspecies. Foods 10: 2448

14. Natera R, Castro R, García-Moreno MV, García-Rowe F, García-Barroso C (2002) Headspace solid-phase microextraction analysis of aroma compounds in vinegar: validation study. J Chromatogr A 967:261–267

Chapter 3

Determination of Volatile Metabolites in Vinegar by Stir Bar Sorptive Extraction–Gas Chromatography–Mass Spectrometry (SBSE–GC–MS)

Enrique Durán-Guerrero and Remedios Castro

Abstract

Stir bar sorptive extraction (SBSE) is a rapid, sensitive, precise, and environmentally friendly extraction technique that, coupled to gas chromatography–mass spectrometry detection (GC–MS), enables the simple determination of volatile organic compounds in liquid samples. The present protocol describes the procedure for the determination of volatile compounds in vinegar by means of SBSE–GC–MS.

Key words SBSE, Gas chromatography, Mass spectrometry, Vinegar, Volatile compounds, Extraction

1 Introduction

Vinegar is today a highly renowned product, very appreciated in gastronomy [1]. Due to the diversity of vinegar on the market and the increase in demand, it is necessary to investigate reliable analytical methods to establish criteria for determining quality and origin since objective authentication remains a pending issue [2, 3]. The aroma of vinegar depends on the raw materials, and the components formed during the acetic fermentation and during aging. Therefore, vinegars could be characterized and differentiated by quantitative and qualitative analysis of their volatile profile, which consists of hundreds of secondary low-molecular-weight metabolites, such as esters, alcohols, furanic compounds, and many others [4].

The extraction and concentration of flavor components, prior to GC analysis, is a problem that has not yet been satisfactorily resolved. The need for sample preparation is due to several factors, such as the low concentration of the analytes, their high variety, the complexity of the matrices, or the low stability of the compounds [5]. In recent years, there has been a growing interest in developing

Raúl González-Domínguez (ed.), *Mass Spectrometry for Metabolomics*,
Methods in Molecular Biology, vol. 2571, https://doi.org/10.1007/978-1-0716-2699-3_3,
© The Author(s), under exclusive license to Springer Science+Business Media, LLC, part of Springer Nature 2023

new analytical techniques for the monitoring of volatile compounds in many matrices. The trend is to develop accurate, sensitive, and easy automatic methods that reduce sample preparation. The preparation of samples for the GC analysis of volatile compounds has been carried out by liquid/liquid or solid-phase extraction, headspace, purge and trap methods, supercritical fluid extraction, or simultaneous distillation-solvent extraction, among others. All of these sample preparation methods have some limitations, such as high cost and long time, possible generation of artifacts, etc. [6].

Stir bar sorptive extraction (SBSE) is a relatively new developed technique proposed by Baltussen et al. [7] in which a stir bar coated with polydimethylsiloxane (PDMS) is used to extract analytes from liquid matrices. The extraction mechanism is similar to that of solid-phase microextraction (SPME) based on PDMS sorption [8]. A magnetic stir bar is added to the sample, where the transfer of analytes to the polymer coating is carried out (*see* Fig. 1). After a certain extraction time, the analytes are thermally desorbed in the GC injector. The advantage of SBSE is the much higher mass of PDMS available compared to SPME, resulting in high recoveries and higher sample capacity. Applications developed with SBSE to different kinds of food samples have shown a high potential of this technique [9–18].

Fig. 1 Stir bar sorptive extraction procedure. *PDMS* Polydimethylsiloxane

2 Materials

Prepare all solutions using ultrapure water and analytical grade reagents. Prepare all reagents at room temperature and store them at 4 °C. Follow all waste disposal regulations when disposing waste materials.

2.1 Stir Bar Sorptive Extraction

1. Stir bar 10 mm × 0.5 mm (*see* **Note 1**).
2. 25-mL pipette.
3. 100-mL Erlenmeyer flask.
4. Sodium chloride.
5. Internal standard solution: 2.2 g/L of 4-methyl-2-pentanol (4M2P) in 80 g/L of HPLC grade acetic acid. Weigh 0.22 g of 4M2P with precision in a beaker (*see* **Note 2**). Dissolve it with 10 mL of Milli-Q water and immediately cover the beaker with parafilm to avoid evaporation (*see* **Note 3**). Add 7.6 mL of acetic acid to the solution (*see* **Note 4**). Make up to 100 mL with Milli-Q water and store at 4 °C in a glass container.
6. Micropipette 10–100 μL.
7. Parafilm.
8. Magnetic stirrer.
9. Tweezers.
10. A lint-free tissue.

2.2 Desorption Procedure

1. The chromatographic instrument should be coupled to a thermal desorption unit (TDU) and to a cooled injection system (CIS) programmed-temperature vaporization (PTV) inlet supplied by Gerstel. It is also coupled to an MPS autosampler supplied by Gerstel (*see* **Note 5**).
2. Glass liner for injection.
3. Liquid nitrogen.
4. Helium.

2.3 GC–MS Analysis

1. Gas chromatograph coupled to mass spectrometry detection (*see* **Note 6**).
2. Wax-type capillary column with 60 m length × 0.25 mm ID and with a 0.25 μm coating.
3. Helium.
4. Calibration curves: Prepare individual stock standard solutions of each aroma compound by weight in ethanol. Prepare a global stock standard solution containing all the analytes in a synthetic vinegar solution (2 g/L of tartaric acid, 80 g/L of acetic acid, 1 g/L of ethyl acetate, and 10 mL/L of ethanol, in

ultrapure water). Prepare calibration solutions by diluting different amounts of the global standard solution in a synthetic vinegar solution. Five levels of concentration were tested in triplicate (*see* **Notes 7** and **8**).

3 Methods

Carry out all the procedures at room temperature.

3.1 Stir Bar Sorptive Extraction

1. Weigh 5.85 g of NaCl in the 100-mL Erlenmeyer flask.
2. Pipette 25 mL of vinegar into the Erlenmeyer flask.
3. Add 84 μL of the internal standard solution.
4. Cover the Erlenmeyer flask with parafilm.
5. Agitate manually the Erlenmeyer flask until the complete dissolution of the NaCl in the sample (*see* **Note 9**).
6. Insert the stir bar into the Erlenmeyer flask avoiding the contact with the hands and cover again with parafilm (*see* **Note 10**).
7. Place the Erlenmeyer flask onto the magnetic stirrer and stir at 1250 rpm for 120 min.
8. Remove the stir bar from the sample (*see* **Note 11**) and wash it with distilled water for a few seconds to remove salt and sample remains (*see* **Note 12**).
9. Dry gently with a lint-free tissue (*see* **Note 13**).
10. Transfer the stir bar to a glass liner for the desorption process and place it in the autosampler tray.

3.2 Desorption Procedure

The desorption procedure is divided into the TDU procedure followed by the CIS procedure.

TDU Parameters:

1. Initial temperature of 40 °C.
2. Increase the temperature at 60 °C/min up to 300 °C.
3. Hold this temperature for 10 min.
4. Helium flow employed is 75 mL/min.
5. 0.5 min of delay time to start the CIS procedure.
6. The injection mode is solvent vent (*see* **Note 14**).

CIS Parameters:

1. Criofocusing initial temperature of −140 °C employing liquid nitrogen (*see* **Note 15**).
2. Increase the temperature at 10 °C/min up to 300 °C.
3. Hold this temperature for 5 min.
4. Helium flow for injection is 75 mL/min.

3.3 GC–MS Analysis

The GC–MS analysis should employ the following parameters:

1. The carrier gas is helium with a flow of 1 mL/min.
2. The transfer line temperature is 300 °C.
3. The initial temperature of the oven is 35 °C.
4. Hold this temperature for 10 min.
5. Increase the temperature at 5 °C/min up to 100 °C.
6. Increase the temperature at 3 °C/min up to 210 °C.
7. Hold this temperature for 40 min.
8. The mass detector is operated in EI+ mode at 70 eV in a range of 30–400 amu.
9. Full scan mode can be employed (*see* **Note 16**).

3.4 Obtention of Quantitative Information

1. The chromatograms obtained should be integrated to obtain quantitative information. The integration is always carried out with the base peak of each volatile compound to be quantified (*see* **Note 17**).
2. Table 1 shows some values of m/z that should be employed for some typical compounds, together with their retention times.
3. The area of each volatile compound has to be related to the area of the internal standard to obtain relative areas.
4. The relative area of each volatile compound is transformed into concentration values by means of the calibration curves previously constructed for each compound. Some calibration curves can be found in previous research [6].

4 Notes

1. Supplied by Gerstel (Mülheim a/d Ruhr, Germany).
2. Due to the liquid state of the standard at room temperature, the weight of the compound should be done employing a Pasteur pipette. Is not necessary that the amount weighted is exactly the indicated, but in that case, the concentration should be recalculated.
3. To avoid possible losses of the volatile compound by evaporation, this step should be done as fast as possible.
4. It is simpler to measure a volume of acetic acid, instead of a weight. The density value at 25 °C (1.05 g/mL) should be employed.
5. It is not strictly necessary to use an automated system, but it facilitates the procedure. In addition, the instrument employed for the SBSE analysis usually includes the MPS autosampler in its configuration.

Table 1
Retention times and m/z employed for quantification purposes of some typical volatile compounds that can be found in vinegar samples by means of SBSE–GC–MS

Volatile compound	Retention time (min)	m/z
Isobutyl acetate	13.93	43
Ethyl butyrate	14.97	71
Camphene	16.17	93
Disulfide dimethyl	16.18	94
Ethyl isovaleriate	16.37	88
4-Methyl-2-pentyl acetate	17.99	43
Isoamyl acetate	18.73	43
Allyl sulfide	19.11	73
beta-Myrcene	19.65	93
alpha-Terpinene	20.18	121
Amyl acetate	20.28	43
2,6-Dimethyl-4-heptanone	20.38	85
Eucaliptol isomer 1	21.06	111
4-Methyl-2-pentanol (Internal Standard)	21.48	45
Limonene	21.56	68
Herboxide isomer 1	21.92	79
Eucalyptol isomer 2	22.49	43
2-Methyl-1-butanol	22.49	57
Hexanoic acid ethyl ester	22.49	88
3-Methyl-1-butanol	22.72	55
Herboxyde isomer 2	23.03	67
gamma-Terpinene	23.05	93
3-Octanone	23.42	99
m-Cymene	23.49	119
p-Cymene	23.68	119
Methyl allyl sulfide	23.80	120
Acetic acid hexyl ester	23.89	43
Octanal	24.50	45
2,4-Dithiapentane	24.61	61
Acetoin	25.31	43
3-Hexen-1-ol acetate	25.38	67

(continued)

Table 1
(continued)

Volatile compound	Retention time (min)	m/z
1-Hexanol	27.22	56
Dimethyl trisulfide	27.39	126
3-Hexen-1-ol (Z)	28.26	67
3-Octanol	28.71	59
D-Fenchone	29.15	81
Acetic acid	29.90	43
Dimethyl styrene	30.09	117
Filifolone	30.10	80
Octanoic acid ethyl ester	30.18	88
Furfural	30.45	95
1-Octen-3-ol	30.82	57
Neroloxide	31.39	68
Allyl disulfide	31.45	41
Diallyl disulfide	31.45	41
Linalool oxide	31.53	43
1,2-Propanediol diacetate	32.37	43
Benzaldehyde	32.89	106
Camphor	33.23	95
Vitispirane	33.73	192
Linalool	33.81	71
Isobutyric acid	34.20	47
Pinocamphone	34.36	83
1-Octanol	34.48	56
5-Methyl furfural	34.57	110
2-Hydroxypropyl acetate	34.57	43
Terpinene-1-ol	35.30	81
D-Fenchyl alcohol	35.31	81
Butyric acid	36.40	60
4-Terpineol	36.55	71
Myrtenal	37.09	79
Decanoic acid ethyl ester	37.49	88

(continued)

Table 1
(continued)

Volatile compound	Retention time (min)	m/z
Safranal	37.66	107
beta-Terpineol	37.67	93
Disulfide, methyl(methylthio)methyl	37.82	61
Isovaleric acid	37.82	60
Allyl anisole	38.20	148
Butanedioic acid diethyl ester isomer 1	38.26	101
Isoborneol	38.34	95
Carvotanacetone	39.14	82
alpha-Terpineol	39.43	121
Borneol	39.57	95
Benzyl acetate	40.03	108
Umbellulone	40.03	108
Verbenone	40.04	107
3-Vinyl-1,2-dithiocyclohexen-4-ene	40.41	111
Carvenone	40.42	110
Piperitenone isomer 1	40.83	82
butanedioic acid diethyl ester isomer 2	41.05	99
Naphtalene,1,2-dihydro-1,1,6-trimethyl	41.06	157
trans-Anethole	41.13	148
Nerol acetate	41.28	69
Geraniol acetate	41.28	69
beta-Citronellol	41.64	69
Methyl salicylate	41.79	120
4-Methylacetophenone	41.97	91
Ethyl phenyl acetate	42.02	119
Nerol	42.79	69
Phenethyl acetate	43.21	104
beta -Damascenone	43.70	121
Hexanoic acid	44.00	60
p-Cymen-8-ol	44.28	135
Geraniol	44.34	69

(continued)

Table 1
(continued)

Volatile compound	Retention time (min)	m/z
alpha-Ionone	44.86	121
Benzyl alcohol	45.03	79
N-(3-methylbuthyl)acetamide	45.34	73
alpha-Ionol	46.23	95
Phenethyl alcohol	46.34	91
2-Phenyl-2-butenal	46.62	117
4-Isopropyl acetophenone	47.02	147
Piperitenone isomer 2	47.41	150
Hexanoic acid, 2-ethyl	47.56	88
Heptanoic acid	47.61	60
beta-Ionone	47.76	177
Methyleugenol	49.69	178
6-Methyl-alpha-ionone	50.02	121
Ethylguaiacol	50.38	137
beta-Patchoulene	50.54	189
Anisaldehyde	50.71	135
Octanoic acid	51.16	60
beta-Panasinsene	51.30	161
Caryophyllenyl alcohol	51.73	111
Cinnamic acid methyl ester	51.89	131
p-Menth-1,8-diol	52.68	81
Eugenol	54.30	164
4-Ethylphenol	54.49	107
Nonanoic acid	54.50	73
m-Thymol	54.81	135
delta-Selinene	55.30	189
Carvacrol	55.71	135
alpha-Copaene	55.91	119
Elemicin isomer 1	56.34	208
Clovanol	57.66	166
Decanoic acid	57.88	73

(continued)

Table 1
(continued)

Volatile compound	Retention time (min)	m/z
Phenol,2,4-bis(1,1-dimethylethyl)	58.68	191
Chavicol	59.34	134
9-Decenoic acid	59.45	55
Isoeugenol	59.75	164
Methyl jasmonate	59.83	83
Elemicin isomer 2	61.43	208
Apiol	64.07	222
Lauric acid	64.47	73
Hydroxymethyl furfural	64.51	97
Methoxy eugenol	66.20	194
Nootkatone	66.22	147
Benzyl benzoate	70.46	105
Vanillic acid ethyl ester	70.61	151
Nonyl phenol	73.78	135
Tetradecanoic acid	74.83	73
Pentadecanoic acid	82.51	73
Hexadecanoic acid	92.73	73

6. The sample injection system has to be coupled to the chromatograph.
7. The concentrations cover the concentration ranges expected for the various volatile metabolites in vinegar.
8. All the solutions must be stored at 4 °C.
9. The presence of undissolved salt in the sample could damage the stir bar. In addition, the modification of the ionic strength of the medium has a certain influence for a better extraction of the most polar components of the sample so that the NaCl has to be completely dissolved.
10. A good option is to use latex gloves and tweezers to avoid the contact with the hands that could contaminate the stir bar.
11. Due to the magnetic nature of the stir bar, magnetic tweezers can also be employed to facilitate its removal from the sample.
12. Due to the nonpolar nature of the extracting polymer (PDMS), the water is not retained at all.

13. It is very important that the stir bar does not present water remains because it could provoke serious damage in the chromatograph during the ulterior criofocusing step due to the formation of ice particles in the liner.
14. This is the usual injection mode employed in PTV injectors.
15. Although other refrigerating substances can be employed for the criofocusing step, in this case, in order to arrive at such low temperatures (−140 °C), liquid nitrogen is needed. The use of a higher temperature in this step would provoke losses of the most volatile compounds.
16. Full scan mode is usually valid to determine volatile compounds up to ppb concentration levels. To increase the sensitivity, single ion monitoring (SIM) could also be used.
17. In this way, we reduce possible interferences of neighboring compounds. The internal standard is also integrated by its base peak.

References

1. Solieri L, Giudici P (2009) Vinegars of the world. In: Solieri L, Giudici P (eds) Vinegars of the world. Springer, Milan
2. Chinnici F, Guerrero ED, Sonni F, Natali N, Marín RN, Riponi C (2009) Gas chromatography-mass spectrometry (GC-MS) characterization of volatile compounds in quality vinegars with protected European geographical indication. J Agric Food Chem 57: 4784–4792
3. Durán-Guerrero E, Schwarz M, Fernández-Recamales MÁ, Barroso CG, Castro R (2019) Characterization and differentiation of Spanish vinegars from jerez and condado de huelva protected designations of origin. Foods 8:341
4. Marrufo-Curtido A, Cejudo-Bastante MJ, Durán-Guerrero E, Castro-Mejías R, Natera-Marín R, Chinnici F, García-Barroso C (2012) Characterization and differentiation of high quality vinegars by stir bar sorptive extraction coupled to gas chromatography-mass spectrometry (SBSE-GC-MS). LWT Food Sci Technol 47:332–341
5. Castro R, Natera R, Durán E, García-Barroso C (2008) Application of solid phase extraction techniques to analyse volatile compounds in wines and other enological products. Eur Food Res Technol 228:1–18
6. Guerrero ED, Marín RN, Mejías RC, Barroso CG (2006) Optimisation of stir bar sorptive extraction applied to the determination of volatile compounds in vinegars. J Chromatogr A 1104:47–53
7. Baltussen E, Sandra P, David F, Cramers C (1999) Stir bar sorptive extraction (SBSE), a novel extraction technique for aqueous samples: theory and principles. J Microcolumn Sep 11:737–747
8. Pawliszyn J (1997) Solid phase microextraction: theory and practice. Wiley, New York
9. Guerrero ED, Marín RN, Mejías RC, Barroso CG (2007) Stir bar sorptive extraction of volatile compounds in vinegar: validation study and comparison with solid phase microextraction. J Chromatogr A 1167:18–26
10. Guerrero ED, Castro Mejías R, Marín RN, Barroso CG (2007) Optimization of stir bar sorptive extraction applied to the determination of pesticides in vinegars. J Chromatogr A 1165:144–150
11. Cejudo-Bastante C, Castro-Mejías R, Natera-Marín R, García-Barroso C, Durán-Guerrero E (2016) Chemical and sensory characteristics of orange based vinegar. J Food Sci Technol 53: 3147–3156
12. Hevia K, Castro R, Natera R, González-García JA, Barroso CG, Durán-Guerrero E (2016) Optimization of head space sorptive extraction to determine volatile compounds from oak wood in fortified wines. Chromatographia 79: 763–771
13. Durán Guerrero E, Mejías RC, Marín RN, Bejarano MJR, Dodero MCR, Barroso CG (2011) Accelerated aging of a Sherry wine vinegar on an industrial scale employing

microoxygenation and oak chips. Eur Food Res Technol 232:241–254

14. Ruvalcaba JE, Durán-Guerrero E, Barroso CG, Castro R (2019) Development of a stir bar sorptive extraction method to study different beer styles volatile profiles. Food Res Int 126: 108680
15. Durán Guerrero E, Cejudo Bastante MJ, Castro Mejías R, Natera Marín R, García Barroso C (2011) Characterization and differentiation of sherry brandies using their aromatic profile. J Agric Food Chem 59:2410–2415
16. Vasile-Simone G, Castro R, Natera R, Masino F, Barroso CG, Durán-Guerrero E (2017) Application of a stir bar sorptive extraction method for the determination of volatile compounds in different grape varieties. J Sci Food Agric 97:939–948
17. Cejudo-Bastante MJ, Durán-Guerrero E, Natera-Marín R, Castro-Mejías R, García-Barroso C (2012) Characterisation of commercial aromatised vinegars: phenolic compounds, volatile composition and antioxidant activity. J Sci Food Agric 93:1284–1130
18. Herrera C, Castro R, García-Barroso C, Durán-Guerrero E (2016) Development of a stir bar sorptive extraction method for the determination of volatile compounds in orange juices. J Sep Sci 39:3586–3593

Chapter 4

Discovery of Food Intake Biomarkers Using Metabolomics

Leticia Lacalle-Bergeron, David Izquierdo-Sandoval, Juan V. Sancho, and Tania Portolés

Abstract

Due to the high impact of diet exposure on health, it is crucial the generation of robust data of regular dietary intake, hence improving the accuracy of dietary assessment. The metabolites derived from individual food or group of food have great potential to become biomarkers of food intake (BFIs) and provide more objective food consumption measurements.

Herein, it is presented an untargeted metabolomic workflow for the discovery BFIs in blood and urine samples, from the study design to the biomarker identification. Samples are analyzed by liquid chromatography coupled to high-resolution mass spectrometry (LC-HRMS). A wide variety of compounds are covered by separate analyses of medium to nonpolar molecules and polar metabolites based on two LC separations as well as both positive and negative electrospray ionization. The main steps of data treatment of the comprehensive data sets and statistical analysis are described, as well as the principal considerations for the BFI identification.

Key words Untargeted metabolomics, Biomarkers of food intake, Plasma metabolites, Urinary metabolites, Dietary assessment, LC-HRMS, Ion mobility

1 Introduction

Untargeted metabolomics (or metabolic fingerprinting) is a global screening approach focused on the differentiation of samples based on the comparison of metabolic patterns that change in response to an endogenous or exogenous factor, with the final aim of identifying relevant biomarkers of the altered state under study [1]. Diet is one of the environmental exposures with a higher impact on health, and therefore, its study is crucial for the health-related sciences. Biomarkers of food intake (BFIs) are desirable for their ability to assess more accurately the nutritional intake versus traditional self-reported methodologies such as food frequency questionnaires (FFQ), food diaries, or 24-h recalls [2], although their usefulness will depend on the monitored period of time. The measurement of intake and diet should be both accurate and capable of being

Raúl González-Domínguez (ed.), *Mass Spectrometry for Metabolomics*,
Methods in Molecular Biology, vol. 2571, https://doi.org/10.1007/978-1-0716-2699-3_4,
© The Author(s), under exclusive license to Springer Science+Business Media, LLC, part of Springer Nature 2023

applied to a large number of subjects, requirements that, more often than not, these traditional techniques are unable to meet [3]. BFIs can be classified according to the period of time that they reflect the intake: short-term for the past hours/days, medium-term for the past weeks/months; and long-term markers for the past months and even years, being this last one the BFI less studied [4]. For the experimental design for BFI discovery, two approaches can be employed: observational studies and intervention studies. For both approaches, it is highly important the collection of demographic/physiological/lifestyle metadata to reduce bias and have a better control of confounding factors that might increase the inter-subject metabolic variation and, therefore, hide the metabolic changes produced by the food or diet to be assessed [5]. On the one hand, in observational studies two groups are selected from the information obtained from the traditional dietary assessment methods (FFQ, food diaries, etc.), generally low and high consumers of the food/diet under study. Participants are selected from large cohorts, and the groups are compared at a single time point (cross-sectional study). The selection of the participants has to be exhaustive based on the demographic/physiological/lifestyle metadata of the candidates to reduce the impact of confounding variables in the study. On the other hand, intervention studies generally involve the consumption of a specific standardized diet and/or food by the participants over a defined period of time. The samples are obtained at specific time points, depending on the aim of the study. Among interventional studies, in cross-over trials the participants receive all the treatments to be compared, reducing the inter-subject variation. Between one treatment and another, a washout is highly recommended (short period of time returning to the habitual diet or avoiding the food of interest) in order to return to basal metabolic levels and avoid carryover [6].

Most of the BFIs found are compounds metabolically inert (identical to the compound found in a specific food) or compounds metabolically transformed by endogenous processes (or gut microbiota) [2]. Blood, urine, and feces are the samples most employed for the determination of BFI as they are easily accessible and their use for monitoring the food intake will be easier through them since they are common samples taken in medical check-ups.

Due to the aqueous properties of usual biofluids, such as blood and urine, and the capability for resolving various classes of metabolites (the medium-high polar, low volatility, and/or thermolabile compounds), liquid chromatography (LC) has been the most employed chromatographic separation technique in the BFI discovery through untargeted metabolomics [6]. Its coupling to high-resolution mass spectrometry (HRMS) allows the detection and identification of metabolites previously separated with high sensitivity and selectivity as well as the acquisition of accurate-mass full-spectrum data [7]. Moreover, ion mobility separation (IMS)

recently coupled to HRMS instruments provides additional structural information, enhancing the characterization of biomarkers [8]. This technique measures the drift time (DT), which will depend on the size, shape, and charge of each ionized molecule, and translates into the collision cross-section value (CCS, $Å^2$) of each ion [9]. Additionally, IMS cell layout prior to hybrid HRMS analyzers allows High Definition MS^E acquisition ($HDMS^E$). As conventional MS^E, two functions are acquired at low and high collision energy, but it is possible to obtain cleaner fragmentation spectrum without interfering fragments of co-eluting ions by filtering with the DT of a target molecule [10].

This chapter explains a protocol for the application of untargeted metabolomics in the determination of BFI in blood and urine by LC-HRMS, with additional clues when the recent LC-IMS-HRMS instruments are employed. Specifically, the experiment design is focused on a cross-over intervention study for the determination of short-term BFIs. Then, the experimental details about the sample treatment, data acquisition, data treatment, statistical analysis, and elucidation are described, based on previous works and protocols employed in our laboratory.

2 Materials

2.1 Equipment and Instrumentation

- Vortex mixer.
- Microcentrifuge.
- Ultrasonic bath.
- Micropipette (100 μL and 1 mL).
- Liquid chromatography system: ACQUITY UHPLC I-Class system (Waters, Milford, MA, USA).
- Electrospray ionization interface (ESI).
- High-resolution mass spectrometer: VION® IMS QTof (Waters, Manchester, UK) ion mobility hybrid Quadrupole Time-of-Flight (IMS-QTOF).

2.2 Solvents and Chemicals

- Water (prepared by purifying deionized water to attain a sensitivity of 18 MΩ-cm at 25 °C) and LC-MS grade methanol (MeOH) and acetonitrile (ACN).
- Formic acid (HCOOH) (LC-MS grade) and ammonium acetate (NH_4Ac).
- 50 ng/mL of leucine-enkephalin in ACN:water 50:50 (v/v) plus 0.01% HCOOH.
- Mobile phase A for medium to nonpolar molecules (reversed-phase liquid chromatography, RP-LC): water plus 0.01% of HCOOH.

- Mobile phase B for medium to nonpolar molecules (RP-LC): MeOH plus 0.01% of HCOOH.
- Mobile phase A polar molecules (hydrophilic interaction liquid chromatography, HILIC): ACN:water (95:5, v/v) plus 0.01% of HCOOH and 10 mM NH_4Ac.
- Mobile phase B polar molecules (HILIC): water plus 0.01% of HCOOH and 10 mM NH_4Ac.

2.3 Supplies

- CORTECS® C18 fused-core 2.7-μm particle size analytical column 100 × 2.1 mm (Waters) for medium to nonpolar molecules (RP-LC).
- CORTECS® HILIC fused-core 2.7-μm particle size analytical column 100 × 2.1 mm (Waters) for polar molecules.
- Pipette tips (5–200 μL and 100–1000 μL).
- 9-mm screw-top vial (2 mL and high recovery)
- 9-mm screw-top bonded in PTFE/silicone pre-slit cap.
- Eppendorf™ polypropylene graduated microtubes.

2.4 Software

- UNIFI 1.9.4 software (Waters, Manchester, UK).
- Progenesis QI v3.0 (NonLinear Dynamics, Waters, Newcastle, UK).
- EZinfo 3.0 (Umetrics, Sweden).

3 Methods

3.1 Intervention Cross-Over Trial Study Design and Sampling

1. Define the study food or diet, groups, sample size, and target population.
2. Recruit the participants of the study to conform to a homogeneous group based on demographic/physiological/lifestyle metadata collected, ensuring that covariates and/or cofounders are well characterized and properly distributed across the group (*see* **Note 1**).
3. Randomly divide the participants to start either in the intervention group (to which the study food/diet is administered) or the control group (receiving an equivalent isocaloric solution).
4. Administer to the participants the study food or the isocaloric control after fasting a minimum of 8 h. Collect plasma and/or urine samples at the start of the intervention (t = 0) and 4 h after. No other food is allowed during the intervention. More than one sampling time point can be settled.

5. After the washout period (e.g., a week), the groups are swapped over, and the participants perform the other part of the study (*see* step 4).
6. Store the samples at −80 °C until sample treatment for subsequent metabolomics determinations.

3.2 Sample Treatment

Carry out all procedures at room temperature unless otherwise specified. Samples must be thawed and vortexed at room temperature prior to sample treatment, avoiding long waiting times under these conditions. The order of the samples is randomly selected.

1. Add a total of 400 μL of ACN to 100 μL of sample in a microtube, then vortex for 30 s, and sonicate for 15 min (*see* **Note 2**).
2. Collect the supernatant after centrifuging at 12000 rpm with a radius of 5.5 cm, 8855 g (RCF) for 15 min at 4 °C and store at −20 °C for 1 h to promote protein precipitation (*see* **Note 3**).
3. Perform an additional centrifugation under the same above-mentioned conditions and collect the supernatant (*see* **Note 4**).
4. Aliquot the final extracts in three parts: two vials of 100 μL stored at −20 °C, to be directly injected into the instrument, and one vial of 200 μL stored at −80 °C as a backup.
5. Additionally, pool and mix 20 μL of all final extracts to generate the Quality Control sample (QC) in order to provide a representative average sample.

3.3 Instrumental Analysis

3.3.1 Ultra-High-Performance Liquid Chromatography (UHPLC) Conditions

1. Prepare the mobile phases required (A and B) for each chromatographic separation.
2. Mount the column and equilibrate following the manufacturer's recommendations.
3. Set the column oven at 40 °C and the sample manager at 10 °C.
4. Set the flow rate at 0.3 mL/min and injection volume at 1 μL.
5. The gradient elution for RP-LC (medium to nonpolar molecules) changes from 10% B at 0 min to 90% B at 14 min, 90% B at 16 min, and 10% B at 16.01 min, with a total run time of 18 min. The gradient is the same for both ESI+ and ESI- ionization modes.
6. For HILIC separation, the gradient starts with 2% B until 1 min, 60% B at 10 min, 60% B at 12 min, and finally 2% B at 12.01 min, with a total run time of 15 min. The gradient is the same for both ESI+ and ESI- ionization modes.
7. Inject at least 10 QC samples at the beginning of each sample batch for column stabilization (*see* **Note 5**).

8. Samples are analyzed in a randomized order including a QC sample every ten injections to control possible instrumental drift along the sequence.

3.3.2 High-Resolution Mass Spectrometry (HRMS) Setup

1. Perform routine maintenance and MS instrument calibration (mass axis and CCS calibration if applicable) according to the instrument manufacturer's guidelines (*see* **Note 6**).
2. Set the capillary voltage to 0.7 kV and 2.0 kV for positive (ESI +) and negative (ESI-) electrospray ionization modes, respectively.
3. Set the cone voltage to 30 V, source temperature to 120 °C, desolvation gas at 550 °C, and flow rate of 1000 L/h.
4. Use nitrogen as desolvation, nebulizing, and collision gas (also as mobility gas).
5. HRMS acquires in MS^E mode (with ion mobility in $HDMS^E$). Set a fixed collision energy to 6 eV for low energy function (LE), and a collision energy ramp from 28 to 58 eV for high energy function (HE). Both collision energy functions are acquired sequentially over an *m/z* range of 50–1000 Da and a scan time of 0.3 s.
6. For recalibrating the mass axis and ensuring robust and accurate measurements along runs, pump the leucine-enkephalin solution at 20 μL/min through the lock-spray needle.

3.4 Data Processing

The aim of the data processing is to extract the information of the detected features from the LC-HRMS raw 3D data (or 4D with the incorporation of IMS) and obtain a 2D data matrix to be used for statistical analysis. All calculations are performed using statistical packages developed on free-ware platforms such as the R language, except when the data is acquired with ion mobility, where proprietary software should be used, such as Nonlinear QI Progenesis (Waters, UK). Data matrix will be characterized by the following information of features across the samples: *m/z* ratio, retention time (RT), relative intensities, and CCS values (this last in the case of IMS instruments). The main steps are as follows:

1. Import your instrument data to the processing software used for data treatment in the appropriate format.
2. Perform the peak picking and deconvolution by the detection of each measured ion in a sample and the assignation of a feature (*m/z*, RT, and CCS). The peak picking algorithm and deconvolution parameters will consider maximum *m/z* error, minimum and maximum time width for a chromatographic peak, the minimum height or intensity, and signal to noise ratio (S/N), among other parameters. To apply the deconvolution tool, based on our previous experiences, a selection of

adduct ions species are set: $[M+H]^+$,$[M-H_2O+H]^+$, $[M+Na]^+$, and $[M+K]^+$ (and even dimmers as $[2M+H]^+$ and $[2M+Na]^+$) for positive ionization analysis and $[M-H]^-$, $[M-H_2O-H]^-$, $[M+Cl]^-$, and $[M+HCOOH-H]^-$ for negative ionization analysis.

3. Apply the retention time alignment. The matched peaks with similar retention times and *m/z* ratio across multiple samples are grouped in accordance with a window of *m/z* and RT, to be assigned as the same feature across samples. In the case of Progenesis QI, use the software menu to automatically perform the retention time alignment using the QC samples as reference (do not use the first QC injections employed for column stabilization).
4. Normalize the areas. Among the different types of normalization that can be employed, the normalization "to all compound" from Progenesis QI software will correct the possible signal loss that the equipment may experience along the sequence, based on the intensity of the compounds found throughout the chromatogram.
5. Gap filling: It is applied to correct and fill in the missing peaks (0 signal) or peaks not detected due to the restrictions of the first two steps (lower intensities or bad peak shape in some of the samples), which can affect the power of subsequent statistical analysis.

3.5 Statistical Analysis

Although the main part of statistical data treatment in untargeted metabolomics is based on multivariate statistical analysis, there are several aspects to consider before in advance. Taking into account that in this kind of study the number of variables largely exceeds the number of objects, data cleaning is highly recommendable by the application of one or successive pre-filtration steps to reduce the number of features and eliminate irrelevant signals while avoiding or minimizing relevant chemical information loss.

1. Features that exhibit poor stability, meaning relative standard deviation (%RSD or CV%>30%) on peak area across the QCs, can be removed. Also, those features that show a low fold change or no significant difference among sample groups or among blank runs and any of the sample groups may be avoided.
2. Divide the samples into four groups (Food t = 0 h, Food t = 4 h, Isocaloric Beverage (IB) t = 0 h, IB t = 4 h).
3. As the study carried out is an intervention cross-over study, filter the data by means of repeated measures ANOVA *p-value* ≤ 0.05 to reduce the individual differences and focus on the potential markers (*see* **Note 7**).

4. These tests should be followed by a False Discovery Rate calculation *p-FDR* < 0.05 (*q-value* set at 0.05) normally applying the Benjamini–Hochberg procedure to rectify *p-values*.
5. Load the data matrix to the statistical software (EZ info) for multivariate analysis.
6. Use principal component analysis (PCA) to interrogate the data and to explore the inherent separation of the groups, without the need to previously provide the information of the groups. Verify that the samples are correctly normalized and analyzed QC samples must be grouped in the center of the PCA loading graph since they do not belong to any specific group and must be equidistant from the rest.
7. Perform an OPLS-DA (orthogonal partial least square—discriminant analysis) facing the samples collected 4 h after food intake vs. the other samples. Information of the compounds that are statistically more different between the groups faced will be obtained. The usually minimum accepted limits for variance explained (R^2Y) and variance predicted (Q^2) for biological models are 50% and 40%, respectively [11].
8. Use S-Plot generated from the OPLS-DA analysis to select those features with more discrimination power between the two groups, e.g., those with a *p(corr)* value near 1 or −1. Use a threshold *p(corr)* $\geq |0.6|$ for further identification.

3.6 Biomarker Identification

Structural characterization and elucidation of potential markers highlighted in the statistical data analysis is commonly a challenge in untargeted metabolomics and can become the bottleneck of the overall metabolomics process. In HRMS, accurate mass measurement is the gold standard for the identification procedure, and it is essential for succeeding in this process.

1. First recognize the (quasi-)molecular ion in the accurate mass spectrum (typically, protonated or deprotonated molecule in LC-MS), where the presence of adducts must be also taken into account. Progenesis QI, as well as other tools such as CAMERA for XCMS, allows componentization, which means that different signals from the same metabolite are grouped together offering greater confidence to the annotation.
2. Calculate the most likely elemental composition according to the mass error and isotope pattern. After that, use fragment ion information based on MS/MS or DIA (MS^E or $HDMS^E$ acquisition mode) data to establish the fragmentation pathways and discard possible chemical structures. To this aim, the use of offline/online and commercially/freely/in-house available spectral databases are of great help. In-silico fragmentation tools might also be useful (*see* **Notes 8** and **9**).

3. Every annotation for each marker has to be carefully reviewed. Assure the full identification of the marker by the injection of a reference standard, if commercially available.

4 Notes

1. The study protocol and procedures are required to be approved according to the ethical standards of the Helsinki Declaration, and subjects are required to provide written informed consent.
2. It is recommended to perform the sample treatment in a randomized order and treat and analyze the samples in a single batch. If the samples have to be analyzed in different batches (different days) or very large number of samples, it is highly advisable the use of internal standards.
3. As protein presence in urine samples is considerably lower than in the plasma sample, the time in the freezer can be reduced (e.g., only 30 min) and the subsequent centrifugation can be avoided.
4. In the case of desiring to pre-concentrate the samples or if a volume injection above 5 μL (e.g., higher sensitivity is needed) is required, the following steps are recommended:
   (a) Since two types of chromatography are performed in this protocol, aliquot 220 μL in two Eppendorfs.
   (b) Bring the extracts to dryness with a stream of nitrogen at a temperature between 30 and 40 °C.
   (c) Reconstitute each aliquot with a solvent mixture according to the initial conditions of the following chromatographic separation employed. For pre-concentration, reconstitute with a volume less than 220 μL.
   (d) Divide each extract into two aliquots of equal volume (one to be injected and the other to store) and a smaller one to generate the QC pooled sample. In the case of having reconstituted with the same initial volume, 100 μL is to be stored at −20 °C until injection, 100 μL is to be stored at −80 °C, and 10 μL is to generate the QC.
5. While for RP-LC, 10 QC samples are injected at the beginning of the batch for column stabilization, it is recommended to increase the number at least to 15 for HILIC.
6. Inject the system-suitability standard solution to confirm that the instrument is properly calibrated. System-suitability standard solution contains a selection of metabolites at the concentration of 10 μg/mL considering the mass range that will be used during the metabolomics analysis (e.g., acetaminophen,

caffeine, sulfaguanidine, verapamil, terfenadine, leucine enkephalin, reserpine, and Val-Tyr-Val). This standard solution is employed to check the stability of the exact mass and CCS measurements, and it is analyzed before and after each metabolomic batch analysis (in both types of LC and both electrospray ionizations).

7. In this type of experimental design, where samples of each participant are taken under different conditions, it has to be specified not only the time the sample was taken but also from which subject comes from. In these cases, the standard ANOVA is not appropriate because the data violates the independence ANOVA assumption. It must be specified not only the groups each sample belongs to but also the subject from which it comes.
8. The following are the most employed databases in food-related studies with untargeted metabolomics: METLIN repository database (https://metlin.scripps.edu), Human Metabolome Database (HMDB) (http://www.hmdb.ca), Kyoto Encyclopedia of Genes and Genomes (KEGG) (https://www.genome.jp/kegg/), FooDB (https://foodb.ca), Chemspider (http://www.chemspider.com/), PubChem Compound database, LIPID Metabolites and Pathways Strategy (LIPID MAPS) (https://www.lipidmaps.org/), Chemical Entities of Biological Interest (ChEBI), SIRIUS, CSI:Finger ID, MassBank (http://www.massbank.jp), and FooDB (http://foodb.ca). The following are the in-silico fragmentation tools used: MetFrag (https://msbi.ipb-halle.de/MetFrag/), MetFusion, and Competitive Fragmentation Modeling for metabolite identification (CFM-ID) (http://cfmid.wishartlab.com/), among others.
9. When IMS information is available, CCS values can be compared with online, commercial, or home-made CCS libraries. Moreover, there are more and more CCS prediction tools that can be employed to increase the confidence in the identification [12–14].

References

1. Weckwerth W (2003) Metabolomics in sytems biology. Annu Rev Plant Biol 54:669–689. https://doi.org/10.1146/annurev.arplant.54.031902.135014
2. Maruvada P, Lampe JW, Wishart DS, Barupal D, Chester DN, Dodd D, Djoumbou-Feunang Y, Dorrestein PC, Dragsted LO, Draper J, Duffy LC, Dwyer JT, Emenaker NJ, Fiehn O, Gerszten RE, Hu FB, Karp RW, Klurfeld DM, Laughlin MR, Little AR, Lynch CJ, Moore SC, Nicastro HL, O'Brien DM, Ordovás JM, Osganian SK, Playdon M, Prentice R, Raftery D, Reisdorph N, Roche HM, Ross SA, Sang S, Scalbert A, Srinivas PR, Zeisel SH (2019) Perspective: dietary biomarkers of intake and exposure—exploration with omics approaches. Adv Nutr 11:200–215. https://doi.org/10.1093/advances/nmz075
3. O'Gorman A, Brennan L (2017) The role of metabolomics in determination of new dietary biomarkers. Proc Nutr Soc 76:295–302. https://doi.org/10.1017/S0029665116002974

4. Hedrick VE, Dietrich AM, Estabrooks PA, Savla J, Serrano E, Davy BM (2012) Dietary biomarkers: advances, limitations and future directions. Nutr J 11:1. https://doi.org/10.1186/1475-2891-11-109
5. Dunn WB, Wilson ID, Nicholls AW, Broadhurst D (2012) The importance of experimental design and QC samples in large-scale and MS-driven untargeted metabolomic studies of humans. Bioanalysis 4:2249–2264. https://doi.org/10.4155/bio.12.204
6. Lacalle-Bergeron L, Izquierdo-Sandoval D, Sancho JV, López FJ, Hernández F, Portolés T (2021) Chromatography hyphenated to high resolution mass spectrometry in untargeted metabolomics for investigation of food (bio)-markers. TrAC Trends Anal Chem 135: 116161. https://doi.org/10.1016/j.trac.2020.116161
7. Castro-Puyana M, Pérez-Míguez R, Montero L, Herrero M (2017) Application of mass spectrometry-based metabolomics approaches for food safety, quality and traceability. TrAC Trends Anal Chem 93:102–118. https://doi.org/10.1016/j.trac.2017.05.004
8. Segers K, Declerck S, Mangelings D, Vander HY, Van EA (2019) Analytical techniques for metabolomic studies: a review. Bioanalysis 11: 2297–2318. https://doi.org/10.4155/bio-2019-0014
9. Mairinger T, Causon TJ, Hann S (2018) The potential of ion mobility–mass spectrometry for non-targeted metabolomics. Curr Opin Chem Biol 42:9–15
10. Paglia G, Smith AJ, Astarita G (2021) Ion mobility mass spectrometry in the omics era: challenges and opportunities for metabolomics and lipidomics. Mass Spectrom Rev: mas.21686. https://doi.org/10.1002/mas.21686
11. Worley B, Powers R (2012) Multivariate analysis in metabolomics. Curr Metabolomics 1: 92–107. https://doi.org/10.2174/2213235X130108
12. Bijlsma L, Bade R, Celma A, Mullin L, Cleland G, Stead S, Hernandez F, Sancho JV (2017) Prediction of collision cross-section values for small molecules: application to pesticide residue analysis. Anal Chem 89: 6583–6589. https://doi.org/10.1021/acs.analchem.7b00741
13. Zhou Z, Tu J, Xiong X, Shen X, Zhu Z-J (2017) LipidCCS: prediction of collision cross-section values for lipids with high precision to support ion mobility–mass spectrometry-based lipidomics. Anal Chem 89:9559–9566. https://doi.org/10.1021/acs.analchem.7b02625
14. Plante P-L, Francovic-Fontaine É, May JC, McLean JA, Baker ES, Laviolette F, Marchand M, Corbeil J (2019) Predicting ion mobility collision cross-sections using a deep neural network: DeepCCS. Anal Chem 91: 5191–5199. https://doi.org/10.1021/acs.analchem.8b05821

Chapter 5

Metabolomic Characterization of the Antiproliferative Activity of Bioactive Compounds from Fruit By-Products Against Colon Cancer Cells

Gerardo Alvarez-Rivera, Alberto Valdés, and Alejandro Cifuentes

Abstract

This methodological work demonstrates the potential of metabolomic approaches based on liquid chromatography coupled to high-resolution mass spectrometry (LC-ESI(+/−)-HRMS) to investigate the antiproliferative capacity of underexplored biomasses (e.g., *Passiflora mollissima* seeds and *Physalys peruviana* calyx), by evaluating the molecular changes induced at the metabolite expression levels on HT-29 human colon cancer cells. This protocol describes in detail the optimal conditions to obtain bioactive extracts by pressurized liquid extraction (PLE), the experimental procedure to grow and treat HT-29 human colon cancer cells and CCD-18Co normal human colon fibroblasts with the target extracts, the metabolites extraction from the cytosolic fraction, and subsequent metabolomic fingerprinting. After treatment for 48 and 72 h, the viability of HT-29 colon cancer cells is markedly affected, and metabolites can be extracted for investigation. Following the proposed metabolomic data analysis and interpretation workflow, altered cellular redox homeostasis, as well as inactivation or dysfunction on other metabolic pathways, constitutes valuable biological information to understand the mechanisms underlying the antiproliferative effect.

Key words Colon cancer, Antiproliferative activity, Metabolomics, Liquid chromatography, High-resolution mass spectrometry, Food by-products, Metabolites extraction, Cytosolic fraction, Bioactive compounds

1 Introduction

Metabolomics is the scientific study of metabolites, or the end products of cellular processes, which are defined as those small molecules with a relative molecular weight of less than 1000 Da. Metabolomics studies can be performed on different biological materials, such as fluids (blood, urine, saliva, etc.) or solids (tissue, cells, feces, etc.), from human, animal, or plant origin. A typical workflow in metabolomics research involves sample preparation, sample analysis, data processing, statistical analysis, compound identification, and data interpretation. Sample preparation is a

Raúl González-Domínguez (ed.), *Mass Spectrometry for Metabolomics*,
Methods in Molecular Biology, vol. 2571, https://doi.org/10.1007/978-1-0716-2699-3_5,
© The Author(s), under exclusive license to Springer Science+Business Media, LLC, part of Springer Nature 2023

critical step because it varies depending on the analytical method and the type of metabolite to be analyzed. Due to the broad physicochemical diversity of the metabolome and the wide concentration range of the metabolites in the biological samples, a plethora of extraction methods can be used mainly based on the polarity of the metabolites. A mixture of solvents such as methanol/water, acetonitrile/water, or a two-phase extraction using chloroform or methyl-tert-butyl ether and water are frequently used to increase the coverage for metabolite extraction, but not a single method or platform can extract the full metabolome [1, 2]. Apart from the sample preparation, different separation techniques such as gas chromatography (GC), liquid chromatography (LC), or capillary electrophoresis (CE), as well as different detection techniques, such as nuclear magnetic resonance (NMR) or high-resolution mass spectrometry (HR-MS), can be used for sample analysis. For instance, GC-MS is ideal for identifying and quantitating small acids, alcohols, hydroxyl acids, amino acids, sugars, fatty acids, sterols, catecholamines, drugs, and toxins. However, some compounds cannot be analyzed by GC-MS, and several LC-MS-based analytical methods have been developed for this aim [1, 2]. One specific and important aspect in LC-MS analyses is the selection of the stationary phase, and reverse-phase (RP) stationary phases are frequently used to study nonpolar metabolites (such as nonpolar vitamins, sterols, or triacylglycerols), while hydrophilic interaction liquid chromatography (HILIC) is preferred for the study of very-polar metabolites (such as amino acids, sugars, or acylcarnitines). Another important parameter to be selected when analyzing metabolites is the ionization mode. Electron ionization (EI) is usually selected for the analyses of small, nonpolar and volatile compounds by GC-MS, and soft ionization techniques, including electrospray ionization (ESI) and atmospheric pressure chemical ionization (APCI), are used to ionize thermally labile and moderately polar organic analytes [3].

Together with the above-mentioned analytical advances, several bioinformatics tools and processing strategies have been developed to process a large amount of generated analytical data sets. Considering the complexity of the metabolomics data matrices, containing thousands of m/z features, data processing and data pre-treatment, including noise filtering, overlapping the peak resolution, peak alignment, peak matching, and normalization, is an essential requirement to allow the identification of significant metabolites. Subsequent multivariate data analysis for pattern recognition usually involves unsupervised models and supervised classification tools. Unsupervised models including principal component analysis (PCA), cluster analysis (HCA), and nonlinear mapping (NLM) are used as a first step in the data analysis to detect sample clustering in the measured data. Afterward, supervisory models such as linear discriminant analysis (LDA), partial least

discriminant analysis (PLS-DA), or orthogonal partial least discriminant analysis (OPLS-DA) can be used as statistical model validation in order to find differences between the known groups and to detect the differential metabolites [4].

The main limitation still resides on the identification and accurate quantification of the metabolites. Metabolite identification is usually performed by calculating the molecular formula (based on the exact mass obtained from high-resolution MS instruments), and by comparing the experimental MS/MS spectra against EI mass spectral fragmentation or MS/MS fragmentation databases [5]. Nowadays, the most important metabolomics databases are METLIN, Human Metabolome Database (HMBD), Mass Bank of North America (MoNA), NIST, and mzCloud, which have been continuously growing during the last decade, both in coverage and chemical diversity. However, confident metabolite identification is still a bottleneck in the metabolomics process. For this reason, a combination of approaches is required, including new analytical strategies, computational algorithms, and database resources, as well as the development of biostatistical methods [6], and the improvement and application of quantum chemistry and computational methods for elucidating or predicting the structures of novel metabolites [7]. Finally, another major challenge in the metabolomic field is the biological interpretation of the observed metabolic changes since most metabolites can potentially participate in multiple functional roles within a biological system, and several tools and algorithms have been programmed to solve such problems [8, 9].

In this regard, the present methodological proposal describes the optimal conditions to obtain bioactive extracts by pressurized liquid extraction (PLE), the experimental procedure to grow and treat HT-29 human colon cancer cells and CCD-18Co normal human colon fibroblasts, the metabolites extraction from the cytosolic fraction, and the subsequent metabolomic fingerprinting analysis using liquid chromatography coupled to high-resolution mass spectrometry (LC-HRMS), following a recursive analysis strategy to investigate the antiproliferative capacity of *Passiflora mollissima* seeds and *Physalys peruviana* calyx at the metabolite expression level [10, 11].

2 Materials

2.1 Pressurized Liquid Extraction

1. Knife mill (Grindomix GM200-Retsch GmbH, Haan, Germany).
2. Assisted solvent extractor (Dionex™ ASE™ 200) equipped with an 11-mL stainless steel extraction cell.
3. Sea sand (50–70 mesh).
4. Ethyl acetate, cyclohexane, and ethanol.

2.2 Cell Lines and Antiproliferative Activity Assay

1. HT-29 (human colon adenocarcinoma) and CCD-18Co (normal human colon fibroblast) cells from the American Type Culture Collection.
2. Growth medium: RPMI 1640 medium supplemented with 25 mM HEPES, 2.05 mM L-glutamine, 10% fetal bovine serum, and 50 μg mL^{-1} gentamicin.
3. PBS solution: 138 mM sodium chloride, 2.7 mM potassium chloride, and 10 mM sodium hydrogen phosphate at pH 7.4.
4. 1.0% Triton X-100 in PBS solution.
5. 0.1% dimethyl sulfoxide in RPMI medium.
6. MTT solution: 0.25 mg mL^{-1} 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide in RPMI medium.
7. Microtiter plate reader at 570 nm (Tecan, Infinite® 200 PRO, Austria GmbH).
8. GraphPad Prism 7.0 software (GraphPAD Corp., San Diego, CA, USA).

2.3 Metabolites Extraction from the Cytosolic Fraction of HT-29 Cells

1. Distilled water from the Milli-Q system (Millipore, Bedford, MA, USA).
2. PBS solution.
3. 0.1% dimethyl sulfoxide in RPMI 1640 medium.
4. 0.3 g glass beads (212–300 μm).
5. Liquid nitrogen for metabolite quenching.
6. Centrifuge (Heraeus Fresco 21, Thermo Scientific).
7. Centrifugal filters (3 kDa Amicon Ultra 0.5 mL, Millipore, Billerica, MA, USA).

2.4 UHPLC-QTOF-MS/MS Analysis

1. Agilent 1290 ultrahigh-performance liquid chromatography system (Agilent Technologies, Santa Clara, CA, USA).
2. Agilent 6540 quadrupole-time-of-flight mass spectrometer (q-TOF MS) controlled by a PC equipped with MassHunter Workstation data acquisition software (Agilent).
3. Orthogonal electrospray source (ESI Agilent Jet Stream, AJS, Santa Clara, CA, USA) operated in positive (ESI+) and negative (ESI−) ionization modes.
4. Zorbax Eclipse Plus C18 column (2.1 × 100 mm, 1.8 μm particle diameter, Agilent Technologies, Santa Clara, CA) at 40 °C.
5. Mobile phase A: Milli-Q water with 0.1% v/v formic acid.
6. Mobile phase B: Acetonitrile with 0.1% v/v formic acid.

2.5 Software and Databases for Data Processing and Interpretation

1. MassHunter software from Agilent.
2. Mass Profiler Professional (MPP) software from Agilent.
3. *HMDB—Human metabolome database (http://www.hmdb.ca/).*
4. *METLIN database (http://metlin.scripps.edu/).*
5. *NIST database (https://chemdata.nist.gov).*
6. *KEGG compound database (http://www.genome.jp/kegg/compound/).*
7. *KEGG Pathway database (http://www.genome.jp/kegg/pathway.html).*
8. *SMPDB—The Small Molecule Pathway Database (https://smpdb.ca).*
9. *PubChem (http://pubchem.ncbi.nlm.nih.gov/).*

3 Methods

3.1 Extraction of Bioactives from Fruit By-Products

1. Wash and dry the target fruit by-product samples (e.g., banana passion fruit seeds and goldenberry calyx) at 60 °C for 12 h in the darkness. Ground samples with knife mill using dry ice and then vacuum-pack and store at −20 °C [12, 13].
2. Mix the powdered sample with sea sand (1:2 w/w) into the extraction cell for pressurized liquid extraction.
3. Defat *Passiflora mollissima* seeds by pressurized liquid extraction, using cyclohexane at 100 bar and 100 °C for 20 min. Extract polar fractions of *Passiflora mollissima* seeds using ethanol as solvent at 150 °C and 100 bar for 20 min.
4. Extract polar fractions of goldenberry calyx in static mode using an ethanol:ethyl acetate mixture (75:25 v/v) as solvent at 125 °C and 100 bar for 20 min.

3.2 Cell Lines Treatment with Tropical Fruit By-Products Extracts

1. Grow human HT-29 cells in growth medium. Maintain in a humidified atmosphere containing 5% CO_2 at 37 °C.
2. Dissolve the extract in dimethyl sulfoxide to make a 10 mg mL^{-1} concentration, and make dilutions using cell culture medium to final concentrations of 100, 50, 25, 12.5, and 6.25 μg mL^{-1}.
3. Seed HT-29 and CCD-18Co cells in 96-well plates at a density of 1.0×10^4 cells/well and 4.5×10^3 cells/well, respectively, allowing the attachment for 24 h.
4. Treat the cells with the series of prepared extracts, using 1.0% Triton X-100 as the positive control, and 0.1% dimethyl sulfoxide in RPMI medium as the untreated control for 24, 48, and 72 h.

5. At the end of the treatment period, replace the culture medium by 100 μL of MTT solution and incubate the plates for 4 h.
6. Discard the medium and add 100 μL of dimethyl sulfoxide.
7. Measure the absorbance of formazan produced using a microtiter plate reader at 570 nm.
8. Express cell viability as a percentage of live cells relative to controls and plot dose-response curves of the percentage of cell viability against concentrations of the extracts.

3.3 Metabolites Extraction Procedure

1. Wash cells with PBS solution and centrifuge at 1000 g for 10 min at 25 °C to form the pellets.
2. Remove supernatant and resuspend cells in 300 μL of PBS solution.
3. Count cells using the trypan blue exclusion method.
4. Centrifuge the estimated volume at 1000 g for 10 min at 25 °C to have 10×10^6 cells for metabolite recovery [14].
5. Mix cell pellets with 300 μL of ultrapure water and 0.3 g of glass beads (212–300 μm).
6. Vortex the mixture and snap-frozen in liquid nitrogen for 1 min for metabolite quenching.
7. Remove glass beads by centrifugation at 24,000 g for 10 min at 4 °C. Discard the pellet (nuclear fraction) and collect the supernatant (cytosolic fraction).
8. Add 350 μL of methanol to 150 μL of cytosolic fraction obtained from the cell culture.
9. Incubate the solution at −20 °C for 2 h to reduce protein solubilization.
10. Centrifuge the suspension for 14,000 × *g* for 10 min at 4 °C.
11. Discard the pellet and collect the supernatant fraction.
12. Store the supernatant (cytosolic fraction) at −80 °C until analysis.

3.4 UHPLC-QTOF-MS/MS Metabolic Analysis of Human HT29 Colon Cancer Cells

1. Run the samples using the following elution gradient in the HPLC system: 0–30% B in 0–10 min, 30–80% B in 10–10.5 min, 80–100% B in 10.5–11 min, 100% B in 11–13 min, and 0% B in 13–14 min.
2. Set the flow rate at 0.5 mL min^{-1} and the sample injection volume at 2 μL.
3. Operate the QTOF-MS system in MS and MS/ MS modes using the following parameters: capillary voltage, 3000 V; nebulizer pressure, 40 psi; drying gas flow rate, 11 L/min; gas temperature, 300 °C; skimmer voltage, 45 V; fragmentor voltage, 110 V.

4. Operate the ESI source in positive (ESI+) and negative (ESI−) ionization modes to widen the range of detectable metabolites (*see* **Note 1**).
5. Set MS and Auto MS/MS modes to acquire m/z values ranging between 50–1100 and 50–800 Da, respectively, at a scan rate of 5 spectra per second. Make internal calibration by introducing the calibrant into the ESI source (*see* **Note 2**).
6. Select 4 precursor ions per cycle at a threshold of 200 counts in Auto MS/MS mode.
7. Randomize the sequence of samples, including three experimental replicates of control and treated cell experiments.
8. Regularly inject a pool of all samples throughout the sample sequence as quality control to monitor the stability of the analysis.
9. After each run, inject a blank sample to avoid the carryover effect in the chromatographic system.
10. Each sample is injected in triplicate by UHPLC-QTOF-MS/MS.

3.5 Metabolomic Fingerprinting Data Analysis

1. Use the MassHunter Qualitative (MHQ) and Mass Profiler Professional (MPP) software from Agilent for post-acquisition HRMS data treatment, following a recursive analysis strategy (*see* Fig. 1).
2. Select a list of MS features for each sample in MHQ software running an untargeted MS feature detection method (cutoff threshold > 200 counts).
3. Export the extracted MS features to MS exchange format file (.cef files).
4. Import the previous .cef files in MPP software for convenient peak alignment (RT window: 0.2 min and Mass window: 20 ppm).
5. Generate a defined list of metabolites in MPP and export for targeted data processing and metabolites detection in MHQ software (.cef files).
6. After recursive analysis, generate a table of metabolite ion intensities filtered by frequency (retaining entities that appear in at least 20% of the samples) in MPP software and exported as .txt file for further filtering.
7. Remove duplicated entities and peaks showing a low signal to noise ratio (average of samples/average of blanks < 3.0).
8. Apply gap filling (minimum value/2) before normalization.
9. Export the resulting data table (.csv file) of high-quality time-aligned detected peaks with their corresponding retention time, m/z, and peak intensity obtained for each sample.

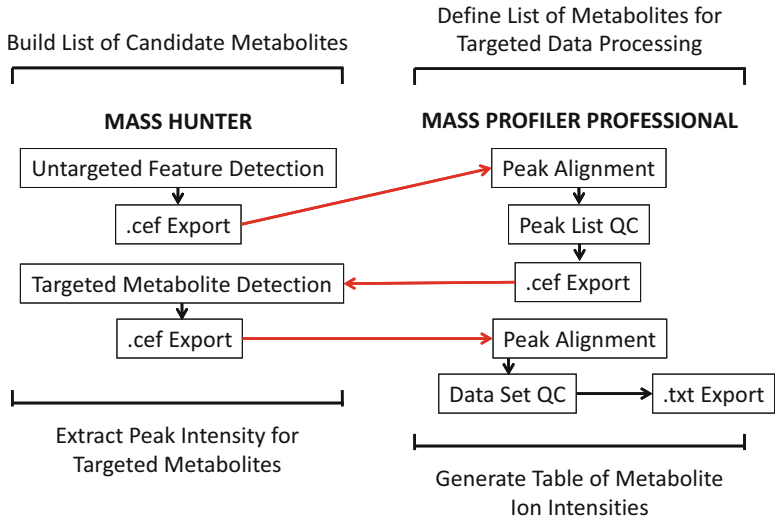

Fig. 1 Untargeted data processing strategy with Agilent software. Recursive analysis with Mass Hunter Qualitative (MHQ) and Mass Profiler Professional (MPP)

10. Submit the data matrix to statistical analysis using the online Metaboanalyst program (www.metaboanalyst.ca).
11. Apply univariate data analysis based on T-test and fold change (FC) analysis to detect significant molecular features (p-value (FDR) cutoff: 0.05) with a defined absolute value of change (FC > 1.5) between control and treated cell samples.
12. Apply partial least squares-discriminant analysis (PLS-DA). MS entities showing high VIP score in the PLS-DA can be considered as relevant metabolites, as they allow explaining the differences between both groups of samples.
13. Select significant MS features from the resulting volcano plot and the multivariate PLS-DA analysis for further identification.
14. Generate molecular formulas using the Molecular Formula Generator algorithm within the MHQ software to enhance the metabolite database search and mass accuracy calculation (<5 ppm).
15. Identify the selected metabolites on the basis of the information provided by MS data (accurate mass, isotopic distribution), positive match of the MS/MS fragmentation pattern with online MS databases (e.g., HMDB, Metlin), and information reported in the literature.

3.6 Metabolomic Data Interpretation

1. After the post-acquisition data processing described in Subheading 3.5, a total of 99 and 88 MS features are obtained in positive and negative ionization mode, respectively, in the studied extracts.

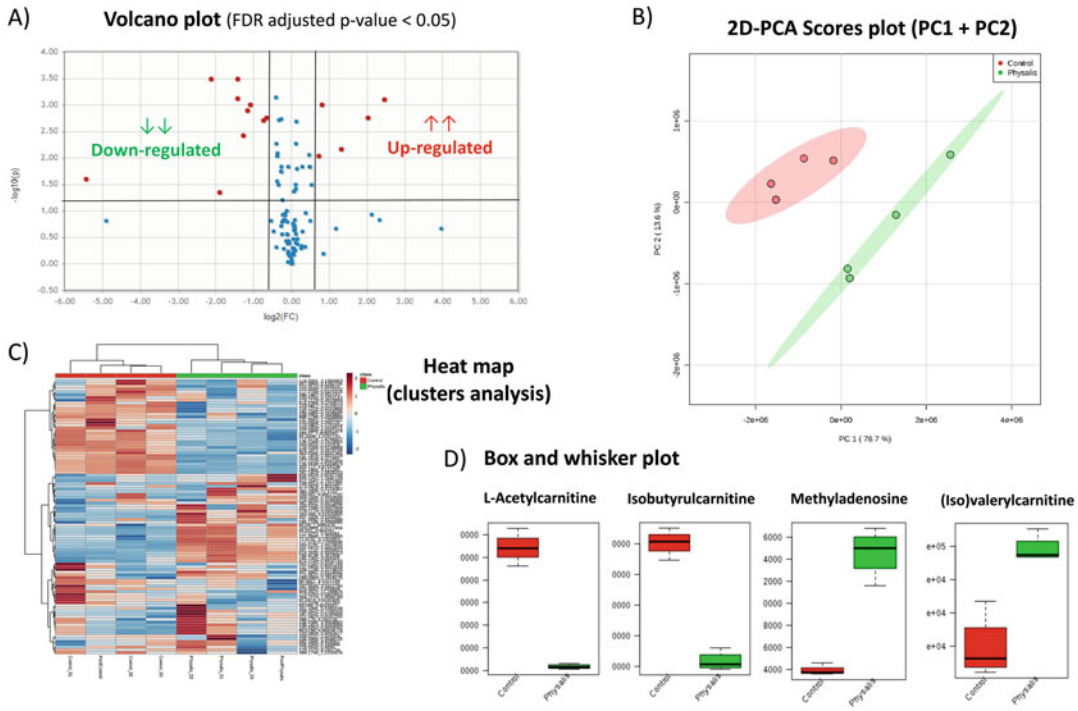

Fig. 2 Metabolomic data interpretation. (**a**) Volcano plot, (**b**) 2D-PCA score plot, (**c**) heat map with clusters analysis, and (**d**) box and whisker plot

2. The 2D-PCA score plot (Fig. 2b) classifies the samples in two groups, suggesting a clear difference between treated and untreated HT-29 cell samples in terms of expressed metabolites.
3. T-tests reveal a total of 38 differentially expressed metabolites (p value FDR adjusted < 0.05) detected in ESI+ mode, whereas operating in ESI (−), a total of 25 compounds showed significant ($p < 0.05$) variation upon treatment with the calyx PLE-extract, most of them exhibiting FC values of > 1.5 ($-0.585 > \text{Log2 FC} > 0.585$).
4. Among the differentially expressed metabolites are carnitine derivatives such as acetyl-, propionyl-, (iso)valeryl-, (iso)-butyryl-, and hydroxybutyryl-L-carnitine; amino acids L-phenylalanine and L-tyrosine; adenosine monophosphate, and purine nucleosides such as inosine, xanthine, and guanosine monophosphate (*see* Fig. 2c).
5. Despite showing lower FC ratios, a group of three metabolites (xanthine, L-leucine, L-valine) can also be considered, as they were shown to be significantly altered compounds with large weight in the discriminant analysis, according to the high

variable importance in projection value (VIP > 1) in the first component of the PLS-DA analysis (*see* **Note 3**).

6. Quantitative pathway and enrichment analysis are carried in the Metaboanalyst platform, integrating high-quality KEGG and SMPDB databases as the backend knowledge bases, in an attempt to tentatively identify the most relevant metabolic pathways in which the differentially expressed metabolites are involved.
7. Beta-oxidation of fatty acids, aminoacyl-tRNA biosynthesis, branched-chain (Val, Leu) and aromatic (Phe, Tyr) amino acids, glutathione, pyrimidine, and purine metabolic pathways are the most affected routes.
8. Metabolomics data also indicate altered cellular redox homeostasis, as deducted from the shift in the cellular GSH-to-GSSG redox balance in favor of the oxidized species. The dramatic upregulated levels of GSSG in treated cells constitute an important signal of cell impairment by oxidative stress. The deregulated levels of essential branched-chain (Val, Leu) and aromatic (Phe, Tyr) amino acids, along with the downregulated levels of short-chain acyl-carnitines, were identified as early indicators of dysfunctions in metabolic pathways involved in protein degradation and turnover, glycogen synthesis, and energy metabolism.

4 Notes

1. Although methods based on LC-ESI-MS are frequently run in positive ionization mode for metabolomics studies, the metabolome coverage is improved in this work by operating in both positive and negative ESI modes. It is worthy of highlighting that only 4 compounds out of 24 target metabolites were detected by both ESI+ and ESI- modes, evidencing the advantage of applying both complementary acquisition modes.
2. The reference compound solution for internal mass calibration of the Q/TOF mass spectrometer containing 5 μM purine ($[C_5H_5N_4]^+$ at 121.050873 Da) and 2.5 μM HP-0921, hex-S-5akis(1H,1H,3H-tetra- fluoropropoxy) phosphazine ($[C_{18}H_{19}O_6N_3P_3F_{24}]^+$ at 922.009798 Da) in acetonitrile-water (95:5, v/v) and was also from Agilent.
3. Despite the low abundance of some MS features, small changes in the levels of some metabolites might have remarkable effects on the metabolism, as discussed by Ballesteros-Vivas et al. [10].

Acknowledgments

This work was supported by AGL2017-89417-R project (Ministry of Science and Innovation, Spain). G.A.-R. and A.V. would like to acknowledge the Ministry of Science and Innovation (Ministry of Science and Innovation, Spain) for their "Juan de la Cierva-Incorporación" postdoctoral grants IJC2019-041482-I and IJC2018-037560-I, respectively.

References

1. Zeki ÖC, Eylem CC, Reçber T et al (2020) Integration of GC–MS and LC–MS for untargeted metabolomics profiling. J Pharm Biomed Anal 190:113509. https://doi.org/10.1016/J.JPBA.2020.113509
2. Cajka T, Fiehn O (2015) Toward merging untargeted and targeted methods in mass spectrometry-based metabolomics and lipidomics. Anal Chem 88:524–545. https://doi.org/10.1021/ACS.ANALCHEM.5B04491
3. Halket JM, Waterman D, Przyborowska AM et al (2005) Chemical derivatization and mass spectral libraries in metabolic profiling by GC/MS and LC/MS/MS. J Exp Bot 56: 219–243. https://doi.org/10.1093/JXB/ERI069
4. Xia J, Sinelnikov IV, Han B, Wishart DS (2015) MetaboAnalyst 3.0—making metabolomics more meaningful. Nucleic Acids Res 43: W251–W257. https://doi.org/10.1093/NAR/GKV380
5. Kind T, Tsugawa H, Cajka T et al (2018) Identification of small molecules using accurate mass MS/MS search. Mass Spectrom Rev 37: 513–532. https://doi.org/10.1002/MAS.21535
6. Blaženović I, Kind T, Sa MR et al (2019) Structure annotation of all mass spectra in untargeted metabolomics. Anal Chem 91: 2155–2162. https://doi.org/10.1021/ACS.ANALCHEM.8B04698
7. Wang S, Kind T, Tantillo DJ, Fiehn O (2020) Predicting in silico electron ionization mass spectra using quantum chemistry. J Chem 12(1):1–11. https://doi.org/10.1186/S13321-020-00470-3
8. Pang Z, Chong J, Zhou G et al (2021) MetaboAnalyst 5.0: narrowing the gap between raw spectra and functional insights. Nucleic Acids Res 49:W388–W396. https://doi.org/10.1093/NAR/GKAB382
9. Barupal DK, Fiehn O (2017) Chemical Similarity Enrichment Analysis (ChemRICH) as alternative to biochemical pathway mapping for metabolomic datasets. Sci Rep 7(1):1–11. https://doi.org/10.1038/s41598-017-15231-w
10. Ballesteros-Vivas D, Alvarez-Rivera G, León C et al (2019) Anti-proliferative bioactivity against HT-29 colon cancer cells of a withanolides-rich extract from golden berry (Physalis peruviana L.) calyx investigated by Foodomics. J Funct Foods 63:103567. https://doi.org/10.1016/j.jff.2019.103567
11. Ballesteros-Vivas D, Alvarez-Rivera G, León C et al (2020) Foodomics evaluation of the antiproliferative potential of Passiflora mollissima seeds. Food Res Int 130:108938. https://doi.org/10.1016/j.foodres.2019.108938
12. Ballesteros-Vivas D, Alvarez-Rivera G, Ibánez E et al (2019) Integrated strategy for the extraction and profiling of bioactive metabolites from Passiflora mollissima seeds combining pressurized-liquid extraction and gas/liquid chromatography–high resolution mass spectrometry. J Chromatogr A 1595: 144–157. https://doi.org/10.1016/J.CHROMA.2019.02.031
13. Ballesteros-Vivas D, Álvarez-Rivera G, del Pilar S-CA et al (2019) A multi-analytical platform based on pressurized-liquid extraction, in vitro assays and liquid chromatography/gas chromatography coupled to high resolution mass spectrometry for food by-products valorisation. Part 1: Withanolides-rich fractions from golde. J Chromatogr A 1584:155–164. https://doi.org/10.1016/j.chroma.2018.11.055
14. Ibáñez C, Simó C, Palazoglu M, Cifuentes A (2017) GC-MS based metabolomics of colon cancer cells using different extraction solvents. Anal Chim Acta 986:48–56. https://doi.org/10.1016/J.ACA.2017.07.019

Chapter 6

Untargeted Metabolomics by Liquid Chromatography–Mass Spectrometry in Biomedical Research

Caridad Díaz and Carmen González-Olmedo

Abstract

Metabolomics, alone or in combination with other omics sciences, has shown great relevance in a large number of investigations in different branches of biomedicine, often providing novel discoveries and helping to expand the knowledge. Metabolomics analyses are carried out using different techniques, but in this chapter, we focus on liquid chromatography coupled to high-resolution mass spectrometry. The designated methodology consists of an untargeted approach for the analysis of plasma samples. The use of this method, with a reverse-phase column and electrospray ionization in positive mode, covers the detection of a broad range of metabolites, mainly of nonpolar and of intermediate polarity. This chapter also reviews the mass fragmentation spectra for the identification of bile acids, acylcarnitines, and glycerophospholipids.

Key words Untargeted metabolomics, Mass spectrometry, Reverse phase, Plasma, Lipids, Bile acids, Acylcarnitines, Glycerophospolipids

1 Introduction

Metabolomics is a powerful tool for finding new biomarkers in the biomedicine area. During the last decades, research in this discipline has shown extensive development, with a significant amount of data generated where the use of metabolomics has soared in popularity for the diagnosis or prognosis of different diseases, such as cancer s [1], diabetes [2], autism [3], or Alzheimer [4]. Other applications have emerged, including drug metabolism [5], study of mechanisms of action [6], toxicology [7], and comparative analysis between mutants that explain pathogenesis and/or virulence [8], among others, which highlight the importance of this -omics science. Metabolomics analyses are carried out using different techniques, mainly mass spectrometry (MS) and nuclear magnetic resonance (NMR). MS shows certain advantages compared to NMR, such as higher sensitivity and selectivity, since it covers a wide range of metabolites in just one analysis. Furthermore, the

Raúl González-Domínguez (ed.), *Mass Spectrometry for Metabolomics*,
Methods in Molecular Biology, vol. 2571, https://doi.org/10.1007/978-1-0716-2699-3_6,
© The Author(s), under exclusive license to Springer Science+Business Media, LLC, part of Springer Nature 2023

combination of MS with other techniques increases the detection capacity and its versatility in the analysis, being habitual to working with separation techniques such as gas chromatography, liquid chromatography (LC), and capillary electrophoresis. On the other hand, NMR is the tool par excellence in structural identification. It requires a small quantity of samples and provides a large amount of information, although the analysis is usually very complex. The combination of these two techniques is of important relevance since NMR and high-resolution (HR) MS offer a perfect solution for the discovery of new biomarkers. However, in this chapter, we will focus on the LC coupled to HRMS. The use of this analytical technique allowed us to work with different types of biological samples, including blood, plasma, serum, urine, tissues, feces, saliva, sweat, exhaled breath, gastrointestinal fluid, breastmilk, cerebrospinal fluid, or even placentas, also with a simple processing in most cases. A significant number of biomedical research that used this type of analysis have led to results in some cases of great scientific relevance [9–11]. Metabolomics is divided into two well established methodologies: (1) untargeted—it was the preferred approach in the first metabolomics analyses and covers a wide range of metabolites; and (2) targeted—totally necessary to validate new metabolic signatures for their translation to the clinic, and some authors believe that it will replace the untargeted analyses in the next years [12]. In this regard, our opinion supports untargeted metabolomics as a useful method that has a lot to explain in some scientific fields—as a practical example, in exposomics [13], the changes that occur in different metabolic pathways during the exposition to pollutants. This process may influence the development of important pathological states that are not always described [14]. Although we cannot ignore that the compound identification task may suppose the bottleneck of this approach, structural information available in databases is increasing, as well as the possibility of acquiring new analytical standards.

Currently, metabolomics is an expanding research area that has re-emerged as a powerhouse in personalized medicine [15, 16], a major driver of advances in the microbiome field [17, 18], and ultimately, as a great tool that is able to provide an enormous amount of data to be fed into the system and applied in the age of artificial intelligence [19–21]. In this chapter, we described an untargeted metabolomics method using LC coupled to HRMS in plasma samples that covers the detection of a broad range of metabolites, and which can be applied in different areas of the biomedicine field. Herein, although the method is described for plasma, it could be used in different biological matrices, considering that, depending on the sample type, the conditions of the sample preparation will have to be modified.

2 Materials

2.1 Metabolite Extraction and Protein Precipitation

1. Blood collecting tubes containing $EDTAK_2$ anticoagulant.
2. Vials for HPLC.
3. Centrifuge (Sorvall ST26R Thermo Scientific).
4. Acetonitrile.
5. Internal standard stock solution (ISS): 1 mg/mL L-tryptophan-3d, roxithromycin, and [^{13}C6] hydroxydiclofenac in dimethylsulfoxide.
6. Shaker (multi-tube vortexer, Fisher Scientific).
7. Evaporator (GeneVac HT-8 evaporator: Savant, Holbrook, NY, USA).
8. Reconstitution solution: 200 ppb ISs (L-tryptophan-3d, roxithromycin, and [$^{13}C_6$] hydroxydiclofenac) in acetonitrile/water (50/50, v/v).
9. Phosphate-buffered saline.

2.2 Liquid Chromatography and Mass Spectrometry

1. Agilent 1290 liquid chromatograph.
2. Atlantis T3 HPLC column (C18: 2.1 mm × 150 mm, 3 μm) and Atlantis T3 guard column (2.1 mm × 2 mm, 3 μm) from Waters (Waters Corporation, Milford, MA, USA).
3. Mobile phase A (MPA): water/acetonitrile (90/10, v/v)] plus 0.1% formic acid.
4. Mobile phase B (MPB): acetonitrile/water (90/10, v/v) plus 0.1% formic acid.
5. QTOF5600 mass spectrometer (SCIEX).
6. CTC-PAL HTS-xt injector (Leap Technologies).

2.3 Software for Data Set Curation and Statistical Analysis

1. Peak View software (version 1.1.2, AB SCIEX, Concord, ON).
2. Marker View software (version 1.2.1, AB SCIEX, Concord, ON).
3. SPPS Version 25.0 (Statistical analysis).
4. Metaboanalyst 5.0 (Statistical analysis).

2.4 Databases for the Identification of Metabolites

1. CEU Mass mediator: http://ceumass.eps.uspceu.es/.
2. Metlin: https://metlin.scripps.edu/landing_page.php?pgcontent=mainPage.
3. Lipidmaps: https://www.lipidmaps.org/.
4. Human metabolome database: https://hmdb.ca/.
5. Chemspider: http://www.chemspider.com/.
6. FooDB: https://foodb.ca/about.

7. Pubchem; https://pubchem.ncbi.nlm.nih.gov/.
8. Massbank: https://massbank.eu/MassBank/.
9. NIST 2012 mass spectral library.

3 Methods

3.1 Sample Preparation

1. All blood samples must be collected at fasting and using homogeneous sampling conditions.
2. Blood samples are collected in tubes containing $EDTAK_2$ as an anticoagulant, and plasma is obtained through a centrifugation process at 2000 rpm for 10 min (*see* **Note 1**).
3. For metabolite extraction, an aliquot of 100 μL of plasma is mixed with 900 μL of cold acetonitrile (*see* **Note 2**).
4. Shake all samples for 1 min at 2500 rpm.
5. Centrifuge at 13,700 rpm for 15 min at 4 °C.
6. A volume of 950 μL of supernatant is evaporated using a vacuum evaporator without heating.
7. All the samples are reconstituted in 250 μL of acetonitrile/water (50/50, v/v) containing ISs (*see* **Note 3**).
8. Quality controls (QCs) are prepared by pooling equal aliquots of each sample included in our analysis (*see* **Note 4**).
9. Blank samples are prepared with 100 μL of PBS using the same procedure as for plasma samples (*see* **Note 5**).

3.2 Liquid Chromatography

1. Equilibrate and purge system before analysis (*see* **Note 6**).
2. The chromatographic run lasts 20 min.
3. The gradient elution is performed as follows: $t = 0$–0.5 min, 1% MPB; $t = 11$ min, 99% MPB; $t = 15.50$ min, 99% MPB; $t = 15.60$ min, 1% MPB.
4. The flow rate is set at 0.3 mL/min.
5. The column temperature is maintained at 25 °C.
6. The injected sample volume is 5 μL.
7. Samples are injected randomly.
8. At the end of the analysis, inject QC samples diluted (*see* **Note 7**).

3.3 Mass Spectrometry

1. This method is described to use a positive electrospray ionization mode and a mass range from 50 to 1400 Da.
2. The source parameters used in the analysis are set as follows: declustering potential (DP): 100 eV, ion source gas 1 and ion source gas 2 = 50 psi, temperature at 550 °C, ion spray voltage floating = 5000 eV, and curtain gas = 45 psi.

3. The analysis is carried out in TOF mode with an information-dependent acquisition (IDA) experiment with full MS scan and information-dependent acquisition triggers MS/MS fragmentation results.
4. The accumulation time for MS full scan is 250 ms for scanning and 99 ms for each IDA experiment.
5. Parameters for this experiment are collision energy = 30 eV and collision energy spread = 15 eV.
6. Fragmentation analyses monitor a maximum of 8 candidate ions per cycle with an intensity threshold above 100 cps.
7. Dynamic background subtraction is switched on.
8. Automatic calibration is recommended every 6 injections during analysis (*see* **Note 8**).

3.4 Data Set Creation

1. First, the retention time and m/z variability of the three ISs must be assessed. This information is used to evaluate the analytical performance and correct for variability in retention time.
2. Data mining is accomplished with an automated algorithm in the retention time range between 1.20 and 15.50 min. The rest of the chromatographic run time is excluded because it corresponds to the death volume and the wash solvent of the column.
3. Collection parameters are set as follows: mass window 20.0 ppm, retention time window 0.12 min, and minimum intensity 100 counts in all cases.
4. Only monoisotopic peaks are considered to decrease mass redundancy (*see* **Note 9**).
5. Eliminate from the analysis all signals that are not more than threefold different in the QC samples compared to the blank samples.
6. Variables with unacceptable reproducibility (RSD > 30%) in QC samples are rejected from the data matrix.

3.5 Analytical Validation

1. Precision is evaluated by using random injections of the QC during the sample sequence.
2. Apply principal component analysis (PCA) to corroborate the clustering of the QC samples.

3.6 Statistical Analysis (Univariate)

1. Apply the Kolmogorov–Smirnov test to evaluate data normality.
2. If the data is not normal, it can be normalized, scaled, and transformed by different strategies.

3. When data is normal, use parametric tests and calculate whether the variances are equal using Levene's test. Thus, if p-value > 0.05 (equal variances), use the Student's t-test (two groups) or ANOVA (more than two groups), and if p-value < 0.05 (unequal variances), use the Welch test.
4. When the data is not normal, apply Wilcoxon's nonparametric test for two related groups, Mann–Whitney for two independent groups, Kruskal–Wallis for more than two independent groups, or Friedman for more than two related groups.

3.7 Identification of Metabolites

1. PeakView software is used to determine the molecular formula according to the exact mass and isotope pattern. Elemental formula approximation is completed from both single HRMS and MS/MS accurate mass spectra. Tentative formulas reflect the differences between calculated and measured m/z values for both parent and fragment ions.
2. Identification of molecular components is achieved through comparative searches of available mass spectra using CEU mediator as a tool for searching metabolites in several databases, such as Metlin, Human Metabolome Data Base, Lipid Maps, Chemspider, Pubchem, NIST 2012 mass spectral library, and Mass bank, and/or in the literature search. In addition, there are tools available to interpretate fragmentation patterns, such as Metfrag.
3. Using parent compound information and the experimental conditions, ionization behavior and/or retention time in the chromatographic condition can facilitate the structural assignation. To consider a structural identification with confidence, it is necessary to obtain its fragmentation spectrum or acquire a commercial standard and evaluate it under the same analysis conditions. Figure 1 shows the total ion chromatogram (TIC) obtained from a plasma sample using this methodology.
4. Phenylalanine is retained at 2.3 min, acquiring a proton (+H) with an m/z of 165.086 Da, while tryptophan is retained at 3.5 min and m/z of 188.070 Da, from the capture of a proton and the loss of NH_3 (+H-NH_3) [22]. Another compound derived from tryptophan, such as tryptophan betaine, can be detected at 3.86 min with m/z 247.144 Da.
5. Nutritional components derived from xanthine can be found near these retention times, as the theophylline ingredient of tea at 3.4 min and at m/z 181.071 Da (+H), and caffeine present in coffee at 4.1 min and m/z 195.087 Da (+H).
6. Other families of compounds, produced by the degradation of hemoglobin in red blood cells, appear in the middle of the chromatogram, between 8 and 9 min. Specifically, biliverdin and bilirubin can be detected by acquiring a proton at m/z 583.255 and 585.271 Da, respectively, and by capturing the

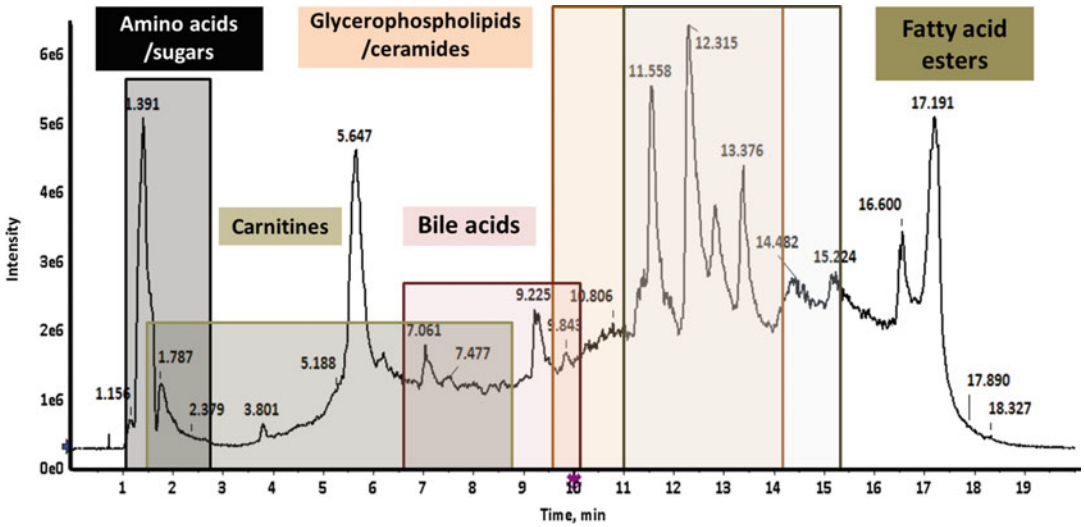

Fig. 1 Total ion chromatogram obtained in reverse-phase chromatography and positive electrospray ionization mode. Each family of molecules is represented in the corresponding area where they elute

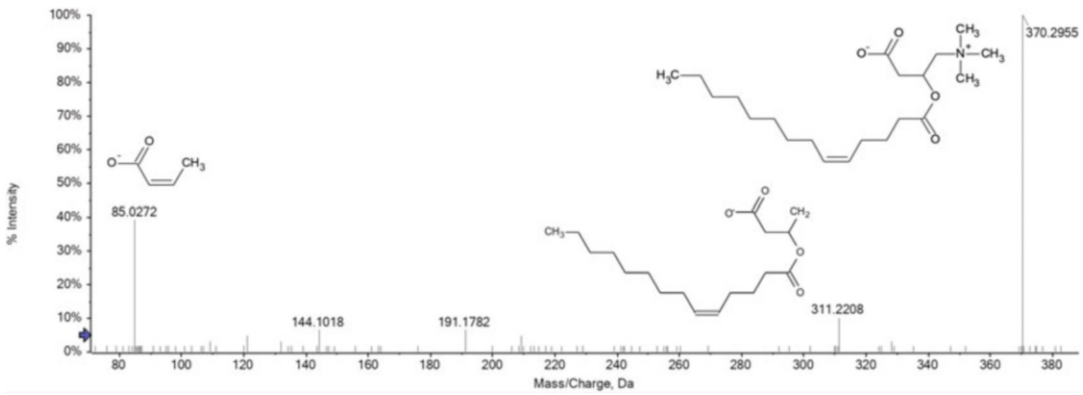

Fig. 2 Fragmentation spectrum of tetradecenoylcarnitine with positive electrospray ionization mode under weak acidic conditions at collision energy = 30 Ev and collision energy spread = 15 Ev. MS/MS interpretation of [M + H] adduct revealed the specific fragment ions at m/z 85.027

sodium cation at m/z 605.630 and 607.650 Da, respectively. Both show a characteristic fragment at m/z 297.123 and 299.147 Da, respectively [23].

7. Eicosanoids show a positive ionization capacity under weak acidic conditions. These compounds elute at the end of the chromatogram.
8. Acylcarnitines can be detected practically throughout the entire chromatogram from 1.52 min, where acetylcarnitine appears, until 9.88 min, where octadecanoylcarnitine is found. To help identifying this metabolite family, it must be considered the appearance of a very intense fragment at m/z 85 Da [24] (Fig. 2 and Table 1).

Table 1
Summary of acylcarnitines found with reverse-phase chromatography and positive electrospray ionization mode analysis

m/z	RT (min)	Molecular formula	Tentative identification	Adduct
204.123	1.52	$C_9H_{17}NO_4$	Acetylcarnitine	[M + H]
232.154	2.97	$C_{11}H_{21}NO_4$	Butirylcarnitine	[M + H]
288.217	6.24	$C_{15}H_{29}NO_4$	Octenoylcarnitine	[M + H]
316.248	7.18	$C_{17}H_{33}NO_4$	Decanoylcarnitine	[M + H]
342.263	7.38	$C_{19}H_{35}NO_4$	Dodecenoylcarnitine	[M + H]
344.279	8.09	$C_{19}H_{37}NO_4$	Lauroylcarnitine	[M + H]
358.295	8.11	$C_{20}H_{39}NO_4$	Tridecanoylcarnitine	[M + H]
424.343	9.38	$C_{25}H_{46}NO_4$	Linoleoylcarnitine	[M + H]
428.373	9.88	$C_{25}H_{49}NO_4$	Octadecanoylcarnitine	[M + H]

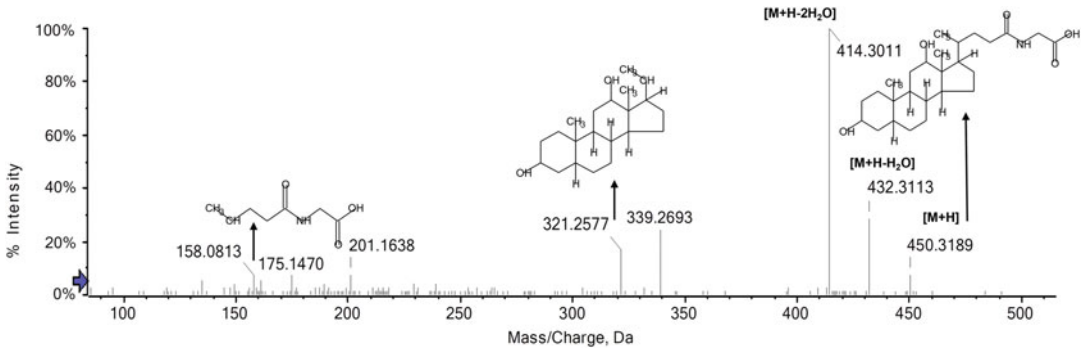

Fig. 3 Fragmentation spectrum of glycodeoxycholic acid with positive electrospray ionization mode under weak acidic conditions at collision energy = 30 eV and collision energy spread = 15 Ev

9. Bile acids can be detected with the loss of one or more molecules of water, but also forming adducts acquiring a proton and/or sodium (*see* **Note 10**), so it is very typical to find several parent ions with different m/z values (Fig. 3 and Table 2).
10. Phosphatidylcholines (PCs) and related lipids are protonated under acidic conditions, acquiring a positive charge on the quaternary nitrogen in the positive ionization mode [25]. The entire product ion spectrum is dominated by the m/z 184 Da ion, representing a phosphocholine fragment. Other characteristic fragments are 104, 125, and 85 Da ions [2], while the structurally informative ions corresponding to the aliphatic chain are in lower abundance [25, 26] (*see* **Note 11**).

Table 2
Summary of bile acids found with reverse-phase chromatography and positive electrospray ionization mode analysis

m/z	RT (min)	Molecular formula	Tentative identification	Adduct
355.263 373.274	8.94 8.96	$C_{20}H_{40}O_5$	Cholic acid	[M + H-3H_2O] [M + H-2H_2O]
466.317	7.62	$C_{26}H_{43}NO_6$	Glycocholic acid	[M + H]
448.305	8.45	$C_{26}H_{43}NO_6$	Glycohyocholic acid	.[M + H]
414.301 432.311 450.323 472.303	8.94 8.96 8.94 8.94	$C_{26}H_{43}NO_5$	(1) Glycodeoxycholic acid	[M + H-2H_2O] [M + H-H_2O] [M + H] [M + Na]
414.301 432.311 450.323 472.303	9.22 9.22 9.22 9.23	$C_{26}H_{43}NO_5$	(2) Glycodeoxycholic acid	[M + H-2H_2O] [M + H-H_2O] [M + H] [M + Na]

11. Alkaline adducts (+Na, +K) of phospholipids can be easily formed, mainly depending on the availability of these in the solution used. The most common adduct formed is [M + Na]. Some specific fragment ions of PCs with these adducts correspond to the loss of trimethylamine $[M + Na/K\text{-}59]^+$, the loss of phosphocholine [M + Na/K-183], and the loss of choline phosphate $[M + Na/K\text{-}(Na/K + 182]^+$. In addition, two typical fragments at m/z 104 and 147 Da are the results of the phosphocholine head group, usually appearing in PC-Na adducts [25–27].
12. Phosphatidylethanolamines (PEs) can positively ionize under acidic conditions acquiring a proton. The positive charge site is on the quaternary nitrogen. Characteristic fragments appear at m/z [M + H-141], corresponding to the neutral loss of the phosphoethanolamine moiety [28, 29] (Fig. 4a). Adducts with Na can also be formed, accompanied by a typical fragment at m/z 195.003 Da (Fig. 4b). Table 3 lists some of the forms of glycerophospholipids that have been identified with this method. Phosphatidylserines show a characteristic daughter ion at [M + H-185], derived from the loss of the phosphate head group.
13. Sphingomyelins (SMs) form the molecular ion [M + H] and show characteristic fragment at m/z 184 $[C_5H_{15}O_4NP]^+$. Other daughter ions could appear, such as the loss of H_2O [30].

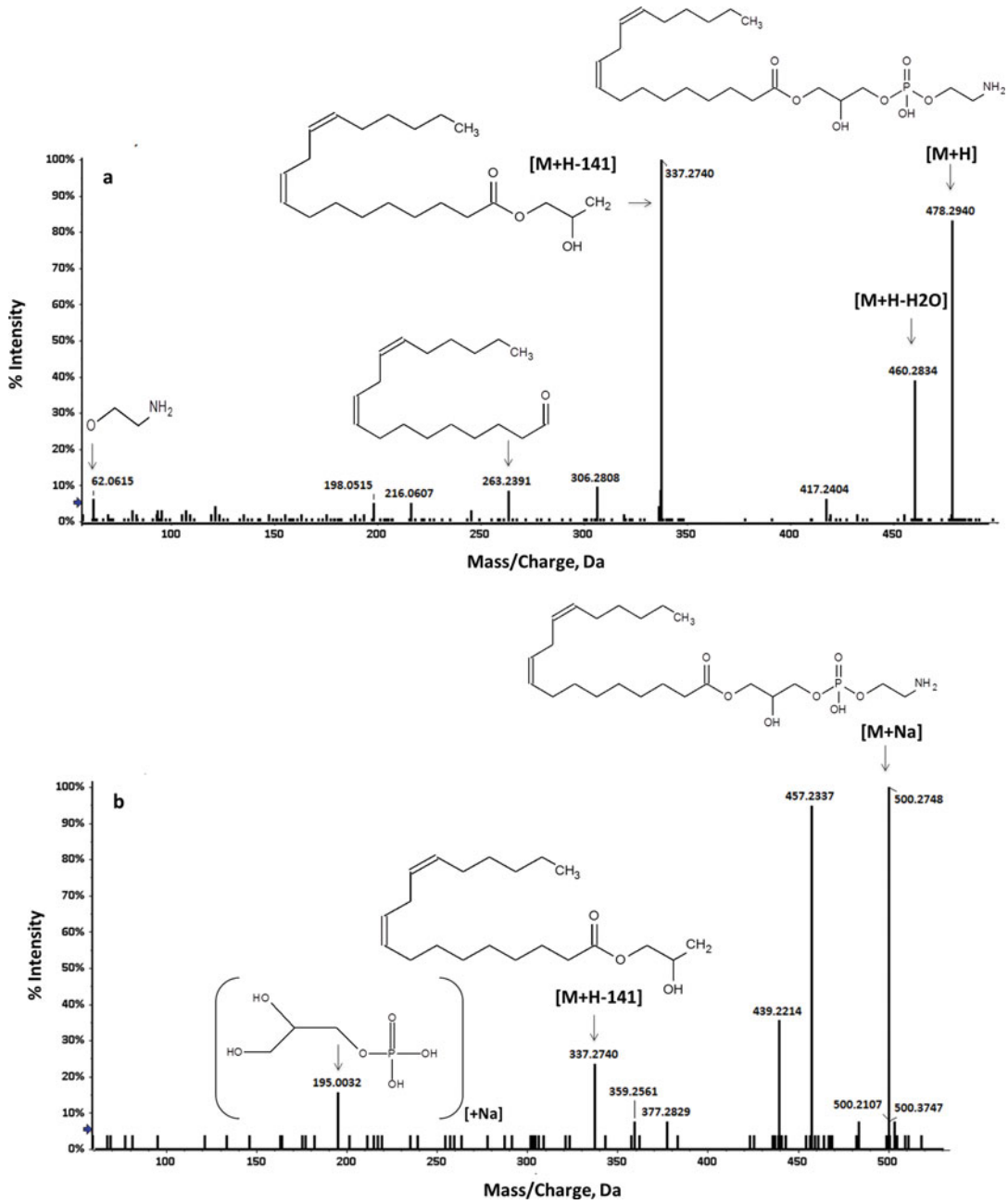

Fig. 4 Fragmentation spectrum of lysophosphatidylethanolamine (18:2(9Z,12Z)/0:0) with positive electrospray ionization mode under weak acidic conditions at collision energy = 30 eV and collision energy spread = 15 Ev. MS/MS interpretation of [M + H] (**a**) and [M + Na] (**b**) adducts revealed specific fragment ions at m/z 195.003

Table 3
Summary of glycerophospholipids found with reverse-phase chromatography and positive electrospray ionization mode analysis

Family	m/z	RT (min)	Molecular formula	Tentative identification	Adduct
Lysophosphatidylcholine	496.338	9.81	$C_{24}H_{50}NO_7P$	PC(O-14:0/2:0)	[M + H]
	991.672	9.83	$C_{24}H_{50}NO_7P$		[2 M + H]
	496.338	11.27	$C_{24}H_{50}NO_7P$	LysoPC(16:0/0:0)	[M + H]
	518.323	11.29			[M + Na]
	508.341	11.01	$C_{25}H_{50}NO_7P$	LysoPC(17:1/0:0)	[M + H]
	510.355	12.19	$C_{25}H_{52}NO_7P$	LysoPC(17:0/0:0)	[M + H]
	1039.67	10.83	$C_{26}H_{50}NO_7P$	LysoPC(18:2/0:0)	[2 M + H]
	522.345	11.61	$C_{26}H_{52}NO_7P$	LysoPC(18:1/0:0)	[M + H]
	546.356	11.4	$C_{28}H_{52}NO_7P$	LysoPC(20:3/0:0)	[M + H]
	566.321	10.74	$C_{28}H_{50}NO_7P$	LysoPC(20:4/0:0)	[M + Na]
	542.323	10.03	$C_{28}H_{48}NO_7P$	LysoPC(20:5/0:0)	[M + H]
	564.305	10.02			[M + Na]
	572.369	11.87	$C_{30}H_{54}NO_7P$	LysoPC(22:4/0:0)	[M + H]
	568.342	10.68	$C_{30}H_{50}NO_7P$	LysoPC(22:6/0:0)	[M + H]
	590.322	10.69			[M + Na]
Lysophosphatidylethanolamine	454.293	11.19	$C_{21}H_{44}NO_7P$	LysoPE(18:2/0:0)	[M + H]
	478.294	10.64	$C_{21}H_{44}NO_7P$	LysoPE(18:2/0:0)	[M + H]
	500.274	10.62			[M + Na]
	480.306	11.76	$C_{23}H_{46}NO_7P$	LysoPE(18:1/0:0)	[M + H]
	502.287	11.78			[M + Na]
	502.292	10.5	$C_{25}H_{44}NO_7P$	LysoPE(20:4/0:0)	[M + H]
	504.31	11.29	$C_{25}H_{46}NO_7P$	LysoPE(20:3/0:0)	[M + H]
	526.291	10.62	$C_{27}H_{44}NO_7P$	LysoPE(22:6/0:0)	[M + H]

4 Notes

1. Plasma samples showing a red color should not be considered in the analysis due to the rupture of red blood cells.
2. To favor the extraction of metabolites, store the acetonitrile at −20 °C before use.
3. The retention time change of these ISs during the analysis will serve to correct that variability in the data set.
4. Make sure when preparing the QC samples to take the necessary volume to avoid the vial running out of enough liquid to inject.
5. The use of blank samples is very important to check for impurities from the solvents or the extraction process and to check carryover contamination from analytes by comparing blank samples with QC samples.
6. At the beginning of the analysis, equilibrate column using 10 QC samples.

7. At the end of the analysis, inject QC samples diluted 1/3, 1/9, and 1/27 to evaluate the possible saturation in some signals that can lead to mistakes in the interpretation of results.
8. Consider a molecular formula as good when the mass error is ≤5 ppm.
9. It is recommended to select only monoisotopic peaks to improve the selection of the true molecular feature and to facilitate the identification.
10. Different signals can correspond to the same molecule due to the formation of several adducts; take this into account when evaluating the candidates.
11. It should be noted that in this class of metabolites the retention times increase with the number of carbons in the chains and decrease with the number of double bonds.

Acknowledgments

We would like to thank Fundación MEDINA for funding and Dr. José Pérez del Palacio for technical support.

References

1. Martín-Blázquez A, Díaz C, González-Flores E et al (2019) Untargeted LC-HRMS-based metabolomics to identify novel biomarkers of metastatic colorectal cancer. Sci Rep 9(1):1–9. https://doi.org/10.1038/s41598-019-55952-8
2. García-Fontana B, Morales-Santana S, Díaz Navarro C et al (2016) Metabolomic profile related to cardiovascular disease in patients with type 2 diabetes mellitus: a pilot study. Talanta 148:135–143. https://doi.org/10.1016/j.talanta.2015.10.070
3. Wang H, Liang S, Wang M et al (2016) Potential serum biomarkers from a metabolomics study of autism. J Psychiatry Neurosci 41(1): 27–37. https://doi.org/10.1503/jpn.140009
4. Niedzwiecki MM, Walker DI, Howell JC et al (2020) High-resolution metabolomic profiling of Alzheimer's disease in plasma. Ann Clin Transl Neurol 7(1):36–45. https://doi.org/10.1002/acn3.50956
5. Dubuis S, Ortmayr K, Zampieri M (2018) A framework for large-scale metabolome drug profiling links coenzyme a metabolism to the toxicity of anti-cancer drug dichloroacetate. Commun Biol 1(1):1–11. https://doi.org/10.1038/s42003-018-0111-x
6. Zampieri M, Szappanos B, Buchieri MV et al (2018) High-throughput metabolomic analysis predicts mode of action of uncharacterized antimicrobial compounds. Sci Transl Med 10(429). https://doi.org/10.1126/scitranslmed.aal3973
7. Steuer AE, Brockbals L, Kraemer T (2019, May) Metabolomic strategies in biomarker research-new approach for indirect identification of drug consumption and sample manipulation in clinical and forensic toxicology. Front Chem 7. https://doi.org/10.3389/fchem.2019.00319
8. Díaz C, Pérez Del Palacio J, Valero-Guillén PL et al (2019) Comparative metabolomics between mycobacterium tuberculosis and the MTBVAC vaccine candidate. ACS Infect Dis 5(8):1317–1326. https://doi.org/10.1021/acsinfecdis.9b00008
9. His M, Viallon V, Dossus L et al (2019) Prospective analysis of circulating metabolites and breast cancer in EPIC. BMC Med 17(1). https://doi.org/10.1186/s12916-019-1408-4
10. Isa F, Collins S, Lee MH et al (2018) Mass spectrometric identification of urinary biomarkers of pulmonary tuberculosis. EBioMedicine

31:157–165. https://doi.org/10.1016/j.ebiom.2018.04.014

11. Wurtz P, Havulinna AS, Soininen P et al (2015) Metabolite profiling and cardiovascular event risk: a prospective study of 3 population-based cohorts. Circulation 131(9):774–785. https://doi.org/10.1161/CIRCULATIONAHA.114.013116
12. Pinu FR, Goldansaz SA, Jaine J (2019) Translational metabolomics: current challenges and future opportunities. Meta 9(6). https://doi.org/10.3390/metabo9060108
13. Walker DI, Valvi D, Rothman N, Lan Q, Miller GW, Jones DP (2019) The metabolome: a key measure for exposome research in epidemiology. Curr Epidemiol Rep 6:93–103. https://doi.org/10.1007/s40471-019-00187-4
14. Gangemi S, Gofita E, Costa C et al (2016) Occupational and environmental exposure to pesticides and cytokine pathways in chronic diseases (review). Int J Mol Med 38(4): 1012–1020. https://doi.org/10.3892/ijmm.2016.2728
15. Beger RD, Dunn W, Schmidt MA, et al. (2016) Metabolomics enables precision medicine: "A White Paper, Community Perspective." Metabolomics 12(10). https://doi.org/10.1007/s11306-016-1094-6
16. Castelli FA, Rosati G, Moguet C et al (2021) Metabolomics for personalized medicine: the input of analytical chemistry from biomarker discovery to point-of-care tests. Anal Bioanal Chem. https://doi.org/10.1007/s00216-021-03586-z
17. Chevre R, Silvestre-Roig C, Soehnlein O (2018) Nutritional modulation of innate immunity: the fat–bile–gut connection. Trends Endocrinol Metab 29(10):686–698. https://doi.org/10.1016/j.tem.2018.08.002
18. Romick-Rosendale LE, Haslam DB, Lane A et al (2018) Antibiotic exposure and reduced short chain fatty acid production after hematopoietic stem cell transplant. Biol Blood Marrow Transplant 24(12):2418–2424. https://doi.org/10.1016/j.bbmt.2018.07.030
19. Sen P, Lamichhane S, Mathema VB et al (2021) Deep learning meets metabolomics: a methodological perspective. Brief Bioinform 22(2):1531–1542. https://doi.org/10.1093/bib/bbaa204
20. Liebal UW, Phan ANT, Sudhakar M, Raman K, Blank LM (2020) Machine learning applications for mass spectrometry-based metabolomics. Meta 10(6):1–23. https://doi.org/10.3390/metabo10060243
21. Pomyen Y, Wanichthanarak K, Poungsombat P, Fahrmann J, Grapov D, Khoomrung S (2020) Deep metabolome: applications of deep learning in metabolomics. Comput Struct Biotechnol J 18:2818–2825. https://doi.org/10.1016/j.csbj.2020.09.033
22. Díaz-Beltrán L, González-Olmedo C, Luque-Caro N et al (2021) Human plasma metabolomics for biomarker discovery: targeting the molecular subtypes in breast cancer. Cancers (Basel) 13(1):1–18. https://doi.org/10.3390/cancers13010147
23. Quinn KD, Nguyen NQT, Wach MM, Wood TD (2012) Tandem mass spectrometry of Bilin tetrapyrroles by electrospray ionization and collision-induced dissociation. Rapid Commun Mass Spectrom 26(16):1767–1775. https://doi.org/10.1002/rcm.6287
24. Giesbertz P, Ecker J, Haag A, Spanier B, Daniel H (2015) An LC-MS/MS method to quantify acylcarnitine species including isomeric and odd-numbered forms in plasma and tissues. J Lipid Res 56(10):2029–2039. https://doi.org/10.1194/jlr.D061721
25. Hsu FF, Turk J (2003) Electrospray ionization/tandem quadrupole mass spectrometric studies on phosphatidylcholines: the fragmentation processes. J Am Soc Mass Spectrom 14(4):352–363. https://doi.org/10.1016/S1044-0305(03)00064-3
26. Pulfer M, Murphy RC (2003) Electrospray mass spectrometry of phospholipids. Mass Spectrom Rev 22(5):332–364. https://doi.org/10.1002/mas.10061
27. Hsu FF, Turk J, Thukkani AK, Messner MC, Wildsmith KR, Ford DA (2003) Characterization of alkylacyl, alk-1-enylacyl and lyso subclasses of glycerophosphocholine by tandem quadrupole mass spectrometry with electrospray ionization. J Mass Spectrom 38(7): 752–763. https://doi.org/10.1002/jms.491
28. Brouwers JFHM, Vernooij EAAM, Tielens AGM, Van Golde LMG (1999) Rapid separation and identification of phosphatidylethanolamine molecular species. J Lipid Res 40(1): 164–169. https://doi.org/10.1016/s0022-2275(20)33352-6
29. Kerwin JL, Tuininga AR, Ericsson LH (1994) Identification of molecular species of glycerophospholipids and sphingomyelin using electrospray mass spectrometry. J Lipid Res 35(6): 1102–1114. https://doi.org/10.1016/s0022-2275(20)40106-3
30. Pi J, Wu X, Feng Y (2016) Fragmentation patterns of five types of phospholipids by ultra-high-performance liquid chromatography electrospray ionization quadrupole time-of-flight tandem mass spectrometry. Anal Methods 8(6):1319–1332. https://doi.org/10.1039/C5AY00776C

Chapter 7

Untargeted Metabolomics for Disease-Specific Signatures

Constantina Chalikiopoulou, José Carlos Gómez-Tamayo, and Theodora Katsila

Abstract

Human diseases account for complex traits that usually exhibit markedly diverse clinical manifestations coming from a series of pathogenic processes that shape heterogeneous phenotypes. Considering that correlation does not imply causation as well as population differences and/or inter-individual variability, disease-specific signatures are becoming critical for biomarker discovery. Untargeted metabolomics is deemed to be a powerful approach to delineate molecular pathways of prime interest. Metabotypes capture the interplay of genomics and environmental influences per se. Untargeted metabolomics share the charm of being not only hypothesis-driven but also hypothesis-generating. Notwithstanding, the applicability of untargeted metabolomics toward clinically relevant outcomes depend on wet- and dry-lab procedures in the context of elegant study designs with clear rationale. As ideal may be far from feasible, herein we provide recommendations to combat sample mishandling that adversely affect data outcomes and if so, deal with imbalanced datasets toward data integrity.

Key words Untargeted metabolomics, Molecular signatures, Disease profiling

1 Introduction

Each disease phenotype shares specific molecular pathways that can be traced via the metabolome [1], while changes in cell bioenergetics unravel inter-individual variability as well as population differences [2]. The accumulation of cell biomass is associated with dramatically increased bioenergetic and biosynthetic demand [3, 4]. Metabolic reprogramming, once thought of as an epiphenomenon, currently, relates to cell processes, often in response to extracellular fate-decisive signals. Omics-based molecular signatures of clinical phenotypes present great opportunities, despite current challenges and hurdles that we need to overcome. To name but a few, solid experimental design, careful data

Authors "Constantina Chalikiopoulou" and "José Carlos Gómez-Tamayo" have equally contributed to this chapter.

Raúl González-Domínguez (ed.), *Mass Spectrometry for Metabolomics*,
Methods in Molecular Biology, vol. 2571, https://doi.org/10.1007/978-1-0716-2699-3_7,
© The Author(s), under exclusive license to Springer Science+Business Media, LLC, part of Springer Nature 2023

procurement, avoidance of over-fitting, validation on independent datasets, and data integration (usually of multiple datasets and data types) need to be considered [5–7]. To meet clinical needs, molecular signatures need to outperform reproducibility, sensitivity, specificity, and robustness [1, 2, 8–11].

To implement such a strategy in full capacity, both wet- and dry-lab approaches should be elegantly designed and performed; sample collection, extraction, and preparation for mass spectrometry analysis and next, sample processing and analysis. At each step, several factors and parameters need to be accounted for (personnel, instrumentation, location, reagents), also depending on the scale of the study (pilot, large-scale, multi-center) as well as its design (patient-cohort, umbrella, basket). Limited sample sizes and imbalanced datasets may hinder data reliability, despite stringent experimental protocols improving reproducibility. For this, the study design should be meticulous to fully address the topic in question and at the same time, account for the limitations present. In particular, when it comes to clinical data, inter-individual variability is a major issue together with anthropometric parameters (such as age and sex) and disease phenotypes (such as disease staging). To fill the current knowledge gap against confounding factors and empower clinical credibility, metabolomics is often employed.

Herein we discuss untargeted metabolomics that offers collective and real-time molecular measurements (metabotypes) for disease-specific profiles and provide our recommendations to translate information growth into knowledge growth accounting for biases. Our recommendations are independent of the technological platform used.

2 Materials

Prepare all solutions using HPLC-grade or LC–MS-grade reagents. Prepare and store all reagents at room temperature (unless indicated otherwise). All waste disposal regulations should be followed when disposing waste materials. All relevant ethical regulations and guidelines for the collection and use of human blood should be followed.

2.1 Solutions and Materials for Blood Collection and Processing

1. 10-mL lavender-top K2 EDTA BD Vacutainer® blood collection tubes.
2. Red cap BD Vacutainer® blood collection tubes.
3. Extraction buffer: 50:30:20 methanol:acetonitrile:water (v:v:v) (*see* **Note 1**).
4. 1 mM zinc acetate (or zinc chloride).

2.2 Liquid Chromatography–Mass Spectrometry

1. Mobile phase A: acetonitrile.
2. Mobile phase B: 20 mM ammonium carbonate, 0.1% ammonium hydroxide (v:v).
3. UPLC 1290 system (Agilent Technologies, Santa Clara, CA USA) coupled to a TripleTOF 5600+ mass spectrometer (SCIEX) equipped with SWATH acquisition, SelexION technology, and an electrospray ionization source (ESI).
4. pZIC-HILIC column (5 μm, 2.1 mm × 150 mm).

3 Methods

Experimental design (wet-lab approach) is deemed as a critical first step in data analysis. This is the fundamental aspect to be considered when there is a question of which approach to be implemented for signal extraction from a given dataset.

3.1 Sample Collection and Storage

Proper sample collection (plasma or serum) is crucial (*see* **Note 2**).

1. Patient must be seated at least 5 min before the draw.
2. Blood can be drawn from median, cubital, basilic, or cephalic veins, but never from a port. For this purpose, apply a tourniquet two inches above the antecubical fossa or above the area to be drawn with enough pressure to provide adequate vein visibility. Select the site for venipuncture.
3. Clean the forearm of the patient with antiseptic wipe in a circular motion beginning at the insertion site. Allow the antiseptic to dry.
4. For plasma collection, pre-chill 10-mL lavender-top K2 EDTA BD Vacutainer® venous blood collection tubes on wet ice for at least 5 min. For serum, use red cap collection tubes (another option may be red cap/black ring serum clot activator tubes).
5. Draw blood into the corresponding tube by aspirating until the tube is completely filled (*see* **Note 3**). Carefully remove the tubes when full without dislodging the needle.
6. Immediately after allowing the tube to completely fill, slowly and gently invert the tube 8–10 times.
7. For plasma, immediately insert the tube into wet ice. For serum, place tubes vertically on the bench at room temperature and let peripheral blood samples to clot naturally (>30 min).
8. Centrifuge the tubes at 2000 *g* for 15 min in a refrigerated centrifuge (4 °C). EDTA tubes can also be directly frozen before centrifugation (*see* Subheading 3.3 about how to process further frozen blood samples).

9. Transfer the supernatant plasma/serum (using a sterile disposable 10-mL pipette) to 3-mL cryovials (sterile cryovials with silicone washer seal and external threads; self-standing; certified DNase-free, RNase-free, DNA-free, and Pyrogen-free), taking care to not disturb the buffy coat.
10. Store plasma and serum samples at 80 °C until their further analysis. Sample processing and freezing must be completed within 90 min of collection.

3.2 Sample Processing (Plasma/Serum)

1. Allow plasma/serum samples to thaw on wet ice at 4 °C for 30–60 min (*see* **Note 4**).
2. Aliquot 50 μL of plasma/serum samples into a labeled 1.5-mL microcentrifuge tube.
3. Add 750 μL of the ice-cold extraction buffer to each sample (*see* **Note 5**). Shake the mixture at 4 °C for 15 min.
4. Centrifuge samples at 14,000 *g* for 10 min and collect supernatants.
5. Dry down (lyophilize) each sample in a centrifugal vacuum evaporator for 18 h. Apply no heating during the drying process.
6. Store samples at 4 °C no more than 12 weeks (*see* **Note 6**).

3.3 Sample Processing (Frozen Blood)

1. Thaw samples at 4 °C overnight.
2. Label 15-mL Falcon tubes and 1.5-mL or 2.0-mL low adherence Eppendorf tubes.
3. Centrifuge the thawed EDTA tubes at 2000 *g* for 5 min at room temperature (*see* **Notes 3** and **7**). Collect the supernatant in new 15-mL Falcon tubes. Avoid the interface (buffy coat = platelets, RBC debris, WBCs).
4. To remove hemoglobin, add 1.0 mL of zinc acetate/chloride solution (*see* **Note 8**). Mix and allow standing for 10 min. Incubate for 5 min at 37 °C for hemoglobin-specific precipitation (*see* **Note 9**). Centrifuge at 2500 *g* for 10 min at room temperature. Collect the supernatant in new 15-mL Falcon tubes (*see* **Note 10**).
5. Mix gently and thoroughly with a pipette and collect 1.0 mL in 1.5-mL or 2.0-mL low adherence Eppendorf tubes. Centrifuge at 500 *g* for 10 min at 4 °C to remove cells (*see* **Note 11**). Collect the supernatant in new 1.5-mL or 2.0-mL low adherence Eppendorf tubes.
6. Proceed with protein precipitation as described above (*see* Subheading 3.2).

3.4 Sample Randomization

1. Mark sample types (*see* **Note 12**) and replicates with their corresponding "sample IDs."
2. Check that the sample list consists of the unknown samples (samples of biological interest), the blank samples, and the quality control (QC) samples. Prepare QCs as a pool of samples; add 2-μL small aliquots of each biological sample. Mix thoroughly.
3. Generate a random sample list using MATLAB, Excel, or any available random sample number/sequence generator. For example, in Excel, add a new column within the spreadsheet and name it Random_number. In the first cell underneath the heading row, type "= RAND()". Press "Enter," and a random number will appear in the cell. Copy and paste the first cell into the other cells in this column. Once each row contains a random number, sort the records by Random_number column.
4. Check the outcome so that no replicates or samples of the same test-group are adjacent.

3.5 Data Acquisition

1. Perform tuning and mass calibration for U/HPLC–MS at the start of each analytical block (rather than each analytical batch) to avoid large-step changes in the data coming from the analytical batches, which can be detrimental in data preprocessing steps.
2. Chromatography: Perform UPLC separation on metabolite extracts using a pZIC-HILIC column (5 μm, 2.1 mm × 150 mm) operating at 45 °C by directly injecting 10 μL of samples. The flow rate should be 0.2 mL/min by applying a gradient, in both positive and negative ion mode, from 20% to 80% B in 15 min.
3. Mass spectrometry: Acquire data over a mass range of 75–1000 m/z. Automated calibration is performed using an external calibrant delivery system (CDS), which infuses APCI positive or negative calibration solution every five samples. To monitor the instrument performance over time, analyze QC samples every five samples. Perform a TOF MS survey scan experiment with an IDA set to monitor the eight most intense candidate ions (accumulation time of 150 msec in TOF-MS and 50 msec in the IDA experiment), with a collision energy of 35 eV and a collision energy spread of 10 eV, a declustering potential of 80 V, a source temperature (TEM) of 500 °C, and IonSpray Voltage Floating (ISVF) of 5500 V (positive polarity) or 4500 V (negative polarity) in high-sensitivity mode.
4. Archive raw data for future use.
5. Perform instrument maintenance at the end of each analytical batch. This involves mass spectrometer ion source and liquid chromatography column cleaning.

3.6 Data Preprocessing and Analysis

Herein, we provide recommendations and describe a data (pre)-processing and analysis protocol when imbalanced datasets are at hand using Python 3.6.0, following up on the use of any commercially available or open-source software (*see* **Note 13**):

1. Process raw data to provide a data matrix using MetaboKit (https://github.com/MetaboKit).
2. Remove metabolic features detected in <50% of QC samples.
3. Calculate the relative standard deviation (RSD) for all metabolic features in QC samples. Remove metabolic features with RSD > 20% (*see* **Note 14**).
4. Remove metabolic features detected in <33% of samples.
5. To assess if noise is decreased, plot metabolite distributions for the test-groups studied; determine the number of samples in which each metabolic feature is detected and plot their count distributions to check (Fig. 1).
6. Impute missing values with a small value (i.e., 1).
7. Perform data centering and unit variance scaling.
8. Perform log2fold calculation and select only features with a log2fold ≥ 2 for subsequent enrichment analysis.

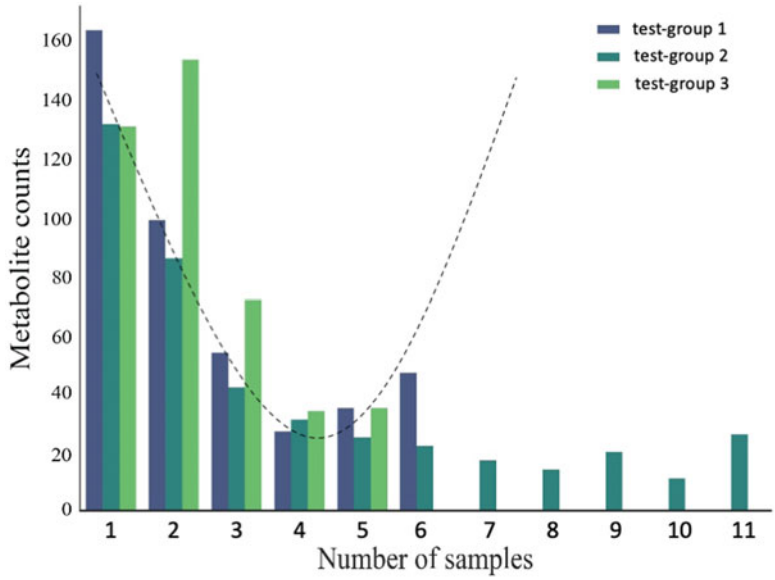

Fig. 1 Count distributions for metabolic features in the test-groups studied. The trend (dotted line) shows that noise is decreased after the removal of all metabolic features detected in <33% of samples. The plot was generated using the python seaborn plotting library (distplot function) and depicts the count distribution of metabolic features over the samples in the test-groups studied. Test-groups studied: (1) dark blue; (2) dark green; (3) light green

9. Merge positive- and negative-ion mode datasets for each test-group studied in a single data table by taking into account the maximum absolute log2fold values when a metabolite is found in both datasets.
10. Perform principal component analysis (PCA), partial least squares (PLS) analysis, and PLS-discriminant analysis (PLS-DA) to analyze patterns in the group variance and extract individual metabolite significance (statistical significance of variables in projection, VIP) for all the test-groups studied (*see* **Note 15**).
11. Determine PLS-VIP and PLS-DA-VIP values and include those at the master table.
12. Perform enrichment analysis using Metaboanalyst v5.0 [14] employing pathway-associated metabolite sets (SMPDB). After submitting the list of metabolite names, obtain a graph with enriched pathways sorted by their significance as well as a table containing the data serving the graph (Fig. 2).
13. Use the KEGG pathway identifier to extract metabolites for each pathway using the KEGG REST api. Use this extracted list to match the metabolites identified and generate two new tables. The first one contains the significant metabolites identified together with the significant pathways they are involved in (metabolites vs. pathways). The second table contains the significant pathways and their corresponding metabolites identified (pathways vs. metabolites).

4 Notes

1. To achieve an ice-cold state for the organic solvent, store at −20 °C overnight before use.
2. Special care should be taken when various sample protocols are employed or limited sample sizes or imbalanced datasets are in place. In these cases, it is a good idea to acquire an independent set of samples (at least for data imputation) to be processed together with the majority of the data or used for independent validation of the analysis pipeline. Sample confounding or unbalanced design can also undermine the power of predictive models [12]. Identifying, blocking, and tracking of all technical factors as well as sample randomization may ameliorate or even correct unavoidable biases to preserve the biological information. Furthermore, the reliability of metabolomic datasets is affected by sample quality, handling, preparation, and storage. Different extraction protocols can result in different metabolic patterns, also shaping data interpretation.

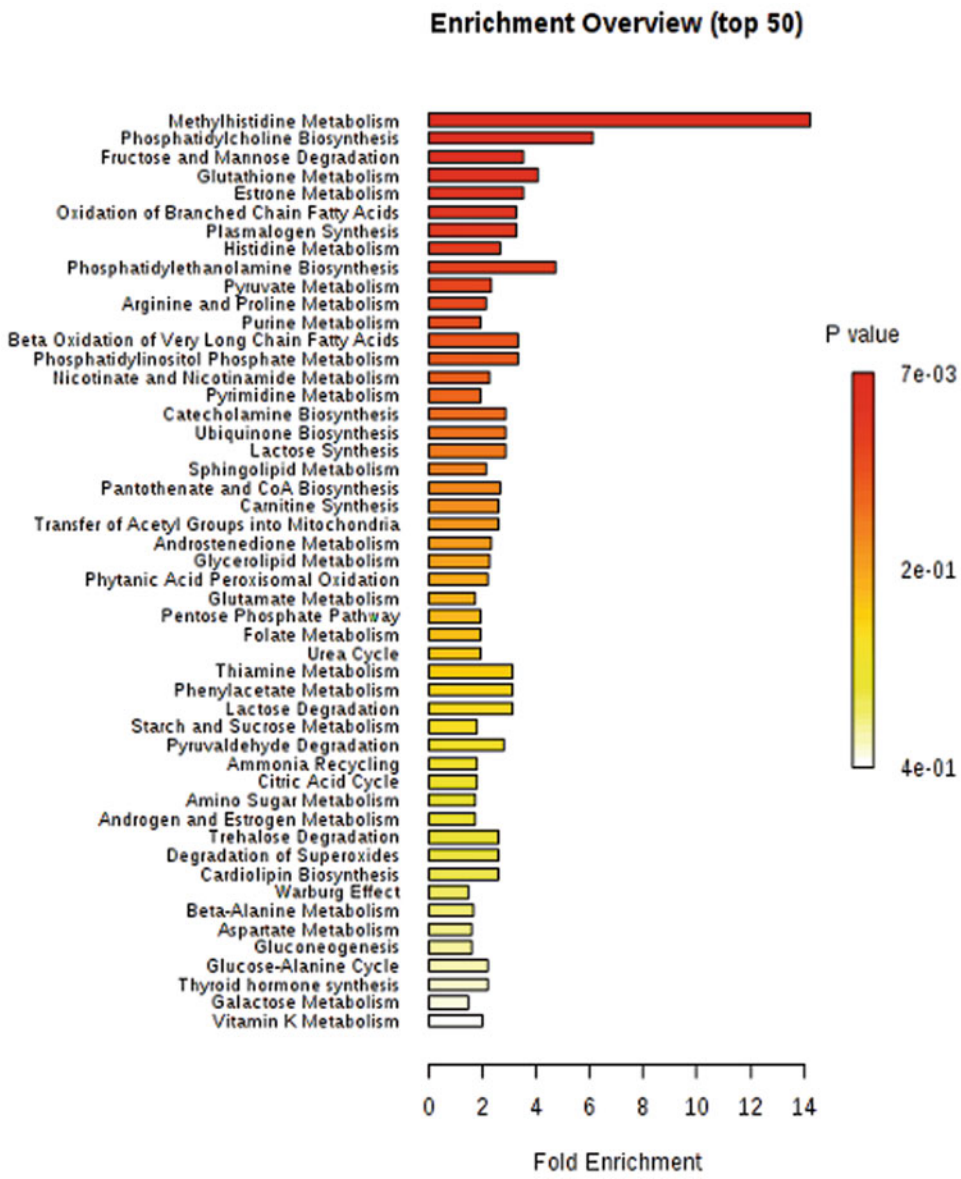

Fig. 2 Metabolite set enrichment analysis (MSEA). This analysis was carried out by MetaboAnalyst v5.0. Enriched pathways were ranked by significance (see color scale) calculated by the MetPa method implemented in MetaboAnalyst v5.0

3. In particular, for human blood samples, different outcomes have been reported even with placing samples on ice or using EDTA-covered storage tubes, instead of refrozen samples [10]. Tube additives (heparin- or EDTA-tubes, citrate-containing tubes, sodium fluoride-containing tubes) used are not always interchangeable options as several studies report that they may affect mass spectrometry-derived metabolomic

profiles (e.g., signal to noise ratio). Freeze and thaw cycles should be avoided. Freeze–thaw cycles of the biospecimens have to be recorded in their data sheets.

4. Enzyme activity is decreased at reduced temperatures, although no enzyme inhibition is achieved until temperatures below −56 °C are set. Sample preparation steps until protein precipitation are best to be carried out at temperatures just above the freezing point of water (sample thawing/preparation on ice is recommended).
5. Protein precipitation results in the separation of the precipitate and the metabolite-containing supernatant. A choice of several organic solvents is available, and each shares advantages and disadvantages. Various solvent ratios (v/v) should also be tested to select the one for best recovery. It is best to add the organic solvent gradually at equal sub-volumes (e.g., 3*400 μL instead of 1200 μl methanol aliquots).
6. A single deproteinization step is described herein as it allows for fast sample preparation (n = 100 samples/daily), thereby removing the variability that might be introduced from preparing samples separately. This accounts for potential errors that can be observed in replicate sample preparation procedures.
7. Please note that platelets get activated at a lower temperature.
8. Hemoglobin types A, B, and D may mask low abundance metabolites.
9. Room temperature also works, but 37 °C gives the best precipitation. Do not use too much zinc that will also precipitate other proteins (mainly HSA).
10. Alternatively, use a Ficoll or Histopaqe gradient for the isolation of the plasma and/or buffy coat. It may be hard because of the hemolysis. Collect the supernatant in new 15-mL Falcon tubes. Another action plan would be to use a Nucleopore filter (or similar) of a size less than 3 microns. Pass the sample through (some types may need centrifugation). Plasma will be collected in the bottom chamber. Blood cells will be captured onto the filter. The "buffy coat" may be transferred to a tube for DNA extraction, etc. Protein precipitation (using organic solvents) may also be of help.
11. Alternatively, centrifuge at 1000 *g* for 15 min to remove cell membranes. Collect supernatant in new 1.5-mL or 2.0-mL Eppendorf tubes (low adherence).
12. Sample randomization is a two-step process; sample preparation order is randomized from sample picking and then re-randomized from sample analysis. This combats no systematic biases. Internal standard may always be a tough decision to make, in particular for untargeted metabolomics. An ideal

choice may not be possible in cases where hundreds to thousands of metabolites are detected. A mix of internal standards may be applied, allowing for each internal standard to compensate for multiple metabolites, a metabolite class, or a subset of metabolites. Quality control (QC) samples are considered identical to the biological samples studied (i.e., unknowns), sharing their metabolic and sample matrix composition. QC samples allow metabolic features with excessive drift in retention time, signal, or accurate mass to be removed prior to data analysis. Thus, QC samples serve as a measure of repeatability within an analytical run. QC samples may be commercially available, consisting of multiple biological samples not present in the study. We recommend using a pooled QC sample that comes from all or a subset of the subject population so that no or minimal metabolic information is lost. Of note, pooled QC samples are anticipated to be closest to the composition of the biological samples upon study. Column mix chromatography test-samples may also be included in the sample list.

13. Similar to the case of a small, focused study, large-scale studies are broken down into smaller ones that allow for reproducibility and robustness [11, 13]. Data coming from multiple analytical blocks as well as positive and negative ionization modes are processed and integrated into a single data set. Experimental design (dry-lab approach) may seem more simplistic in nature in a way that the resulted dataset is the one that shapes what tests should or may be performed as well as their limitations. Upon raw data acquisition, preprocessing refers to data formats appropriate for data analysis leading to matrices of m/z versus index or retention time versus intensity or ion current. Several software options are currently available for data processing resulting in the construction of a 3D matrix (chromatographic peak with related index or retention time, electron-induced fragmentation mass spectrum, and/or accurate mass) versus response versus sample ID [11]. Such software may be open-source or commercially available. Typically, data alignment is secured (i.e., alignment of drift for retention time and accurate mass) and each chromatographic peak is identified in each sample based on the same parameters.
14. Set strict criteria that may sacrifice quantity over quality to obtain a list of reliable features. The latter are used in further statistical analyses for group characterization (i.e., enrichment analysis, partial least squares analysis).
15. The partial least squares discriminant analysis model is able to separate the test-groups studied by their metabolite profiles.

References

1. Wang R, Li B, Lam SM, Shui G (2020) Integration of lipidomics and metabolomics for in-depth understanding of cellular mechanism and disease progression. J Genet Genomics 47(2):69–83
2. Sun YV, Hu Y-J (2016) Integrative analysis of multi-omics data for discovery and functional studies of complex human diseases. Adv Genet 93:147–190
3. Johnson CH, Ivanisevic J, Siuzdak G (2016) Metabolomics: beyond biomarkers and towards mechanisms. Nat Rev Mol Cell Biol 17(7):451–459
4. Lambrecht BN, Hammad H (2017) The immunology of the allergy epidemic and the hygiene hypothesis. Nat Immunol 18(10): 1076–1083
5. Schadt EE (2009) Molecular networks as sensors and drivers of common human diseases. Nature 461(7261):218–223
6. Hood L, Heath JR, Phelps ME et al (2004) Systems biology and new technologies enable predictive and preventative medicine. Science 306(5696):640–643
7. Sung J, Wang Y, Chandrasekaran S et al (2012) Molecular signatures from omics data: from chaos to consensus. Biotechnol J 7(8): 946–957
8. Kumar A, Misra BB (2019) Challenges and opportunities in cancer metabolomics. Proteomics 19(21–22):1900042
9. Luan H, Wang X, Cai Z (2019) Mass spectrometry-based metabolomics: targeting the crosstalk between gut microbiota and brain in neurodegenerative disorders. Mass Spectrom Rev 38(1):22–33
10. Fang Z-Z, Gonzalez FJ (2014) LC–MS-based metabolomics: an update. Arch Toxicol 88(8): 1491–1502
11. Dunn WB, Broadhurst D, Begley P et al (2011) Procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid chromatography coupled to mass spectrometry. Nat Protoc 6(7):1060–1083
12. Zhao Q, Adeli E, Pohl KM (2020) Training confounder-free deep learning models for medical applications. Nat Commun 11(1):1–9
13. Zelena E, Dunn WB, Broadhurst D et al (2009) Development of a robust and repeatable UPLC– MS method for the long-term metabolomic study of human serum. Anal Chem 81(4):1357–1364
14. Pang Z, Chong J, Zhou G et al (2021) MetaboAnalyst 5.0: narrowing the gap between raw spectra and functional insights. Nucleic Acids Res 49(W1):W388–W396

Chapter 8

Metabolomics Analysis of Blood, Urine, and Saliva Samples Based on Capillary Electrophoresis–Mass Spectrometry

Masahiro Sugimoto and Yumi Aizawa

Abstract

Capillary electrophoresis–mass spectrometry (CE–MS) is an ideal method for analyzing various metabolites in biological samples. CE–MS can simultaneously identify and quantify hundreds of charged metabolites using only two acquisition methods for positively and negatively charged metabolites. Furthermore, CE–MS is commonly used for analyzing biological samples to understand the pathology of diseases at the metabolic level and biofluid samples, such as blood and urine, to explore biomarkers. Here, we introduce a protocol that delineates the handling of clinical samples to ensure that the CE–MS analysis yields reproducible quantified data. We have focused on sample collection, storage, processing, and measurement. Although the implementation of rigorous standard operating protocols is preferred for enhancing the quality of the samples, various limitations in an actual clinical setting make it difficult to adhere to strict rules. Therefore, the effect of each process on the quantified metabolites needs to be evaluated to design a protocol with acceptable tolerances. Furthermore, quality controls and assessments to handle clinical samples are introduced.

Key words Capillary electrophoresis–mass spectrometry, Clinical samples, Biofluid

1 Introduction

Among the various analytical approaches in metabolomics, mass spectrometry (MS) is a widely used tool for identifying and quantifying metabolites in biological samples. CE–MS can detect charged metabolites; therefore, metabolites, such as nucleotides and amino acids, of various primary pathways, including glycolysis, tricarboxylic acid (TCA) cycle, pentose phosphate pathway (PPP), urea cycle, polyamine pathway, and one-carbon metabolism pathway, can be analyzed [1, 2].

Thus, CE–MS is extensively used to analyze clinical samples. Drastic differences in the metabolomic profiles of colon cancer tissue samples and adjunct normal tissue samples were observed even in the early stages of cancer, such as adenoma [3]. Different metabolomic profiles were observed in tumor tissues collected from

Raúl González-Domínguez (ed.), *Mass Spectrometry for Metabolomics*,
Methods in Molecular Biology, vol. 2571, https://doi.org/10.1007/978-1-0716-2699-3_8,
© The Author(s), under exclusive license to Springer Science+Business Media, LLC, part of Springer Nature 2023

patients with gastric cancer and colon cancer [4]. Therefore, cancer type-specific metabolomic profiles in biofluids have been vigorously analyzed to realize a minimally invasive diagnosis. Blood samples, urine, and saliva samples have also been analyzed to identify biomarkers for various cancers, for example, urine samples to detect colon cancer [5], and saliva samples to explore the biomarkers of oral cancer, breast cancer, and pancreatic cancer [6–10].

To obtain reproducible quantitative values of the metabolite biomarkers, a standard operating protocol (SOP) integrated with quality control (QC) for each process is required [10, 11]. For example, the duration between sample collection and the last meal of the patient can affect the concentration of salivary oral biomarkers; also, the sample collection method, stimulated or unstimulated saliva, yielded differences in the metabolomic profiles [12, 13]. Furthermore, the storage period and temperature alter the salivary profile [14]; therefore, they must be controlled. For blood samples, the use of plasma and serum showed high similarities in their metabolomic profiles, while the serum metabolite profiles, with clotting time having the largest impact on serum metabolomic profiles, require a rigorous standardization [15]. To identify and validate biomarkers in clinical research, strict rules are preferable to eliminate unexpected biases. However, it is often difficult to follow strict rules in an actual clinical setting, and designing realistic SOPs within acceptable quality tolerance of the quantified data is warranted.

In this study, a protocol to handle clinical samples for CE–MS-based metabolomics is introduced.

2 Materials

2.1 Blood Sample

1. Centrifuge the samples at 1600 × *g* for 10 min to obtain the serum after clotting for 30 min at room temperature (15–25 °C), or plasma after anticoagulation for 30 min at 4 °C (*see* **Note 2**).
2. Transfer the serum or plasma to a polypropylene tube and store them at −80 °C.

2.2 Urine Sample

1. Collect spot urine samples at a fixed time (*see* **Note 3**).
2. Transfer the samples to a polypropylene tube and store them at −80 °C.

2.3 Saliva Sample

1. Ensure the saliva providers are not allowed to consume any food except water intake after 9:00 p.m. on the previous day (*see* **Note 4**).

2. Ensure that the subjects brush their teeth without toothpaste on the day of saliva collection and refrain from drinking water, smoking, tooth-brushing, and intense exercise 1 h before saliva collection.
3. Direct the subjects to gargle with water just before saliva collection.
4. Collect the samples to a polypropylene at a fixed time (8:00 to 11:00 a.m.; *see* **Note 4**).
5. After collection, immediately store the samples at −80 °C (*see* **Note 5**).

3 Methods

3.1 Blood Samples

1. Thaw the frozen serum or plasma samples at 4 °C for approximately 1.5 h.
2. Mix 40 μL aliquots of serum and plasma samples with 360 μL of methanol containing internal standards (IS_1: 20 μM methionine sulfone, 2-morpholinoethanesulfonic acid, and D-camphor-10-sulfonic acid).
3. Mix the samples and add 400 μL of chloroform and 160 μL of Milli-Q water (Millipore, Billerica, MA, USA), followed by centrifugation at 10,000 × *g* for 3 min at 4 °C.
4. Filter the upper aqueous layer of each sample through a centrifugal filter tube with a 5-kDa cutoff (Millipore) at 9100 × *g* for 3 h at 20 °C to remove any large molecules.
5. Centrifuge the filtrates using a vacuum concentrator at 40 °C to concentrate them and resuspend in 40 μL of Milli-Q water containing internal standards (IS_2: 200 μM 3-aminopyrrolidine and trimesate).
6. Transfer all the samples to a vial for analysis by CE–MS (Fig. 1).

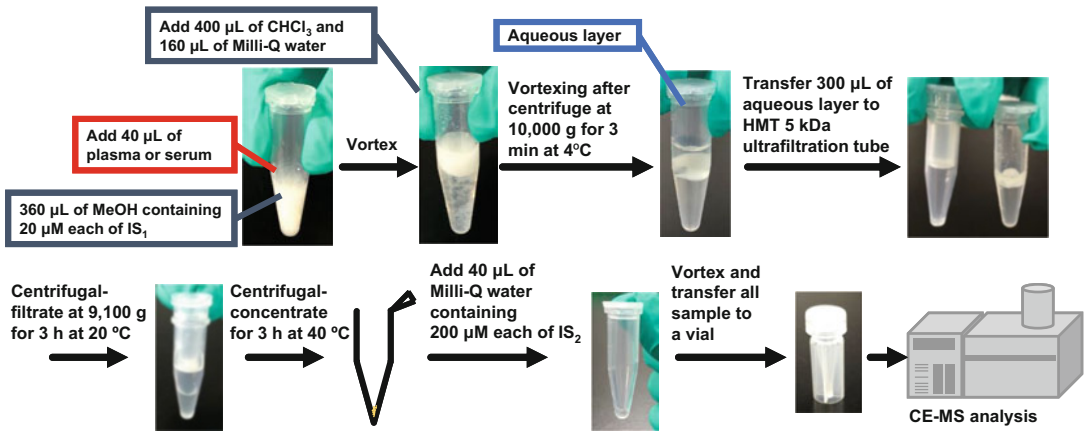

Fig. 1 Sample processing protocol for blood (serum and plasma)

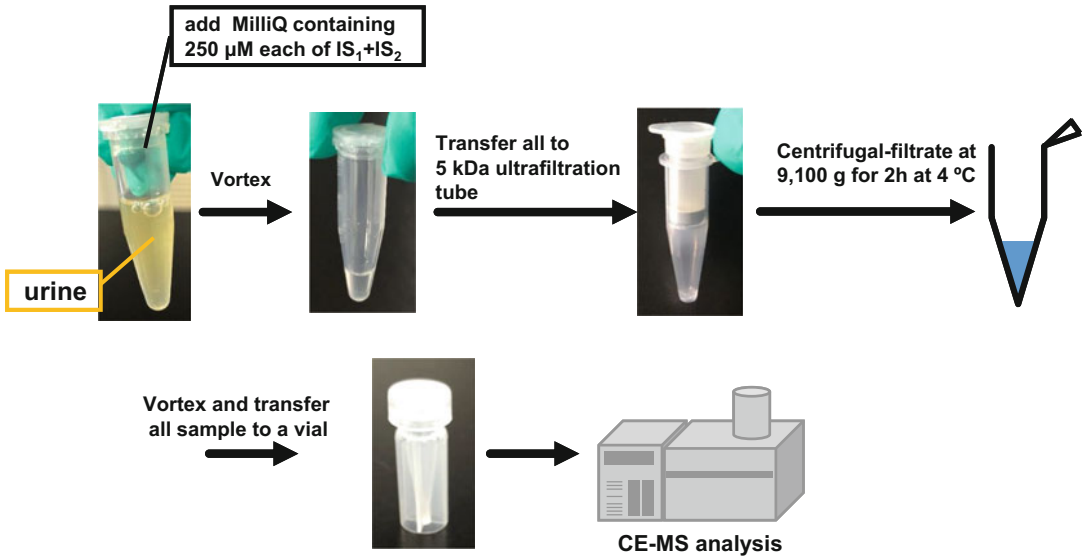

Fig. 2 Sample processing protocol for urine

3.2 Urine Samples

1. Thaw the frozen urine samples at 4 °C for approximately 1.5 h.
2. Homogenize the samples using a vortex mixer at room temperature (15–25 °C).
3. Transfer 20 μL of each sample to a 1.5-mL Eppendorf tube with 80 μL of Milli-Q water containing internal standards (IS_1 + IS_2 250 μM of methionine sulfone, 2-morpholinoethanesulfonic acid, D-camphor-10-sulfonic acid, 3-aminopyrrolidine, and trimesate).
4. Centrifuge the mixture at 9100 × *g* for 2 h at 4 °C and filter it through a 5-kDa cutoff filter (Millipore) to remove any large molecules.
7. Transfer the filtrate to a vial for analysis by CE–MS (Fig. 2).

3.3 Saliva Samples

1. Thaw the frozen samples at 4 °C for approximately 1.5 h.
2. Transfer 100 μL of the saliva sample to a 1.5-mL tube with a 5-kDa cutoff filter (Millipore) using a pipette.
3. Centrifuge the tube at 9100 × *g* for 3 h at 4 °C to eliminate any large molecules.
4. If the filtrate volume is less than 45 μL, independently process the rest of the raw saliva samples, and merge the filtrate. Then, transfer 45 μL of the filtrate to a 1.5-ml microtube using a pipette.

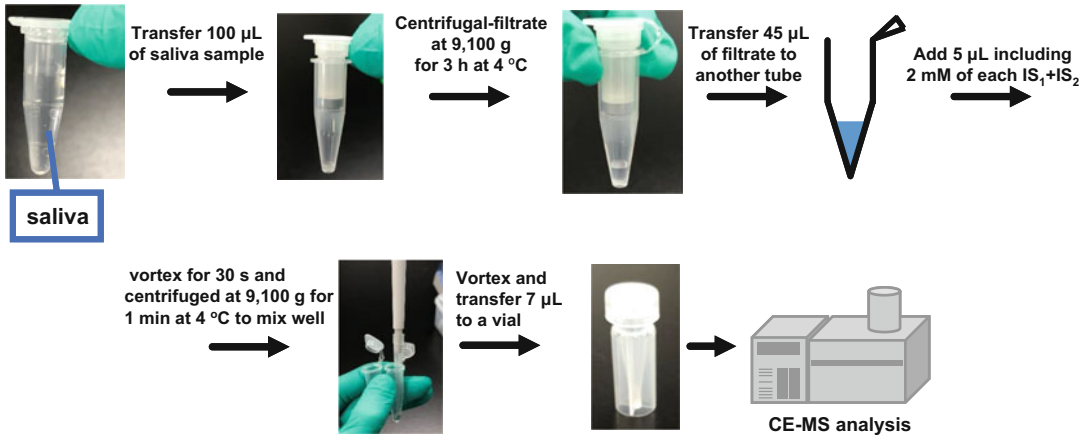

Fig. 3 Sample processing protocol for saliva

5. Add 5 μL of water containing an IS mixture (IS_1 + IS_2: 2 mM methionine sulfone, 2-morpholinoethanesulfonic acid, D-camphor-10-sulfonic acid, 3-aminopyrrolidine, and trimesate) to the tube using a pipette, mix for 30 s using a vortex, and centrifuge at 9100 × *g* for 1 min at 4 °C to mix well.
6. Transfer 7 μL of the mixture to a vial and cover it with a snap cap (Fig. 3).

3.4 Cation Measurement Setting

1. Setup the CE as follows: capillary: fused-silica, i.d. 50 μm × 100 cm; buffer: 1 M formic acid; voltage: positive, 30 kV; temperature: 20 °C; injection: pressure injection 50 mbar, 5 s (approximately 5 nL).
2. Setup time-of-flight (TOF)-MS as follows: ion source: electrospray ionization (ESI); polarity: positive; capillary voltage: 4000 V; fragmentor: 75 V; skimmer: 50 V; OCT RVF: 125 V; drying gas: N_2, 10 L/min; drying gas temperature: 300 °C; nebulizer gas pressure: 7 psig; sheath liquid: cation sheath liquid; flow rate: 10 μL/min; lock mass: 2MeOH ^{13}C isotope *m/z* 66.063061 and Hexakis (2,2-difluoroethoxy) phosphazene *m/z* 622.028963 (*see* **Note 6**).
3. Run cation measurements as follows: Blood and urine samples: 4 min at cation run buffer. Saliva samples: 5 min at Ammonium Formate, 5 min at Mili-Q and 5 min at a cation run buffer.

3.5 Anion Measurement Setting

1. Setup CE as follows: capillary: COSMO(+), i.d. 50 μm × 105 cm; buffer: 50 mM ammonium acetate, pH 8.5; voltage: negative, 30 kV; temperature: 20 °C;

injection: pressure injection 50 mbar, 30 s (approximately 30 nL); preconditioning: 2 min at 50 mM ammonium acetate, pH 3.4, and 5 min at run buffer.

2. Setup TOF-MS as follows: ion source: ESI; polarity: negative; capillary voltage: 3500 V; fragmentor: 100 V; skimmer: 50 V; OCT RVF: 200 V; drying gas: N_2, 10 L/min; drying gas temperature: 300 °C; nebulizer gas pressure: 7 psig; sheath liquid: anion sheath liquid; flow rate: 10 μL/min; lock mass: $2CH_3COOH$ ^{13}C isotope *m/z* 120.038339, Hexakis (2,2-difluoroethoxy) phosphazene+CH_3COOH *m/z* 680.035541; ESI needle: platinum.
3. Run anion measurements as follows: anion for standard mixtures and samples: 2 min at 50 mM ammonium acetate, pH 3.4, and 5 min at an anion run buffer (*see* **Note 7**).

4 Notes

1. Diurnal variation of metabolomic profiles in blood samples has been reported [16, 17]; therefore, a consistent collection time is preferable.
2. The metabolomic profiles of plasma were less sensitive than those of the serum samples [15]. Thus, when serum samples are selected for analysis, the clotting periods should be strictly regulated.
3. The diurnal validation of urinary metabolomic profiles has been previously observed [18]. If the collection time cannot be strictly controlled, the samples should be collected at different times and the variations in metabolomic profiles should be first analyzed. Then, this variation should be used in addition to the comparative statistical analyses to investigate the biomarkers [19].
4. The salivary metabolomic profiles were affected by the duration between saliva collection and the last meal of the patient [20]. A longer duration is preferable.
5. Storing the saliva at room temperature (15–25 °C) affected the salivary metabolomic profiles [14]. Therefore, the saliva should be collected at a lower temperature, for example, using a tube on ice.
6. The total ion electropherogram (TIE) (Fig. 4) and the electric currency of cation measurement are shown in Fig. 5. In both figures, the data with an acceptable quality show flat intensities

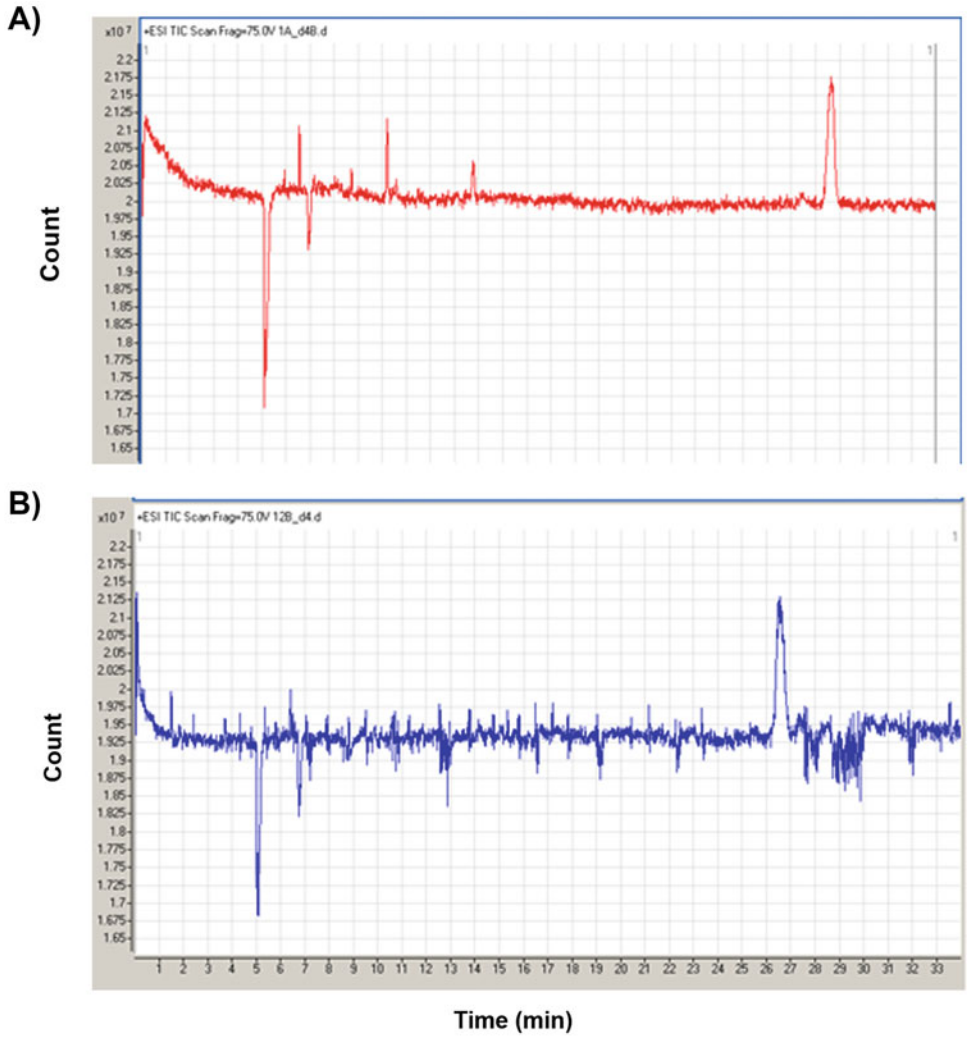

Fig. 4 Total ion electropherogram of cation measurement. (**a**) An example of acceptable quality. (**b**) An example of low quality

and electric currencies during measurement. At an acceptable quality, the electropherogram showed a clear decrease of approximately 5 min, which indicates a large number of Na^+ and K^+ peaks. Furthermore, the clear peak at approximately 29 min indicates neutral substances. Although these areas are not flat, they are acceptable.

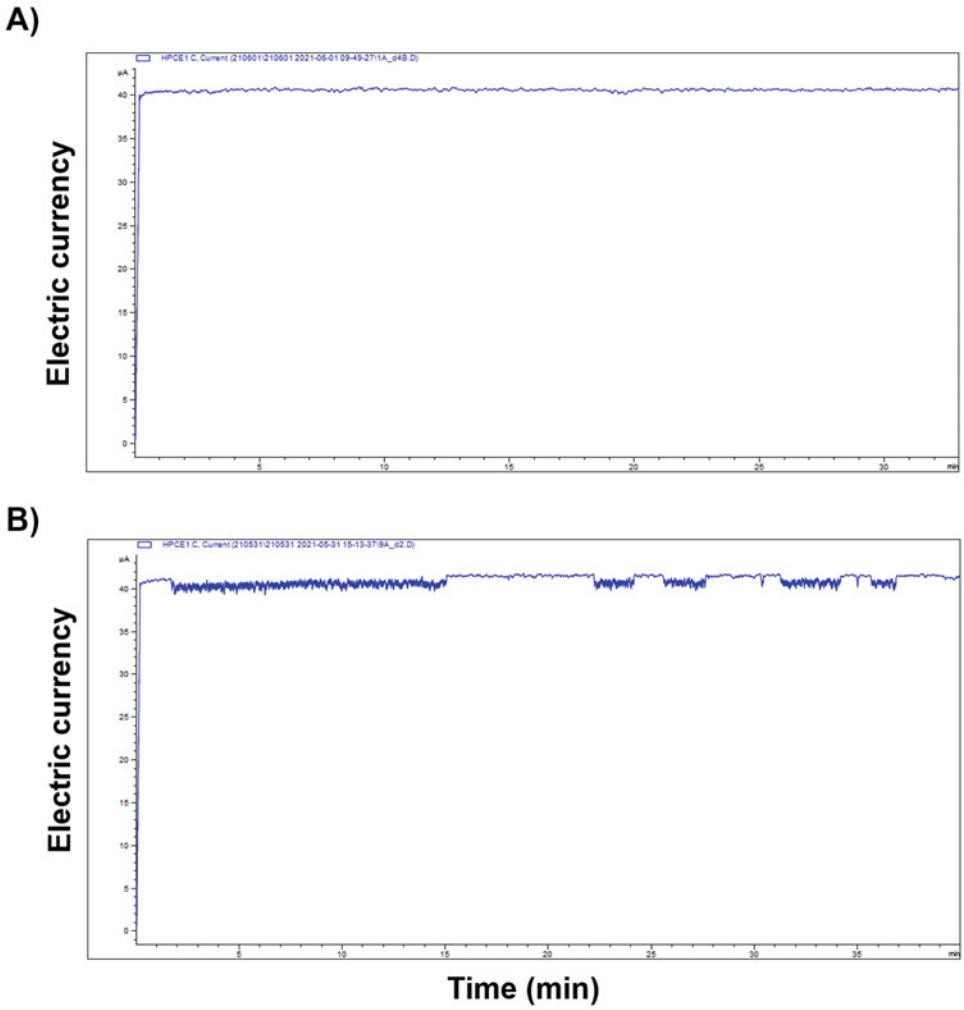

Fig. 5 Electric currency of cation measurement. (**a**) An example of acceptable quality. (**b**) An example of low quality

7. The TIE and the electric currency of anion measurement are shown in Figs. 6 and 7, respectively. In both figures, the data with an acceptable quality show flat intensities and electric currencies during measurement.

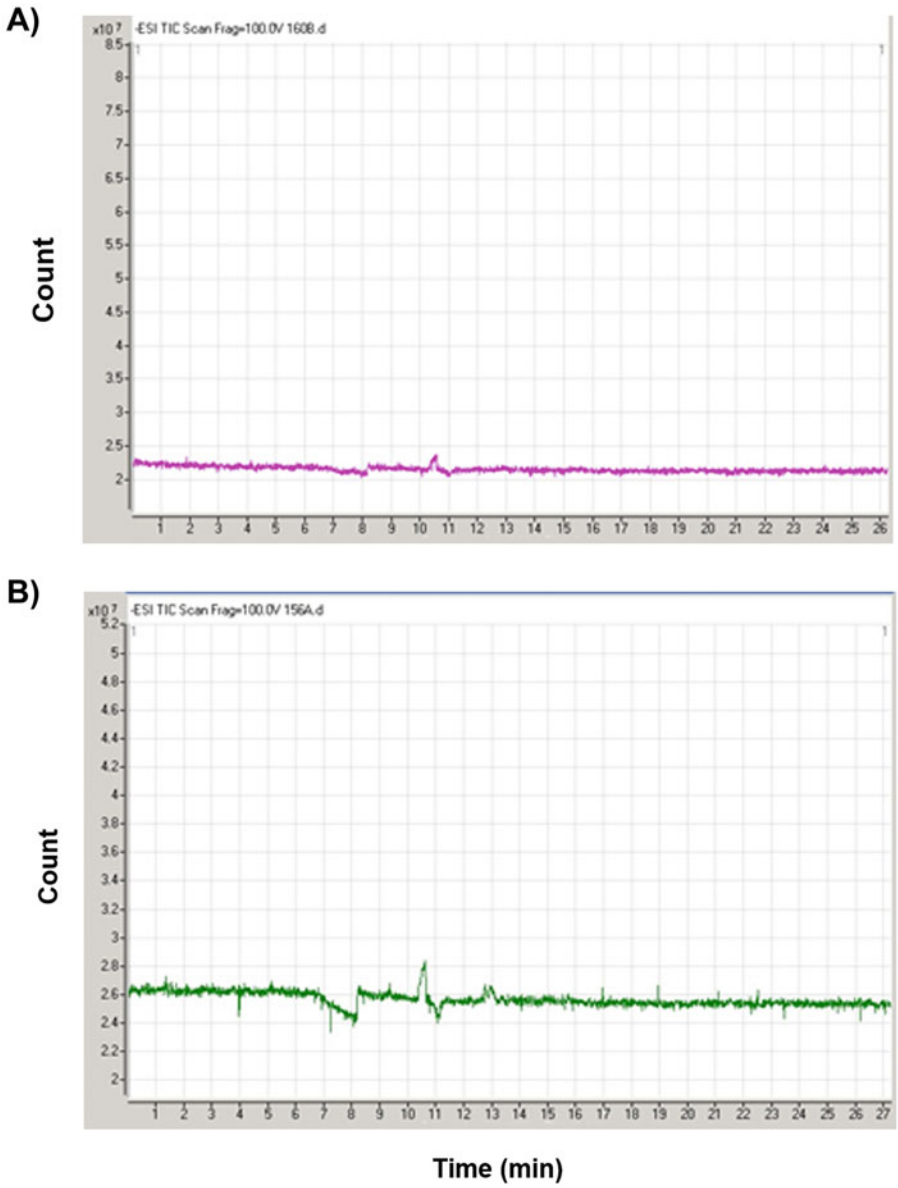

Fig. 6 Total ion electropherogram of anion measurement. (**a**) An example of acceptable quality. (**b**) An example of low quality

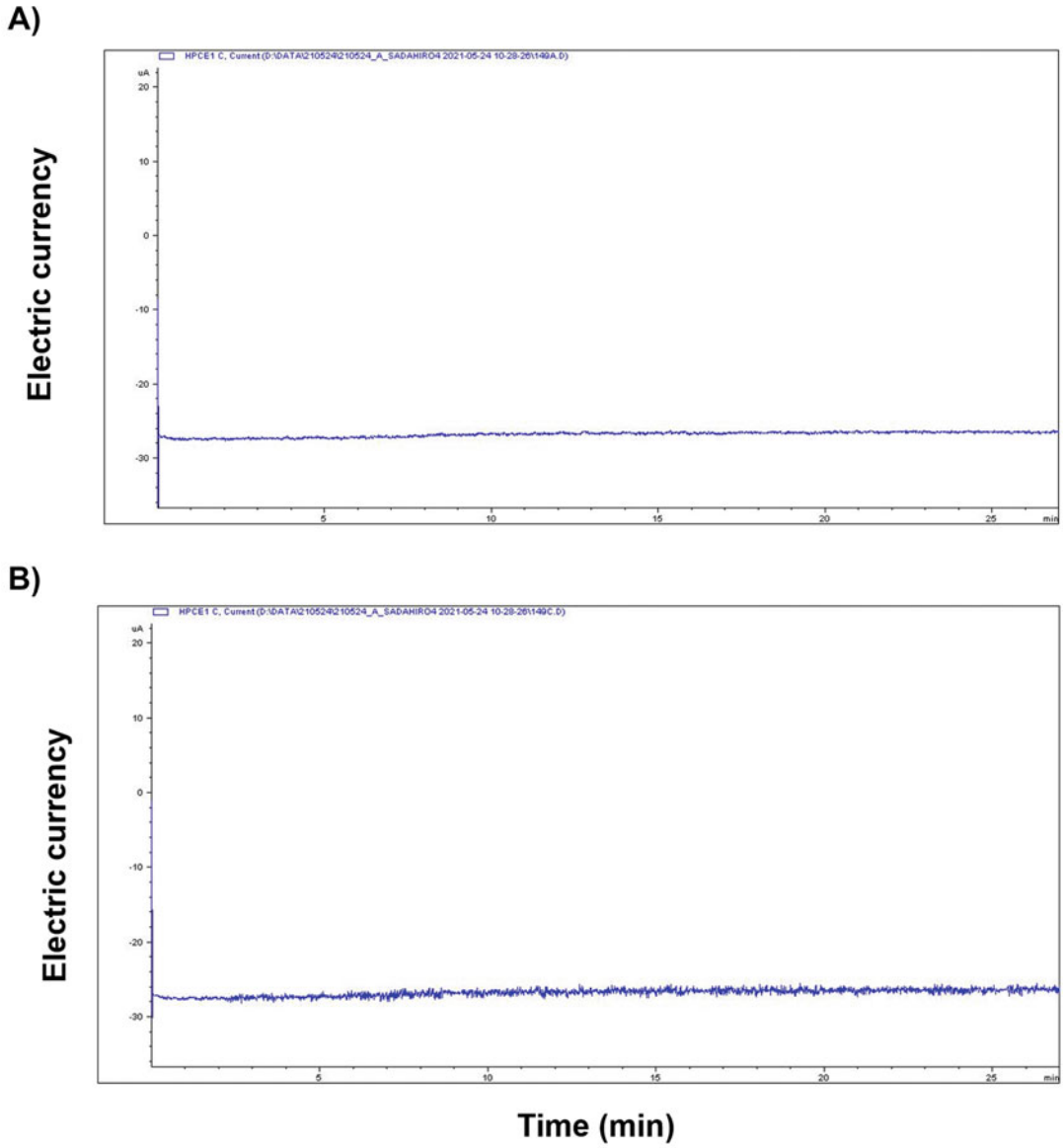

Fig. 7 Electric currency of anion measurement. (**a**) An example of acceptable quality. (**b**) An example of low quality

Acknowledgments

This research was funded by grants from JSPS KAKENHI (grant number 20B205).

References

1. Krait S, Konjaria ML, Scriba GKE (2021) Advances of capillary electrophoresis enantioseparations in pharmaceutical analysis (2017-2020). Electrophoresis. https://doi.org/10.1002/elps.202000359
2. Zhang W, Ramautar R (2021) CE-MS for metabolomics: developments and applications in the period 2018-2020. Electrophoresis 42(4):381–401. https://doi.org/10.1002/elps.202000203
3. Satoh K, Yachida S, Sugimoto M, Oshima M, Nakagawa T, Akamoto S, Tabata S, Saitoh K, Kato K, Sato S, Igarashi K, Aizawa Y, Kajino-Sakamoto R, Kojima Y, Fujishita T, Enomoto A, Hirayama A, Ishikawa T, Taketo MM, Kushida Y, Haba R, Okano K, Tomita M, Suzuki Y, Fukuda S, Aoki M, Soga T (2017) Global metabolic reprogramming of colorectal cancer occurs at adenoma stage and is induced by MYC. Proc Natl Acad Sci U S A 114(37): E7697–e7706. https://doi.org/10.1073/pnas.1710366114
4. Hirayama A, Kami K, Sugimoto M, Sugawara M, Toki N, Onozuka H, Kinoshita T, Saito N, Ochiai A, Tomita M, Esumi H, Soga T (2009) Quantitative metabolome profiling of colon and stomach cancer microenvironment by capillary electrophoresis time-of-flight mass spectrometry. Cancer Res 69(11):4918–4925. https://doi.org/10.1158/0008-5472.Can-08-4806
5. Udo R, Katsumata K, Kuwabara H, Enomoto M, Ishizaki T, Sunamura M, Nagakawa Y, Soya R, Sugimoto M, Tsuchida A (2020) Urinary charged metabolite profiling of colorectal cancer using capillary electrophoresis-mass spectrometry. Sci Rep 10(1):21057. https://doi.org/10.1038/s41598-020-78038-2
6. Sugimoto M, Wong DT, Hirayama A, Soga T, Tomita M (2010) Capillary electrophoresis mass spectrometry-based saliva metabolomics identified oral, breast and pancreatic cancer-specific profiles. Metabolomics 6(1):78–95. https://doi.org/10.1007/s11306-009-0178-y
7. Murata T, Yanagisawa T, Kurihara T, Kaneko M, Ota S, Enomoto A, Tomita M, Sugimoto M, Sunamura M, Hayashida T, Kitagawa Y, Jinno H (2019) Salivary metabolomics with alternative decision tree-based machine learning methods for breast cancer discrimination. Breast Cancer Res Treat 177(3):591–601. https://doi.org/10.1007/s10549-019-05330-9
8. Asai Y, Itoi T, Sugimoto M, Sofuni A, Tsuchiya T, Tanaka R, Tonozuka R, Honjo M, Mukai S, Fujita M, Yamamoto K, Matsunami Y, Kurosawa T, Nagakawa Y, Kaneko M, Ota S, Kawachi S, Shimazu M, Soga T, Tomita M, Sunamura M (2018) Elevated polyamines in saliva of pancreatic cancer. Cancer 10(2):doi:10.3390/cancers10020043
9. Ishikawa S, Wong DTW, Sugimoto M, Gleber-Netto FO, Li F, Tu M, Zhang Y, Akin D, Iino M (2019) Identification of salivary metabolites for oral squamous cell carcinoma and oral epithelial dysplasia screening from persistent suspicious oral mucosal lesions. Clin Oral Investig 23(9):3557–3563. https://doi.org/10.1007/s00784-018-2777-3
10. Sugimoto M (2020) Salivary metabolomics for cancer detection. Expert Rev Proteomics 17(9):639–648. https://doi.org/10.1080/14789450.2020.1846524
11. Sugimoto M, Kawakami M, Robert M, Soga T, Tomita M (2012) Bioinformatics tools for mass spectroscopy-based metabolomic data processing and analysis. Curr Bioinforma 7(1): 96–108. https://doi.org/10.2174/157489312799304431
12. Kawanishi N, Hoshi N, Masahiro S, Enomoto A, Ota S, Kaneko M, Soga T, Tomita M, Kimoto K (2019) Effects of inter-day and intra-day variation on salivary metabolomic profiles. Clin Chim Acta 489:41–48. https://doi.org/10.1016/j.cca.2018.11.030
13. Okuma N, Saita M, Hoshi N, Soga T, Tomita M, Sugimoto M, Kimoto K (2017) Effect of masticatory stimulation on the quantity and quality of saliva and the salivary metabolomic profile. PLoS One 12(8):e0183109. https://doi.org/10.1371/journal.pone.0183109
14. Tomita A, Mori M, Hiwatari K, Yamaguchi E, Itoi T, Sunamura M, Soga T, Tomita M, Sugimoto M (2018) Effect of storage conditions on salivary polyamines quantified via liquid chromatography-mass spectrometry. Sci Rep 8(1):12075. https://doi.org/10.1038/s41598-018-30482-x
15. Hirayama A, Sugimoto M, Suzuki A, Hatakeyama Y, Enomoto A, Harada S, Soga T, Tomita M, Takebayashi T (2015) Effects of processing and storage conditions on charged metabolomic profiles in blood. Electrophoresis 36(18):2148–2155. https://doi.org/10.1002/elps.201400600
16. Kasukawa T, Sugimoto M, Hida A, Minami Y, Mori M, Honma S, Honma K, Mishima K, Soga T, Ueda HR (2012) Human blood

metabolite timetable indicates internal body time. Proc Natl Acad Sci U S A 109(37): 15036–15041. https://doi.org/10.1073/pnas.1207768109

17. Minami Y, Kasukawa T, Kakazu Y, Iigo M, Sugimoto M, Ikeda S, Yasui A, van der Horst GT, Soga T, Ueda HR (2009) Measurement of internal body time by blood metabolomics. Proc Natl Acad Sci U S A 106(24): 9890–9895. https://doi.org/10.1073/pnas.0900617106
18. Slupsky CM, Rankin KN, Wagner J, Fu H, Chang D, Weljie AM, Saude EJ, Lix B, Adamko DJ, Shah S, Greiner R, Sykes BD, Marrie TJ (2007) Investigations of the effects of gender, diurnal variation, and age in human urinary metabolomic profiles. Anal Chem 79(18):6995–7004. https://doi.org/10.1021/ac0708588
19. Nakajima T, Katsumata K, Kuwabara H, Soya R, Enomoto M, Ishizaki T, Tsuchida A, Mori M, Hiwatari K, Soga T, Tomita M, Sugimoto M (2018) Urinary polyamine biomarker panels with machine-learning differentiated colorectal cancers, benign disease, and healthy controls. Int J Mol Sci 19(3):doi:10.3390/ijms19030756
20. Ishikawa S, Sugimoto M, Kitabatake K, Tu M, Sugano A, Yamamori I, Iba A, Yusa K, Kaneko M, Ota S, Hiwatari K, Enomoto A, Masaru T, Iino M (2017) Effect of timing of collection of salivary metabolomic biomarkers on oral cancer detection. Amino Acids 49(4): 761–770. https://doi.org/10.1007/s00726-017-2378-5

Chapter 9

Capillary Electrophoresis–Mass Spectrometry for the Direct Analysis of Metabolites in Highly Saline Samples Using In-Capillary Preconcentration

Marlien van Mever and Rawi Ramautar

Abstract

Capillary electrophoresis–mass spectrometry (CE–MS) is gaining interest for metabolomics studies because of its high separation efficiency, selectivity, and versatility. The ability to inject nanoliters from only a few microliters of sample in the injection vial makes this approach very suited for volume-limited applications. However, the low injection volumes could compromise the detection sensitivity of CE–MS, thereby potentially limiting its scope in metabolomics. To overcome this issue, online sample preconcentration methods have been developed to increase sample-loading volumes without hampering the intrinsic high separation efficiency of CE. In this protocol, online preconcentration with sample stacking based on pH junction was assessed for the direct profiling of endogenous metabolites in rat brain microdialysates. Sample stacking was realized by a pre-injection of ammonium hydroxide, followed by a large sample injection (i.e., about 17% of the total capillary volume). It is shown that this relatively simple and fast preconcentration procedure is fully compatible with the high-salt concentration in microdialysates and significantly improves the detection sensitivity of the CE–MS method.

Key words Capillary electrophoresis, Direct analysis, Mass spectrometry, Metabolic profiling, Brain microdialysate

1 Introduction

Capillary electrophoresis coupled to mass spectrometry (CE–MS) can be considered a strong analytical tool for the highly efficient analysis of polar and charged compounds in biological samples [1–3]. A main advantage of CE is that only small sample amounts are required for sample injection (nL-range), which makes it an attractive tool for the analysis of volume-limited or scarcely available samples. However, a limitation of the low sample-loading capacity of CE is that the detection sensitivity can be compromised, especially in conjunction with a classical sheath-liquid interface. This can be particularly problematic for the analysis of low-molecular-

Raúl González-Domínguez (ed.), *Mass Spectrometry for Metabolomics*,
Methods in Molecular Biology, vol. 2571, https://doi.org/10.1007/978-1-0716-2699-3_9,
© The Author(s), under exclusive license to Springer Science+Business Media, LLC, part of Springer Nature 2023

mass biomolecules in brain fluids such as cerebrospinal fluid (CSF) and microdialysates, as this has shown to be analytically challenging, mainly due to the low concentrations of bioactive compounds in the central nervous system (CNS), but also because of the intrinsically low sample volumes in the case of microdialysates.

Over the past decades, many online sample preconcentration methods have been developed to increase sample-loading volumes without hampering the intrinsic high separation efficiency of CE [4]. It has been shown that the use of sample preconcentration is able to improve detection limits of CE by at least 100 to up to 1000-fold. A few examples of sample preconcentration include field-amplified sample injection (FASI) [5], large-volume sample stacking (LVSS) [6], and electrokinetic supercharging (EKS) [7]. However, not all in-capillary preconcentration procedures are directly suitable for highly conductive or saline sample matrices, resulting in the need for sample clean-up before injection. In the case of low sample volumes (<50 μL), this might be problematic. On the other hand, preconcentration techniques such as transient isotachophoresis (tITP) [8] and preconcentration based on dynamic pH junction [9] have shown to be highly compatible with high saline samples and show potential for the analysis of brain fluids, thereby without performing any sample clean-up steps.

Recently, an online preconcentration procedure with sample stacking based on pH junction was assessed for the direct profiling of endogenous metabolites in low volumes of rat brain microdialysates, i.e., without using any sample preparation or derivatization [10]. Sample stacking was realized by a pre-injection of ammonium hydroxide, followed by a large sample injection (i.e., about 17% of the total capillary volume). The sample stacking parameters were optimized using response surface methodology (RSM). In this chapter, attention is paid to the methodological aspects of the sample stacking procedure for the profiling of cationic metabolites in rat brain microdialysis samples. It is shown that this relatively simple preconcentration procedure is fully compatible with the high-salt concentration in microdialysates and significantly improves the detection sensitivity of the CE–MS method.

2 Materials

A Milli-Q Advantage A10 water purification system with a 0.22-μM pore-size filter is used to obtain pure water. All chemicals used are of analytical grade or higher purity.

2.1 Solutions and Samples for Analysis

1. Background electrolyte (BGE) solution: 10% (v/v) acetic acid, pH 2.2. Add approximately 80 mL of water in a 100-mL volumetric flask and add 10 mL of concentrated acetic acid.

Add water to the gauge line. Mix the solution thoroughly. Store at 4 °C.

2. Sheath-liquid solution for CE–MS analysis: Mix 100 mL of water with 100 mL of isopropanol. Add 60 μL of concentrated acetic acid to this solution. Degas for 10 min in an ultrasonic bath before the first use (*see* **Note 1**).
3. Ammonium hydroxide solution: 5% (v/v) ammonium hydroxide, pH ~12. Add 5 mL of ammonium hydroxide (25%) into a 25-mL volumetric flask and add water up to the gauge line. Mix the solution thoroughly. Store at 4 °C (*see* **Note 2**).
4. Sodium hydroxide solution: 1 M sodium hydroxide. Add 1 g of sodium hydroxide (98%) into a 25-mL volumetric flask and add water up to the gauge line. Mix the solution thoroughly. Store at 4 °C.
5. Hydrochloric acid solutions: 1 M hydrochloric acid, 0.1 M hydrochloric acid. For 1 M hydrochloric acid solution, add 2.1 mL hydrochloric acid (37%) into a 25-mL volumetric flask and add water up to the gauge line. Mix the solution thoroughly. For 0.1 M hydrochloric acid solution, add 2.5 mL of 1 M hydrochloric acid solution into a 25-mL volumetric flask and add water up to the gauge line. Mix the solution thoroughly. Store both solutions at 4 °C.
6. Metabolite standard mixture: Add 5 mL of acetonitrile and 0.5 mL of concentrated formic acid in a 100-mL volumetric flask and add water to the gauge line. Prepare stock solutions of 10 mM for glycine, alanine, sarcosine, gamma-aminobutyric acid, serine, cytosine, creatinine, proline, valine, threonine, homoserine, hydroxyproline, creatine, isoleucine, leucine, asparagine, ornithine, aspartic acid, anthranilic acid, glutamine, lysine, glutamic acid, methionine, histidine, phenylalanine, arginine, tyrosine, and tryptophan in the water:acetonitrile mixture. Make aliquots of stock solutions by dilution with water to obtain a working solution of 50 μM for each analyte. Store stock and working solutions at −80 °C.
7. Internal standard mixture: Prepare stock solutions of 10 mM for [13C2]-glutamine, [13C6]-lysine, [13C2]-aspartic acid, [13C5]-valine, and [13C2]-glutamic acid in BGE. Make aliquots of stock solutions by dilution with BGE to obtain a working solution of 10 μM for each analyte. Store stock and working solutions at −80 °C when not in use.

2.2 Rat Brain Microdialysis Samples

1. Rat brain microdialysis samples: Rat brain microdialysis samples were kindly provided by L. Légat (Department of Pharmaceutical Chemistry, Drug Analysis and Drug Information, Center for Neurosciences, Vrije Universiteit Brussel, Brussel, Belgium) [11] and stored at −80 °C when not in use.

2. Perfusion fluid: 147 mM NaCl, 2.3 mM $CaCl_2$, 4 mM KCl. Add 86 mg of NaCl, 2.6 mg of $CaCl_2$, and 3.0 mg of KCl into a 10-mL volumetric flask and add water to the gauge line. Mix the solution thoroughly (*see* **Note 3**). Store at 4 °C.

2.3 Analytical Equipment

1. CE separation: Fused-silica capillaries (50 μM I.D. × 90 cm total length) are used.
2. CE–MS: A 7100-capillary electrophoresis system from Agilent Technologies hyphenated with a 6230 TOF, also from Agilent.
3. CE–MS coupling: Coupling is realized via a coaxial sheath-liquid ESI interface equipped with a standard triple-tube sprayer.
4. Software: MassHunter version B.06.00 is used for data acquisition, instrument control, and data treatment. Peak integration is performed using extracted ion electropherograms (EIE) based on the m/z for the protonated molecular ion $[M + H]^+$ of each metabolite. Excel is used for further data processing (*see* **Note 4**).

3 Methods

Please use all appropriate laboratory safety procedures, including safety glasses, lab coat, and gloves, when performing the experiments described in this protocol.

The CE–MS setup and the optimization of MS settings have been described previously, and interested readers are referred to another publication [12].

3.1 Preparation of Microdialysates/Perfusion Fluid

1. Prior to CE–MS analysis, mix the microdialysis sample/perfusion fluid with BGE (1:1, v/v), containing internal standards with final concentrations of 5 μM and centrifuge for 10 min at 4 °C and 16,100 × *g* (*see* **Note 5**).

3.2 CE–MS Analysis

3.2.1 Capillary Conditioning and System Performance Check

1. Prior to first use, new capillaries installed in the coaxial sheath liquid interface are conditioned by subsequently rinsing, at 5 bar for 1 min, with methanol, water, 1 M sodium hydroxide, water, 1 M hydrochloric acid, water, 0.1 M hydrochloric acid, water, and background electrolyte (BGE).
2. Perform at least three trial runs with the newly conditioned capillary using the metabolite standard mixture as the sample until a stable current, repeatable migration times, and signal intensities are observed (*see* **Note 6**).

3.2.2 CE–MS Analysis of Biological Samples with Online Sample Preconcentration

1. Add 10 μL of the metabolite standard mixture into an empty 250-μL microvial (*see* **Notes 7**, **8**, and **9**) and put the vial in the inlet sample tray.
2. Fill a CE vial with 5% (v/v) ammonium hydroxide solution (1 mL).
3. Inject a small plug of ammonium hydroxide using an injection volume of ~12 nL (13 s 50 mbar) (*see* **Note 10**).
4. Inject the metabolite standard mixture using an injection volume of ~291 nL (327 s 50 mbar) (*see* **Note 11**).
5. Inject a small plug of BGE using an injection volume of ~5 nL (5 s 50 mbar).
6. Apply a separation voltage of 30 kV to start the electrophoretic process. In-capillary preconcentration by pH junction takes place (*see* **Note 12**). The effect of sample stacking is shown in Fig. 1 for model compounds lysine, isoleucine, leucine, and aspartic acid.

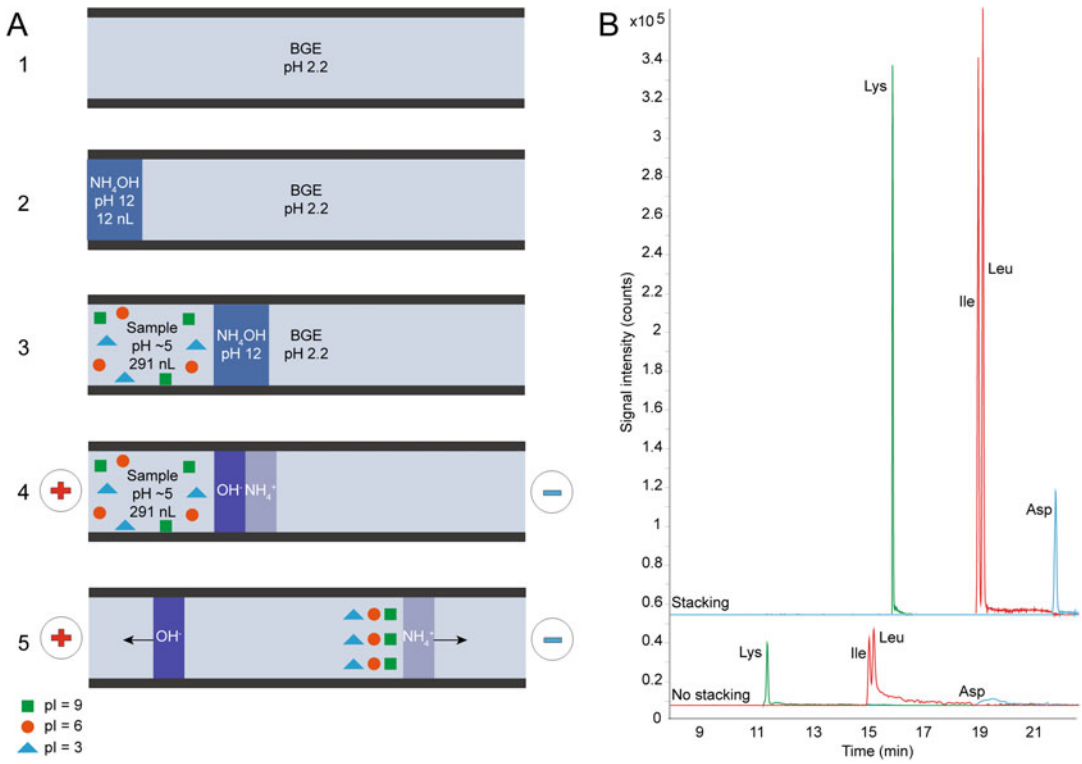

Fig. 1 (**a**) Schematic diagram of sample stacking based on pH junction: (1) fill the capillary with low pH BGE solution, (2) hydronamically inject a plug of ammonium hydroxide (NH_4OH) onto the capillary, (3) hydronamically inject a (large) sample plug, (4) apply a positive voltage: discrete electrolyte zones are formed, analytes are focused, (5) regular electrophoretic separation of analytes. (**b**) EIE for lysine, isoleucine, leucine, and aspartic acid, top: stacking, bottom: no stacking

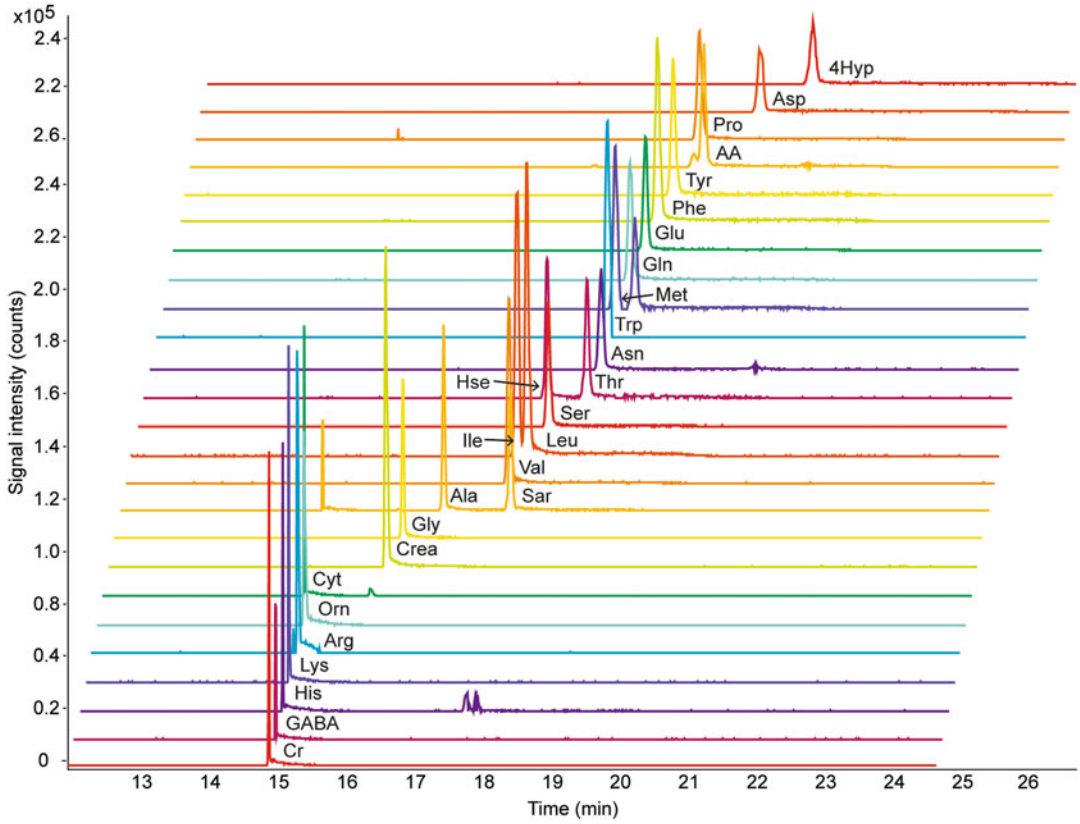

Fig. 2 Extracted-ion electropherograms obtained by CE–MS from the analysis of 28 metabolites (5 μM) spiked in perfusate. Separation conditions: BGE, 10% acetic acid; sample injection volume 291 nL; ammonium hydroxide (concentration: 5%) pre-injection volume, 12 nL. (The figure is re-used from Ref. [10] with permission)

7. Between sample injections, rinse the capillary with milli-Q, methanol, milli-Q, and BGE, each at 5 bar for 60 s (*see* **Note 13**).
8. Repeat **steps 1–7** for metabolites spiked into perfusion fluid. A typical profile obtained using in-capillary preconcentration based on pH junction is shown in Fig. 2 (*see* **Note 14**). Repeat the analysis at least 3 times and check if a typical current profile, as shown in Fig. 3 (*see* **Note 15**), and repeatable migration times (*see* **Note 16**) are observed.
9. After analysis of the perfusion fluid, analyze the actual rat brain microdialysis samples. Figure 4 depicts the electropherograms corresponding to a rat brain microdialysis sample.

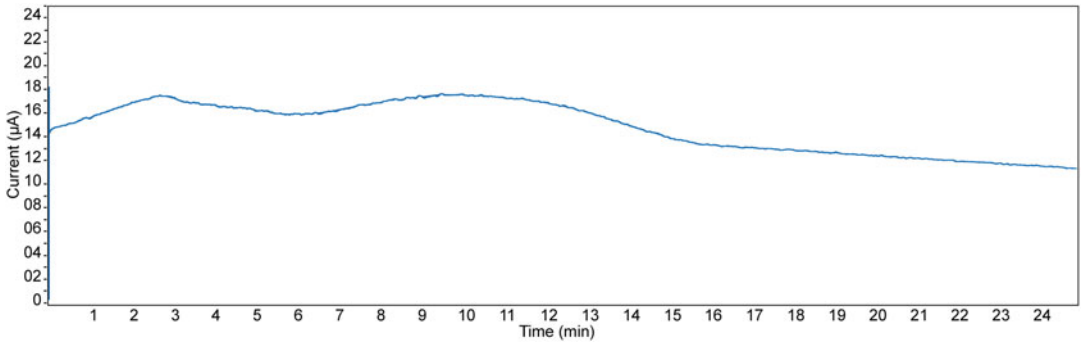

Fig. 3 Typical current profile obtained by CE–MS analysis of perfusate using sample preconcentration based on pH junction. Separation conditions: BGE, 10% acetic acid; sample injection volume 291 nL; ammonium hydroxide (concentration: 5%) pre-injection volume, 12 nL

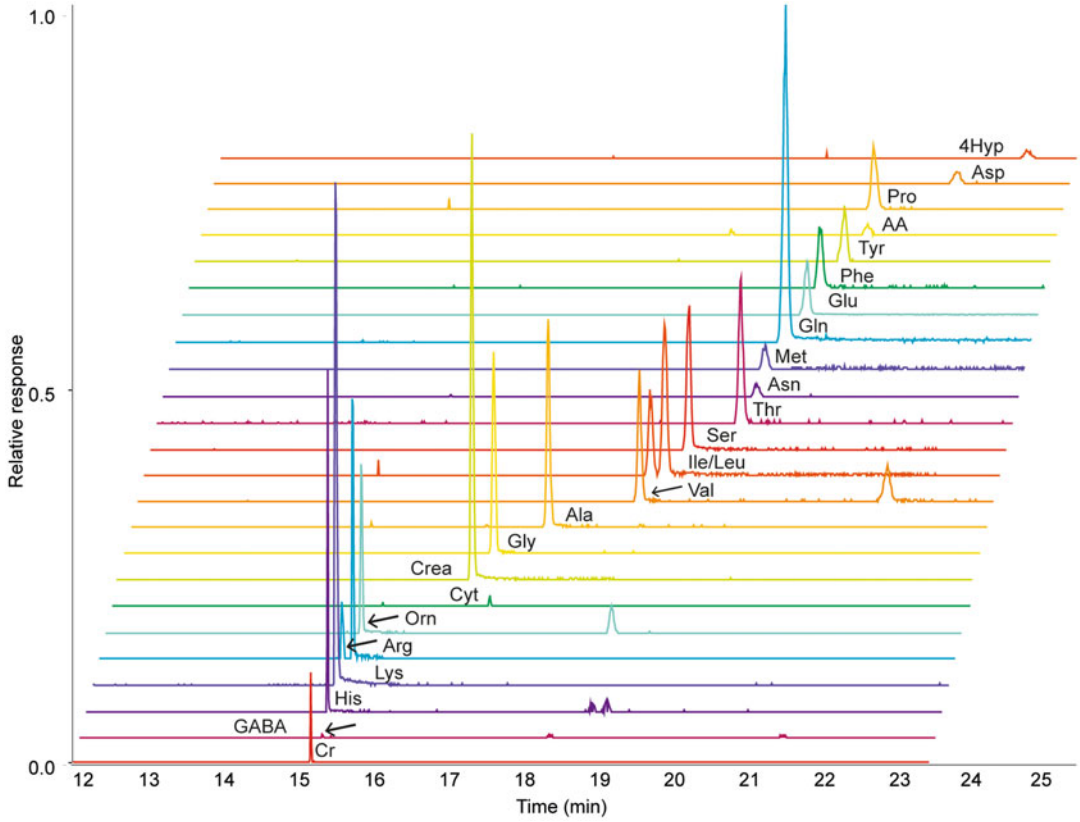

Fig. 4 Extracted-ion electropherograms obtained by CE–MS from the analysis of endogenous metabolites in basal rat brain microdialysate. Separation conditions: BGE, 10% acetic acid; sample injection volume 291 nL; ammonium hydroxide (concentration: 5%) pre-injection volume, 12 nL. (The figure is re-used from Ref. [10] with permission)

4 Notes

1. Prepare fresh BGE weekly.
2. Prepare fresh ammonium hydroxide every 2 weeks and check the pH with pH-fix indicator strips before use. The pH should be ~12.
3. Prepare fresh perfusion fluid weekly.
4. As a microdialysis test matrix, perfusion fluid (modified Ringer's solution) is used, mainly because of the scarce availability of the valuable rat brain microdialysis samples.
5. From the sheath-liquid, isopropanol ($C_3H_8OH^+$) and its clusters ($[(C_3H_8O)_2 + H]^+$ and ($[(C_3H_8O)_3 + H]^+$) with corresponding m/z values of 61.06479, 121.12231, and 181.17982, respectively, are used as lock masses to correct for mass shift.
6. To check system performance, a test mixture containing six metabolites is adequate: procaine (internal standard), paracetamol (internal standard), lysine, aspartic acid, and isoleucine/leucine.
7. When a low sample volume is filled into the sample vial, air bubbles are easily formed on the bottom of the sample insert vial. Check carefully for air bubbles, and if necessary, remove the air bubble using a pipette.
8. The use of build-in microvials ensures proper alignment of the microvial, allowing for repeatable injections and lower sample volumes (down to 5 μL).
9. The use of softer material caps (polyethylene olefin) prevents incomplete puncturing due to resistance when using harder material caps and slows down blunting of the prepuncturer.
10. Include a water-dip step before and after ammonium hydroxide injection to prevent contamination of BGE and sample.
11. In case compatible with the CE system, then it is possible to perform the sample injection using 100 mbar (135 s).
12. Analytes with a high electrophoretic mobility under the current circumstances (e.g., lysine) possess a positive charge at pH 2.2 and show efficient stacking and good peak shapes/intensities. Analytes with a lower electrophoretic mobility, and lower pKa (e.g., aspartic acid), are only partially charged at pH 2.2, and therefore stacking based on pH junction is less efficient, and peak shapes/intensities are less desirable.
13. For optimal results, use a fresh vial with BGE for every analysis.
14. Fast migrating compounds that migrate close to the salt plug (MT ~ 14.0–15.5 min) experience more prominent matrix

effect than slower migrating compounds. Therefore, the use of stable isotope labeled standards for every metabolite (or if not possible, for every pKa/pI range) is advised before quantification to correct for differences in matrix effect and also in stacking efficiency.

15. The typical current profile during CE analysis using sample preconcentration is not completely constant over time, as shown in Fig. 3.
16. When using sample preconcentration, the total analysis time is increased from 20 to 25 min.

Acknowledgments

M. van Mever and Dr. R. Ramautar would thank L. Légat from the Vrije Universiteit Brussel for providing the rat brain microdialysis samples. M. van Mever and Dr. R. Ramautar acknowledge the financial support of the Vidi grant scheme of the Netherlands Organization of Scientific Research (NWO Vidi 723.016.003).

References

1. Zhang W, Ramautar R (2021) CE-MS for metabolomics: developments and applications in the period 2018–2020. Electrophoresis 42(4):381–401
2. Kristoff C, Bwanali L, Veltri L, Gautam G, Rutto P, Newton E et al (2020) Challenging bioanalyses with capillary electrophoresis. Anal Chem 92(1):49
3. Gao Z, Zhong W (2021) Recent (2018–2020) development in capillary electrophoresis. Anal Bioanal Chem 22:1–16
4. Jarvas G, Guttman A, Miękus N, Bączek T, Jeong S, Chung DS et al (2020) Practical sample pretreatment techniques coupled with capillary electrophoresis for real samples in complex matrices. TrAC 122:115702
5. Liu JX, Aerts JT, Rubakhin SS, Zhang XX, Sweedler JV (2014) Analysis of endogenous nucleotides by single cell capillary electrophoresis-mass spectrometry. Analyst 139(22):5835–5842
6. Flores-Aguilar JF, Medrano LC, Perez-Escalante E, Rodriguez JA, Camacho-Mendoza RL, Ibarra IS (2019) Large-volume sample stacking with polarity switching for analysis of azo dyes in water samples by capillary electrophoresis. Int J Environ Anal Chem 99(13):1255–1267
7. Wells SS, Dawod M, Kennedy RT (2019) CE-MS with electrokinetic supercharging and application to determination of neurotransmitters. Electrophoresis 40(22):2946–2953
8. Zhang W, Guled F, Hankemeier T, Ramautar R (2020) Profiling nucleotides in low numbers of mammalian cells by sheathless CE–MS in positive ion mode: circumventing corona discharge. Electrophoresis 41(5–6). https://doi.org/10.1002/elps.201900417
9. Tak YH, Somsen GW, Jong GJ (2011) Optimization of dynamic pH junction for the sensitive determination of amino acids in urine by capillary electrophoresis. Anal Bioanal Chem 401(10):3275
10. van Mever M, Segers K, Drouin N, Guled F, Vander Heyden Y, Van Eeckhaut A et al (2020) Direct profiling of endogenous metabolites in rat brain microdialysis samples by capillary electrophoresis-mass spectrometry with on-line preconcentration. Microchem J 2020: 156
11. Légat L, Brouwers S, Smolders IJ, Dupont AG (2017) Hypotensive response to angiotensin II Type 2 receptor stimulation in the rostral ventrolateral medulla requires functional GABA-A receptors. Front Neurosci 11:346
12. Drouin N, Pezzatti J, Gagnebin Y, González-Ruiz V, Schappler J, Rudaz S (2018) Effective mobility as a robust criterion for compound annotation and identification in metabolomics: toward a mobility-based library. Anal Chim Acta 1032:178–187

Chapter 10

Chemical Derivatization to Enhance Capillary Electrophoresis Mass Spectrometric Analysis of Acidic Metabolites in Mammalian Cells

Marlien van Mever and Rawi Ramautar

Abstract

The simultaneous analysis of cationic and anionic metabolites using capillary electrophoresis–mass spectrometry (CE–MS) has been considered challenging, as often two different analytical methods are required. Although CE–MS methods for cationic metabolite profiling have already shown good performance metrics, the profiling of anionic metabolites often results in relatively low sensitivity and poor repeatability caused by problems related to unstable electrospray and corona discharge when using reversed CE polarity and detection by MS in negative ionization mode. In this protocol, we describe a chemical derivatization procedure that provides a permanent positive charge to acidic metabolites, thereby allowing us to profile anionic metabolites by CE–MS using exactly the same separation conditions as employed for the analysis of basic metabolites. The utility of the overall approach is demonstrated for the analysis of energy metabolism-related metabolites in low numbers of HepG2 cells.

Key words Capillary electrophoresis, Carboxylic acid metabolites, Chemical derivatization, Metabolomics, HepG2 cells

1 Introduction

The separation mechanism of CE and the high selectivity and sensitivity of mass spectrometry (MS) makes CE–MS a well-suited approach for the efficient profiling of polar ionogenic metabolites in metabolomics, thereby providing complementary information to other separation techniques [1, 2]. CE–MS has especially shown good performance metrics for cationic metabolite profiling, as recently demonstrated for the analysis of large sample cohorts and in a Metabo-ring trial [3, 4]. On the other hand, the profiling of anionic metabolites by CE–MS has been considered challenging, mainly due to relatively poor metabolite responses and inadequate repeatability caused by problems related to unstable electrospray

Raúl González-Domínguez (ed.), *Mass Spectrometry for Metabolomics*,
Methods in Molecular Biology, vol. 2571, https://doi.org/10.1007/978-1-0716-2699-3_10,
© The Author(s), under exclusive license to Springer Science+Business Media, LLC, part of Springer Nature 2023

and corona discharge when using reversed CE polarity and detection by MS in negative ionization mode [5, 6].

To circumvent the issues described above, various efforts have been made to improve anionic metabolite profiling by CE–MS over the last decade: in this case, by the use of a platinum electrospray needle [8], by the use of ion-pair reagents [7], by the use of a background electrolyte (BGE) with high pH and an electrophoretic separation in normal polarity mode [8], and by considering the detection of anions as their ammonium adducts [9]. Alternatively, instead of (further) optimizing the instrumental setup and electrophoretic separation conditions, chemical derivatization can be combined with CE–MS to profile anionic metabolites in biological samples. In the latter approach, metabolites are derivatized with one or several chemical reagents prior to analysis by CE–MS. As the physicochemical properties of the metabolites are altered, an enhanced analytical performance can be achieved, including a higher detection sensitivity and improved quantification ability [10]. Various efforts using chemical derivatization have already been applied to analyze anionic metabolites as cationic analytes by CE–MS, for instance, by using a two-step derivatization approach for the analysis of carboxylic acid-containing metabolites [10], by using the reagent maleimide to derivatize thiols in plasma [11], and by using two-stage derivatization to label amines, hydroxyls, and carboxylates [12].

Recently, a novel derivatization reagent based on trimethylmethaneaminophenacetyl bromide (TmAmPBr) was presented, which following an optimized derivatization workflow provides a permanent positive charge to acidic metabolites [13], allowing their analysis by using exactly the same separation conditions as used for the analysis of basic metabolites. The method was used for the analysis of energy metabolism-related metabolites in limited numbers of mammalian cells. In this chapter, the main methodological aspects of the derivatization procedure are described in more detail and critical steps are highlighted. Aspects related to the synthesis of this derivation agent have been explained in [13] and therefore not considered in this chapter. Overall, we hope that the work presented here will further stimulate the use of chemical derivation in CE–MS-based metabolomics studies.

2 Materials

A Milli-Q Advantage A10 water purification system with a 0.22-μM pore-size filter is used to obtain pure water. All chemicals used are of analytical grade or higher purity.

2.1 Solutions and Samples for Analysis

1. Background electrolyte (BGE) solution: 10% (v/v) acetic acid, pH 2.2. Add approximately 80 mL of water in a 100-mL volumetric flask and add 10 mL of concentrated acetic acid. Add water to the gauge line. Mix the solution thoroughly. Store at 4 °C (*see* **Note 1**).
2. Sheath-liquid solution for CE–MS analysis: Mix 100 mL of water with 100 mL of isopropanol. Add 60 μL of concentrated acetic acid to this solution. Degas for 10 min in an ultrasonic bath before the first use.
3. Sodium hydroxide solution: 1 M sodium hydroxide. Add 1 g of sodium hydroxide (98%) into a 25-mL volumetric flask and add water up to the gauge line. Mix the solution thoroughly. Store at 4 °C.
4. Hydrochloric acid solutions: 1 M hydrochloric acid, 0.1 M hydrochloric acid. For 1 M hydrochloric acid solution, add 2.1 mL of hydrochloric acid (37%) into a 25-mL volumetric flask and add water up to the gauge line. Mix the solution thoroughly. For 0.1 M hydrochloric acid solution, add 2.5 mL of 1 M hydrochloric acid solution into a 25-mL volumetric flask and add water up to the gauge line. Mix the solution thoroughly. Store both solutions at 4 °C.
5. Ammonium hydroxide solution: 2.5% (v/v) ammonium hydroxide. Add 5 mL of ammonium hydroxide (25%) into a 50-mL volumetric flask and add water up to the gauge line. Mix the solution thoroughly. Store at 4 °C.
6. Ammonium carbonate solution: 65 mM ammonium carbonate. Add 62.5 mg of ammonium carbonate into a 10-mL volumetric flask and add water up to the gauge line. Mix the solution thoroughly. Adjust the pH to 10 using a 1 M sodium hydroxide solution. Check the pH with a pH meter. Store at 4 °C (*see* **Note 2**).
7. Metabolite standard mixture: Prepare stock solutions of 1 mg/mL for glutaric acid, malonic acid, pyruvic acid, α-ketoglutaric acid, fumaric acid, isocitric acid, succinic acid, malic acid, citric acid, lactic acid, and oxaloacetic acid in a mixture of dimethylformamide:dimethylsulfoxide (DMF:DMSO) (50:50, v/v) (*see* **Note 3**). Make aliquots of stock solutions by dilution with DMF:DMSO to obtain a working solution of 500 μM for each analyte. Store stock and working solutions at −80°.
8. Internal standard mixture: Prepare a stock solution of 1 mg/mL for [D6]-succinic acid in DMF:DMSO. Make aliquots of stock solutions by dilution with DMF:DMSO to obtain a working solution of 125 μM for each analyte. Store stock and working solutions at −80° when not in use.

2.2 Derivatization Reagent

1. TmAmPBr pure reagent: TmAmPBr was synthesized and kindly provided by J.P.D van Veldhoven, A. E. Christina, and D. van der Es. A detailed procedure for the synthesis of TmAmPBr is described in a previous publication [13]. Store TmAmPBr in powder form (*see* **Note 4**) at room temperature (20°), preferably in an air-free environment (*see* **Notes 5** and **6**).
2. TmAmPBr solution: 40 mg/mL TmAmPBr. Add 40 mg of TmAmPBr into a 2-mL Eppendorf vial and add 1 mL of DMF: DMSO (*see* **Note 7**). Mix the solution thoroughly. Store at 4 °C.

2.3 Mammalian (HepG2) Cells

1. Human liver cancer cells (HepG2) cells: HepG2 were cultured, harvested, and collected in-house. The HepG2 cells were collected in aliquots of 2×10^6 cells per vial in the Eppendorf vial and stored at −80 °C until sample preparation was performed. For a detailed protocol concerning this procedure, we would like to refer the readers to [14].

2.4 Analytical Equipment

1. CE separation: Fused-silica capillaries (50 μM I.D. × 70 cm total length) are used.
2. CE–MS: A 7100-capillary electrophoresis system from Agilent Technologies hyphenated with a 6230 TOF, also from Agilent.
3. CE–MS coupling: Coupling is realized via a coaxial sheath-liquid ESI interface equipped with a standard triple-tube sprayer.
4. Software: MassHunter version B.06.00 is used for data acquisition, instrument control, and data treatment. Peak integration is performed using extracted ion electropherograms (EIEs) based on the m/z for the protonated molecular ion $[M + H]^+$ of each metabolite. Excel is used for further data processing (*see* **Note 8**).

3 Methods

Please use all appropriate laboratory safety procedures, including safety glasses, lab coat, and gloves, when performing the experiments described in this protocol.

3.1 Mammalian (HepG2) Cell Lysate Sample Preparation

1. Thaw a HepG2 cell extract aliquot of 2×10^6 cells slowly on ice. Then take 25 μL of the lysed cells' extract and add 225 μL of cold methanol:water (80:20, v/v) so that the diluted mixture has a concentration of 25,000 cells per 10 μL (*see* **Note 9**).

2. Add 50 μL of methanol, 200 μL of water, and 250 μL of chloroform to the mixture to make a final volume of 750 μL in the ratio of methanol:water:chloroform of 1:1:1 (v/v/v) (*see* **Note 10**). Vortex for 30 s and centrifuge for 10 min at 16,100 × *g*.
3. Collect the methanol/water layer and pipette into a new 1.5-mL Eppendorf vial. Evaporate the sample to dryness using a SpeedVac and reconstitute the dry material in 10 μL of DMF/DMSO before derivatization.

3.2 Derivatization of Metabolite Standards and Mammalian (HepG2) Cell Extracts

1. Add 10 μL of the metabolite standard mixture into a 1.5-mL Eppendorf vial (*see* **Note 11**).
2. Add 10 μL of ammonium carbonate (65 mM, pH 10), and add 5 μL of internal standard solution (125 μM) (*see* **Notes 12** and **13**) and vortex for 30 s.
3. Add 20 μL of TmAmPBr (40 mg/mL) to the mixture, and vortex for 1 min.
4. Place the Eppendorf vial into a shaking incubator (*see* **Notes 14** and **15**) at maximum speed (900 RPM) (*see* **Note 16**) for 50 min at 90 °C.
5. After incubation, centrifuge for 5 min at 16,100 × *g* (*see* **Note 17**), and then add 10 μL of acetic acid (10% v/v, 1.7 M) to the vial to quench the reaction. Incubate for an additional 20 min of incubation under the same conditions (*see* **Note 18**).
6. Use the procedures described in Subheading 3.1, followed by **steps 1–5** described above for HepG2 cell extracts. The workflow for HepG2 cell lysate sample preparation and derivatization is summarized in Fig. 1.

3.3 Sample Analysis of Derivatized Metabolite Standards and Mammalian (HepG2) Cell Extracts Using CE–MS

1. Prior to first use, new capillaries installed in the coaxial sheath liquid interface are conditioned by subsequently rinsing, at 5 bar for 1 min, with methanol, water, 1 M sodium hydroxide, water, 1 M hydrochloric acid, water, 0.1 M hydrochloric acid, water, and BGE.
2. Add 10 μL of the derivatized metabolite standard mixture into an empty 250-μL microvial and put the vial in the inlet sample tray.
3. Inject the derivatized metabolite standard mixture using an injection volume of ~27 nL (24 s 50 mbar).
4. Inject a small plug of BGE using an injection volume of ~5 nL (5 s 50 mbar). Apply a separation voltage of 30 kV.
5. Start acquiring MS data in the m/z range from 50 to 1000 by using an ESI voltage of −5.5 kV.

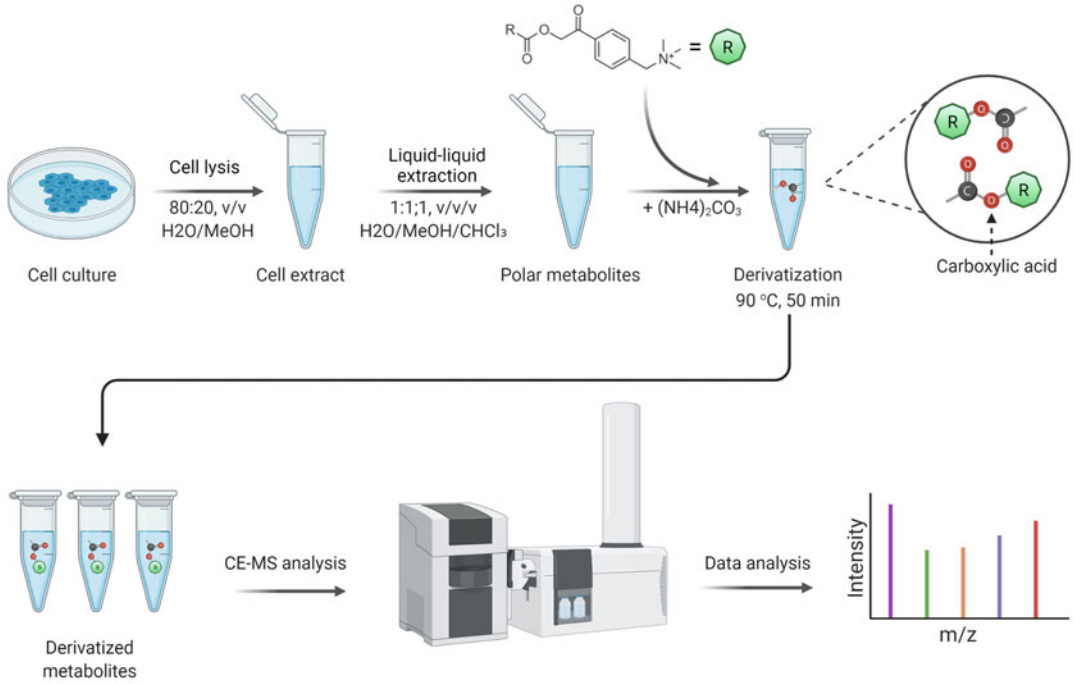

Fig. 1 An overview of the analytical workflow, including sample preparation of the HepG2 cell extracts, derivatization, and CE–MS sample analysis. (Created with BioRender.com)

6. A typical profile for standards that have been derivatized by TmAmPBr and analyzed by CE–MS is shown in Fig. 2 (*see* **Note 19**).
7. Between sample injections, rinse the capillary with milli-Q, methanol, milli-Q, ammonium hydroxide (2.5%), milli-Q, and BGE, each at 5 bar for 60 s.
8. Repeat **steps 2–4** for a derivatized HepG2 cell extract sample. Figure 3 depicts the electropherograms corresponding to the derivatized anionic metabolites extracted from a HepG2 cell extract corresponding to 25,000 cells.

4 Notes

1. Prepare fresh BGE weekly.
2. Prepare a fresh ammonium carbonate solution weekly.
3. A solution of dimethylformamide:dimethylsulfoxide (DMF:DMSO) is used in this protocol for two reasons: (1) to ensure dissolvation of metabolites with different logP values, and (2) to ensure that no unnecessary additional water is added to the derivatization reaction.

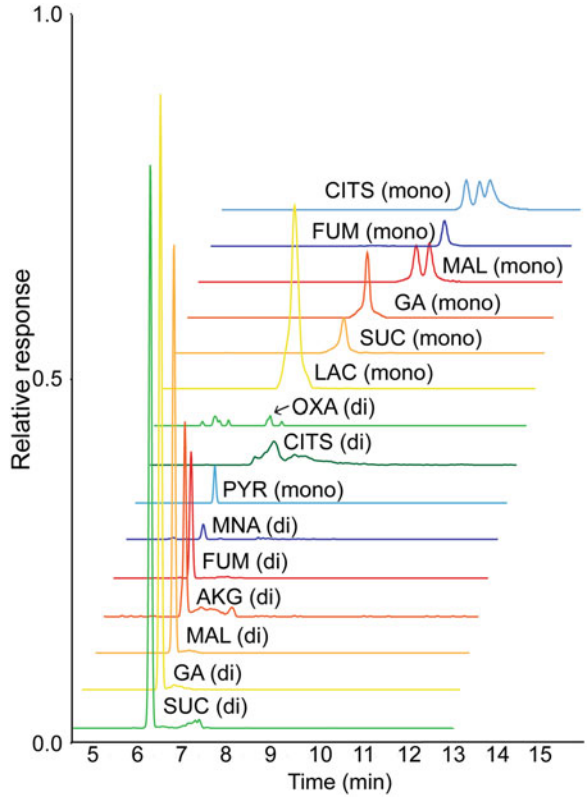

Fig. 2 Extracted-ion electropherogram obtained by CE–MS from the analysis of 12 metabolites (90 μM) after derivatization with TmAmPBr. Separation conditions: BGE, 10% (v/v) acetic acid (1.7 M); sample injection volume 27 nL. (Reproduced with permission from Ref. [13])

4. The stability of the new TmAmPBr reagent over time is not yet studied in detail. Using standards, it was observed that the reactivity of the reagent was acceptable for at least 4 weeks. Therefore, synthesize fresh reagent monthly.
5. Store TmAmPBr in powder form in a water-free environment, and use parafilm to close the container tightly. TmAmPBr is prone to hydrolysis, so contact with water needs to be prevented in order to preserve the integrity of this reagent for derivatization reactions.
6. To prevent hydrolysis of the reagent, prevent the frequent and long opening of the container with TmAmPBr.
7. To prevent contact and potential reaction with water, DMF/DMSO is used as a solvent.
8. From the sheath-liquid, isopropanol ($C_3H_8OH^+$) and its clusters ($[(C_3H_8O)_2 + H]^+$ and ($[(C_3H_8O)_3 + H]^+$)) with corresponding *m/z* values of 61.06479, 121.12231, and

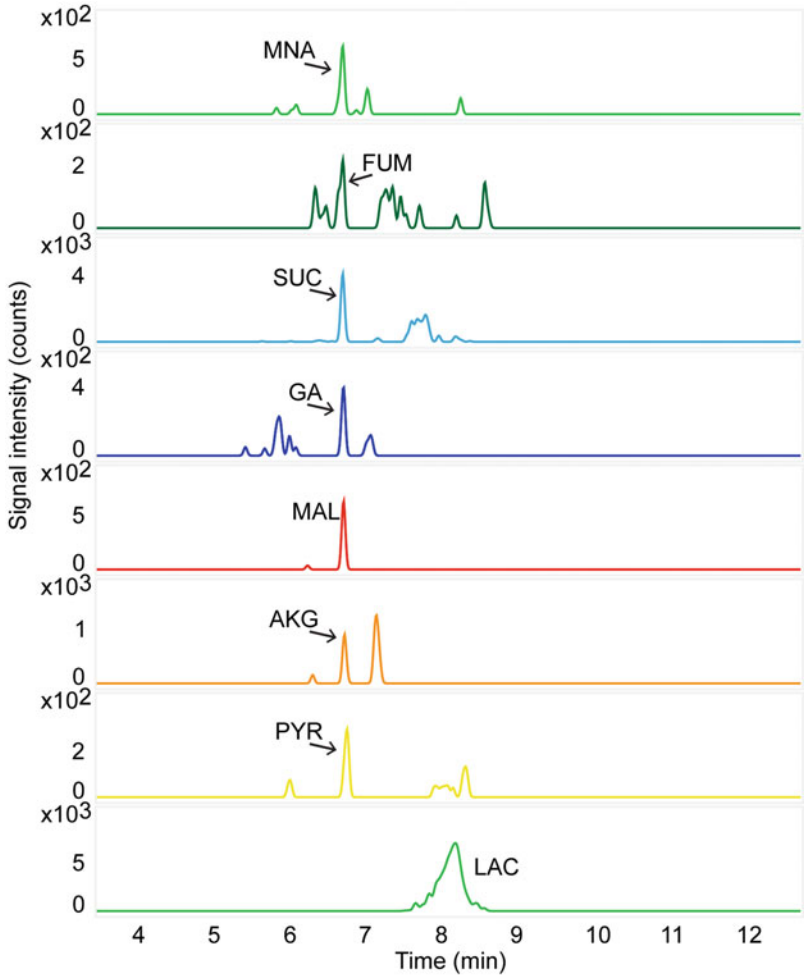

Fig. 3 Extracted-ion electropherograms obtained by CE–MS from the analysis of 25,000 HepG2 cell lysates after derivatization with TmAmPBr. Separation conditions: BGE, 10% (v/v) acetic acid (1.7 M); sample injection volume 27 nL. (Reproduced with permission from Ref. [13])

181.17982, respectively, are used as lock masses to correct for potential mass shifts.

9. The dilution can be altered to acquire the desired cell concentration.
10. A liquid–liquid extraction using water, methanol, and chloroform is used to remove proteins and lipids from the sample matrix. When also removing lipids, the CE–MS migration time repeatability is higher than without removing lipid content.
11. Always use safe-lock Eppendorf vials during this protocol.
12. When internal standards are added at this stage, only the metabolite response during derivatization and CE–MS analysis is corrected by the internal standards response. If it is desired to

also correct metabolite responses during sample preparation, already add internal standards during Subheading 3.1, **step 1**.

13. To ensure more reliable quantification, separate stable isotope labeled internal standards need to be considered for every metabolite.
14. Before placing the Eppendorf vials into the shaking incubator, make sure the vials are tightly closed. Otherwise, sample might be lost due to evaporation during high-temperature incubation.
15. If it is possible, use a shaking incubator where the Eppendorf vials are completely encapsulated and heated. In an open Eppendorf shaker, heat could be inconsistently divided, leading to inconsistent derivatization within a sample set.
16. If the Eppendorf shaker can shake at higher RPM, this can also be increased.
17. The centrifugation step is included to make sure all evaporated (and then condensed after removing from the incubator) sample is at the bottom of the Eppendorf vial before undertaking further steps.
18. The additional incubation is necessary to completely quench the derivatization reaction.
19. If more resolution between metabolites is desired, use a longer capillary length.

Acknowledgments

M. van Mever and Dr. R. Ramautar acknowledge the financial support of the Vidi grant scheme of the Netherlands Organization of Scientific Research (NWO Vidi 723.016.003).

References

1. Garcia A, Godzien J, Lopez-Gonzalvez A, Barbas C (2017) Capillary electrophoresis mass spectrometry as a tool for untargeted metabolomics. Bioanalysis 9(1):99–130
2. Drouin N, Ramautar R (2021) Capillary electrophoresis-mass spectrometry for metabolomics: possibilities and perspectives. In: Colnaghi Simionato AV (ed) Separation techniques applied to omics sciences: from principles to relevant applications. Springer, Cham, pp 159–178
3. Harada S, Hirayama A, Chan Q, Kurihara A, Fukai K, Iida M et al (2018) Reliability of plasma polar metabolite concentrations in a large-scale cohort study using capillary electrophoresis-mass spectrometry. PLoS One 13(1):e0191230
4. Drouin N, van Mever M, Zhang W, Tobolkina E, Ferre S, Servais A-C et al (2020) Capillary electrophoresis-mass spectrometry at trial by metabo-ring: effective electrophoretic mobility for reproducible and robust compound annotation. Anal Chem 92(20): 14103–14112
5. Soga T, Igarashi K, Ito C, Mizobuchi K, Zimmermann H-P (2009) Tomita M. Metabolomic profiling of anionic metabolites by capillary electrophoresis mass spectrometry. Anal Chem 81(15):6165
6. van Mever M, Hankemeier T, Ramautar R (2019) CE-MS for anionic metabolic profiling: an overview of methodological developments. Electrophoresis 40(18–19):2349–2359

7. Lee J-H, Kim S-J, Lee S, Rhee J-K, Lee SY, Na Y-C (2019) Saturated fatty acid determination method using paired ion electrospray ionization mass spectrometry coupled with capillary electrophoresis. Anal Chim Acta 984:223–231
8. Zhang W, Guled F, Hankemeier T, Ramautar R (2020) Profiling nucleotides in low numbers of mammalian cells by sheathless CE-MS in positive ion mode: circumventing corona discharge. Electrophoresis 41(5–6):360–369
9. Drouin N, Pezzatti J, Gagnebin Y, González-Ruiz V, Schappler J, Rudaz S (2018) Effective mobility as a robust criterion for compound annotation and identification in metabolomics: toward a mobility-based library. Anal Chim Acta 1032:178–187
10. Yang WC, Regnier FE, Adamec J (2008) Comparative metabolite profiling of carboxylic acids in rat urine by CE-ESI MS/MS through positively pre-charged and 2H-coded derivatization. Electrophoresis 29(22):4549–4560
11. D'Agostino LA, Lam KP, Lee R, Britz-McKibbin P (2011) Comprehensive plasma thiol redox status determination for metabolomics. J Proteome Res 10(2):592–603
12. Huang T, Armbruster M, Lee R, Hui DS, Edwards JL (2018) Metabolomic analysis of mammalian cells and human tissue through one-pot two stage derivatizations using sheathless capillary electrophoresis-electrospray ionization-mass spectrometry. J Chrom A 1567: 219–225
13. van Mever M, Willacey CCW, Zhang W, Drouin N, Christina AE, Lindenburg PW et al (2021) Profiling acidic metabolites by capillary electrophoresis-mass spectrometry in low numbers of mammalian cells using a novel chemical derivatization approach. Anal Sci Adv:1–11
14. Zhang W, Hankemeier T, Ramautar R (2019) Capillary electrophoresis-mass spectrometry for metabolic profiling of biomass-limited samples. Methods Mol Biol 1972:165–172

Chapter 11

Untargeted Metabolomics Based on Liquid Chromatography–Mass Spectrometry for the Analysis of Plasma and Erythrocyte Samples in Childhood Obesity

Álvaro González-Domínguez, Marina Armeni, Otto Savolainen, Alfonso María Lechuga-Sancho, Rikard Landberg, and Raúl González-Domínguez

Abstract

The circulating metabolome of human peripheral blood provides valuable information to investigate the molecular mechanisms underlying the development of diseases and to discover candidate biomarkers. In particular, erythrocytes have been proposed as potential systemic indicators of the metabolic and redox status of the organism. To accomplish wide-coverage metabolomics analysis, the combination of complementary analytical techniques is necessary to manage the physicochemical complexity of the human metabolome. Herein, we describe an untargeted metabolomics method to capture the plasmatic and erythroid metabolomes based on ultrahigh-performance liquid chromatography coupled to high-resolution mass spectrometry, combining reversed-phase liquid chromatography and hydrophilic interaction liquid chromatography. The method provides comprehensive metabolomics fingerprinting of plasma and erythrocyte samples, thereby enabling the elucidation of the distinctive metabolic disturbances behind childhood obesity and associated comorbidities, such as insulin resistance.

Key words Metabolomics, Mass spectrometry, Liquid chromatography, Plasma, Erythrocyte, Childhood obesity

1 Introduction

Obesity has become a major problem for public health in the last years due to its growing prevalence in both adult and child populations. The rise in obesity is paralleled by an increased risk of developing other noncommunicable diseases, subsequently leading to higher mortality rates [1]. In particular, childhood obesity has been associated with a greater predisposition to develop a myriad of diseases during adulthood, such as type 2 diabetes mellitus, cardiovascular diseases (e.g., hypertension), and other metabolic disorders (e.g., metabolic syndrome) [2]. The characteristic pathogenic

Raúl González-Domínguez (ed.), *Mass Spectrometry for Metabolomics*,
Methods in Molecular Biology, vol. 2571, https://doi.org/10.1007/978-1-0716-2699-3_11,
© The Author(s), under exclusive license to Springer Science+Business Media, LLC, part of Springer Nature 2023

events behind obesity are usually related to various metabolic comorbidities, being insulin resistance (IR) one of the most relevant expressions of these alterations. However, it should be noted here that some obese individuals (ca. 10–30%) are not affected by metabolic disturbances and paradoxically present low cardiometabolic risk, even showing the same severity of weight excess, constituting the so-called "metabolically healthy obesity" [3]. Within this tangled meshwork, the specific molecular mechanisms that underlie the development of obesity and its comorbidities are still largely unknown, so there is a great need to discover potential biomarkers for early detection in the pediatric age, as well as for monitoring their progression.

Metabolomics can be defined as the simultaneous and holistic analysis of the metabolites that are present in a biological sample (i.e., the metabolome). As metabolites are read-outs from up-stream events in the central dogma that are directly involved in the central metabolic processes of all living systems, metabolomics stands out as a powerful tool in biomedicine to elucidate the molecular mechanisms underlying the pathogenesis and progression of diseases, as well as to study the efficacy of pharmacological and nutritional treatments [4]. Nowadays, mass spectrometry has become the most widely employed analytical tool in metabolomics research because of its sensitivity, selectivity, and broad coverage, thanks to the availability of multiple instrumental configurations, such as direct infusion analysis [5, 6] or the coupling with orthogonal separation systems (e.g., liquid or gas chromatography, capillary electrophoresis) [7]. However, the combination of various complementary techniques is usually required to deal with the high physicochemical complexity of the human metabolome [8, 9]. In this respect, the potential of metabolomics for the investigation of obesity and associated comorbidities is nowadays beyond any doubt [10, 11], although its application in the field of childhood obesity is not yet widespread. Of note, many of the metabolomics-based studies previously published on childhood obesity relied on targeted approaches, which focus on a predetermined set of metabolites [12–16]. On the other hand, some authors have also proposed the application of nontargeted metabolomics based on mass spectrometry for the characterization of different biological matrices, such as blood [17–19] or urine [19, 20]. Metabolomics has also been employed to investigate the metabolic changes provoked by an oral glucose tolerance test (OGTT). Interestingly, great similarities have been described between the characteristic metabolic disturbances detected in patients affected by insulin resistance and the OGTT-associated metabolomics profiles [21–25], suggesting the excellent utility of acute challenge trials to decipher the pathological mechanisms underlying this metabolic disorder [26].

This chapter details methods and tips for comprehensive metabolomics analysis of plasma and erythrocyte samples by applying an untargeted multi-platform based on ultrahigh-performance liquid chromatography coupled to high-resolution mass spectrometry (UHPLC-HRMS) and the combination of complementary chromatographic approaches, namely reversed-phase liquid chromatography (RPLC) and hydrophilic interaction liquid chromatography (HILIC).

2 Materials

High purity reagents and solvents (e.g., MS grade) must be used throughout the entire analytical pipeline.

2.1 Sample Collection and Preparation

1. BD Vacutainer tubes with EDTA K2 and Advance vacuum system.
2. Glucose syrup: 500 g L^{-1} of glucose.
3. Saline solution: 9 g L^{-1} of sodium chloride.
4. Extraction solvent for plasma: 100% acetonitrile.
5. Extraction solvent for erythrocytes: 50% methanol, 30% acetonitrile, 20% water (v:v:v).

2.2 Metabolomics Analysis

1. UHPLC instrumentation: Agilent 1290 II UHPLC system (Agilent Technologies, Santa Clara, CA).
2. QTOF-MS instrumentation: Agilent 6550 iFunnel QTOF-MS system equipped with electrospray ion (ESI) source (dual Agilent Jet Stream Source, AJS) (Agilent Technologies, Santa Clara, CA).
3. Reversed-phase column: Waters Acquity UPLC HSS T3, 100 × 2.1 mm, 1.8 μm (Waters, Milford, MA).
4. HILIC column: Waters Acquity UPLC BEH Amide NH2, 100 × 2.1 mm, 1.7 μm (Waters, Milford, MA).
5. Mobile phase A for RPLC: 0.04% formic acid in water (v:v).
6. Mobile phase B for RPLC: 0.04% formic acid in methanol (v:v).
7. Needle wash solvent for RPLC: 5% methanol (v:v).
8. Mobile phase A for HILIC: 10 mM ammonium formate in water.
9. Mobile phase B for HILIC: 90% acetonitrile, 10% 100 mM ammonium formate in water (v:v).
10. Needle wash solvent for HILIC: 90% acetonitrile (v:v).

3 Methods

3.1 Sample Collection

1. Collect blood samples at different time points along an oral glucose tolerance test (OGTT) from normal weight and obese children (*see* **Note 1**), as follows:
   (a) Insert a catheter into the antecubital vein.
   (b) Collect a baseline blood sample, after at least 8 h of fasting, using a BD Vacutainer tube with EDTA K2 and Advance vacuum system.
   (c) Ask the participant to drink 150 mL of the glucose syrup (i.e., 75 g of glucose) within 5 min.
   (d) Collect blood samples after 30, 60, 90, and 120 min of the OGTT as detailed under Subheading 3.1, **step 1.b** (*see* **Note 2**).
2. After blood collection at the different time points, gently invert the tubes 8–10 times.
3. Centrifuge blood samples at 1500 *g* for 10 min at 4 °C.
4. Aliquot the supernatant plasma into cryotubes and store them at −80 °C until analysis (*see* **Note 3**).
5. Add 5 mL of cold saline solution (4 °C) to the Vacutainer tube containing the pelleted blood cells.
6. Gently invert the tube 8–10 times.
7. Centrifuge at 1500 *g* for 5 min at 4 °C and discard the supernatant.
8. Repeat **steps 5**–7 two more times.
9. Aliquot the erythrocyte fraction into cryotubes and store them at −80 °C until analysis (*see* **Note 4**).

3.2 Extraction of Plasma Samples

1. Thaw plasma samples at 4 °C (i.e., in a refrigerator or in an ice bath).
2. Vortex the sample tubes and aliquot 30 μL into a new Eppendorf tube (*see* **Note 5**).
3. Add 200 μL of cold acetonitrile (4 °C).
4. Vortex for 3 min at 1000 rpm using an orbital shaker, preferably in cold.
5. Centrifuge at 10000 *g* for 10 min at 4 °C.
6. Transfer the supernatant to LC injection vials (*see* **Note 6**).
7. Keep the sample extracts at −80 °C until analysis.

3.3 Extraction of Erythrocyte Samples

1. Thaw erythrocyte samples at 4 °C (i.e., in a refrigerator or in an ice bath).
2. Vortex the sample tubes and aliquot 20 μL into a new Eppendorf tube (*see* **Note 5**).
3. Add 180 μL of cold extraction solvent (methanol:acetonitrile: water 5:3:2, at 4 °C).
4. Vortex for 30 min at 1000 rpm using an orbital shaker in cold.
5. Centrifuge at 10000 *g* for 10 min at 4 °C.
6. Transfer the supernatant to LC injection vials (*see* **Note 6**).
7. Keep the sample extracts at −80 °C until analysis.

3.4 Metabolomics Analysis by UHPLC–MS

1. Before starting the analysis, clean the ESI source (i.e., spray shield, nebulizer) using a mixture of isopropanol, water, and formic acid.
2. Turn the QTOF-MS instrument on and let it equilibrate for 30 min.
3. Tune the instrument to check the MS resolution and sensitivity (*see* **Note 7**).
4. Flush the UHPLC system with a freshly prepared mobile phase until the pressure stabilizes.
5. Set the UHPLC operating conditions as follows:
   (a) Injection volume: 2 μL in the positive ion mode, 4 μL in the negative ion mode.
   (b) Multisampler temperature: 4 °C.
   (c) Column compartment temperature: 45 °C.
   (d) Flow rate: 0.4 mL min^{-1}.
   (e) Gradient program for RPLC: 0–6 min, 5–100% B; 6–10.5 min, 100% B; 10.5–10.51 min, 100–5% B; 10.51–13 min, 5% B.
   (f) Gradient program for HILIC: 0–1 min, 100% B; 1–8 min, 100–70% B; 8–8.1 min, 70–100% B; 8.1–11 min, 100% B.
6. Set the QTOF-MS operating conditions as follows:
   (a) Acquisition mode: full scan within the m/z range 50–1700 Da (centroid mode).
   (b) Ion polarity: positive and negative ionization modes in separate runs.
   (c) Acquisition rate: 1.67 spectra s^{-1}.
   (d) Gas temperature: 175 °C.
   (e) Gas flow: 12 L min^{-1}.
   (f) Nebulizer pressure: 45 psi.

(g) Sheath gas temperature: 350 °C for RPLC, 375 °C for HILIC.

(h) Sheath gas flow: 11 L min^{-1}.

(i) Capillary voltage: 3500 V.

(j) Nozzle voltage: 300 V.

(k) Fragmentor voltage: 175 V.

7. Analyze the samples in random order and intercalate quality control samples across the entire sequence to monitor the system stability (e.g., one quality control every 10 samples) (*see* **Note 8**).

4 Notes

1. Childhood obesity is defined as presenting a body mass index that surpasses two times the standard deviation in the reference population, adjusted for sex and age [27].
2. The measurement of blood insulin levels along the OGTT allows diagnosing insulin resistance when at least one of the following criteria is fulfilled: (i) homeostasis model assessment of insulin resistance (HOMA-IR) score above 3.5, (ii) fasting insulin levels above 15 μUI mL^{-1}, (iii) insulin levels above 75 μUI mL^{-1} at 120 min of the OGTT, and (iv) insulin levels above 150 μUI mL^{-1} at any time point of the OGTT curve [28].
3. Discard hemolyzed samples, as this may strongly alter the circulating metabolomics profile.
4. The pre-analytical phase (i.e., sample collection, pre-processing) must be standardized to minimize inter-sample variability factors and to maintain the metabolic integrity of the biological samples with the aim of ensuring that metabolomics profiles are a direct expression of the in vivo biochemical status [29].
5. Sample treatment can also be performed in 96-well plates, which facilitates the automatization of the extraction process.
6. For each study, use solvents and labware from the same vendor and lot along the entire sample preparation process to minimize inter-sample variability (e.g., background noise, artifact peaks due to contaminations).
7. We use the API-TOF reference mass solution kit from Agilent Technologies, but the reader can opt for other solutions and suppliers depending on the instrumentation to be used.
8. Quality control samples can be prepared by mixing equal aliquots of each of the samples under study.

Acknowledgments

AGD is supported by an intramural grant from the Biomedical Research and Innovation Institute of Cádiz (LII19/16IN-CO24), and RGD is a recipient of a "Miguel Servet" fellowship (CP21/00120) funded by Instituto de Salud Carlos III.

References

1. Ludwig DS (2018) Epidemic childhood obesity: not yet the end of the beginning. Pediatrics 141:e20174078. https://doi.org/10.1542/peds.2017-4078
2. Hossain P, Kawar B, El Nahas M (2007) Obesity and diabetes in the developing world – a growing challenge. N Engl J Med 356: 213–215. https://doi.org/10.1056/NEJMp068177
3. Phillips CM, Dillon C, Harrington JM, McCarthy VJC, Kearney PM, Fitzgerald AP, Perry IJ (2013) Defining metabolically healthy obesity: role of dietary and lifestyle factors. PLoS One 8:e76188. https://doi.org/10.1371/journal.pone.0076188
4. Johnson CH, Ivanisevic J, Siuzdak G (2016) Metabolomics: beyond biomarkers and towards mechanisms. Nat Rev Mol Cell Biol 17:451–459. https://doi.org/10.1038/nrm.2016.25
5. González-Domínguez R, Sayago A, Fernández-Recamales Á (2017) Direct infusion mass spectrometry for metabolomic phenotyping of diseases. Bioanalysis 9:131–148. https://doi.org/10.4155/bio-2016-0202
6. González-Domínguez R, González-Domínguez Á, Segundo C, Schwarz M, Sayago A, Mateos RM, Durán-Guerrero E, Lechuga-Sancho AM, Fernández-Recamales Á (2019) High-throughput metabolomics based on direct mass spectrometry analysis in biomedical research. In: D'Alessandro A (ed) High-throughput metabolomics, Methods in molecular biology, vol 1978. Humana Press, New York. https://doi.org/10.1007/978-1-4939-9236-2_3
7. Kuehnbaum NL, Britz-McKibbin P (2013) New advances in separation science for metabolomics: resolving chemical diversity in a post-genomic era. Chem Rev 113: 2437–2468. https://doi.org/10.1021/cr300484s
8. González-Domínguez Á, Durán-Guerrero E, Fernández-Recamales Á, Lechuga-Sancho AM, Sayago A, Schwarz M, Segundo C, González-Domínguez R (2017) An overview on the importance of combining complementary analytical platforms in metabolomic research. Curr Top Med Chem 17:3289–3295. https://doi.org/10.2174/1568026618666171211144918
9. González-Domínguez R, González-Domínguez Á, Sayago A, Fernández-Recamales Á (2018) Mass spectrometry-based metabolomic multiplatform for Alzheimer's disease research. In: Perneczky R (ed) Biomarkers for Alzheimer's disease drug development, Methods in molecular biology, vol 1750. Humana Press, New York. https://doi.org/10.1007/978-1-4939-7704-8_8
10. Rauschert S, Uhl O, Koletzko B, Hellmuth C (2014) Metabolomic biomarkers for obesity in humans: a short review. Ann Nutr Metab 64: 314–324. https://doi.org/10.1159/000365040
11. Zhang A, Sun H, Wang X (2017) Emerging role and recent applications of metabolomics biomarkers in obesity disease research. RSC Adv 7:14966–14973. https://doi.org/10.1039/C6RA28715H
12. Wahl S, Yu Z, Kleber M, Singmann P, Holzapfel C, He Y, Mittelstrass K, Polonikov A, Prehn C, Römisch-Margl W, Adamski J, Suhre K, Grallert H, Illig T, Wang-Sattler R, Reinehr T (2012) Childhood obesity is associated with changes in the serum metabolite profile. Obes Facts 5:660–670. https://doi.org/10.1159/000343204
13. Englich B, Herberth G, Rolle-Kampczyk U, Trump S, Röder S, Borte M, Stangl GI, von Bergen M, Lehmann I, Junge KM (2017) Maternal cytokine status may prime the metabolic profile and increase risk of obesity in children. Int J Obes 41:1440–1446. https://doi.org/10.1038/ijo.2017.113
14. Lee A, Jang HB, Ra M, Choi Y, Lee HJ, Park JY, Kang JH, Park KH, Park SI, Song J (2015) Prediction of future risk of insulin resistance and metabolic syndrome based on Korean boy's metabolite profiling. Obes Res Clin Pract 9:336–345. https://doi.org/10.1016/j.orcp.2014.10.220
15. Hellmuth C, Kirchberg FF, Lass N, Harder U, Peissner W, Koletzko B, Reinehr T (2016)

Tyrosine is associated with insulin resistance in longitudinal metabolomic profiling of obese children. J Diabetes Res 2016:2108909. https://doi.org/10.1155/2016/2108909

16. Ellul S, Ponsonby AL, Carlin JB, Collier F, Mansell T, Vuillermin P, Burgner D, Saffery R, Barwon Infant Study Investigator Team (2020) Sex differences in infant blood metabolite profile in association with weight and adiposity measures. Pediatr Res 88:473–483. https://doi.org/10.1038/s41390-020-0762-4
17. Szczerbinski L, Wojciechowska G, Olichwier A, Taylor MA, Puchta U, Konopka P, Paszko A, Citko A, Goscik J, Fiehn O, Fan S, Wasilewska A, Taranta-Janusz K, Kretowski A (2022) Untargeted metabolomics analysis of the serum metabolic signature of childhood obesity. Nutrients 14:214. https://doi.org/10.3390/nu14010214
18. Perng W, Gillman MW, Fleisch AF, Michalek RD, Watkins SM, Isganaitis E, Patti ME, Oken E (2014) Metabolomic profiles and childhood obesity. Obesity 22:2570–2578. https://doi.org/10.1002/oby.20901
19. Gallego-Paüls M, Hernández-Ferrer C, Bustamante M, Basagaña X, Barrera-Gómez J, Lau CHE, Siskos AP, Vives-Usano M, Ruiz-Arenas C, Wright J, Slama R, Heude B, Casas M, Grazuleviciene R, Chatzi L, Borràs E, Sabidó E, Carracedo Á, Estivill X, Urquiza J, Coen M, Keun HC, González JR, Vrijheid M, Maitre L (2021) Variability of multi-omics profiles in a population-based child cohort. BMC Med 19:166. https://doi.org/10.1186/s12916-021-02027-z
20. Cho K, Moon J, Kang JH, Jang HB, Lee HJ, Park SI, Yu KS, Cho JY (2017) Combined untargeted and targeted metabolomic profiling reveals urinary biomarkers for discriminating obese from normal-weight adolescents. Pediatr Obes 12:93–101. https://doi.org/10.1111/ijpo.12114
21. Shaham O, Wei R, Wang TJ, Ricciardi C, Lewis GD, Vasan RS, Carr SA, Thadhani R, Gerszten RE, Mootha VK (2008) Metabolic profiling of the human response to a glucose challenge reveals distinct axes of insulin sensitivity. Mol Syst Biol 4:1–9. https://doi.org/10.1038/msb.2008.50
22. Rämö JT, Kaye SM, Jukarainen S, Bogl LH, Hakkarainen A, Lundbom J, Lundbom N, Rissanen A, Kaprio J, Matikainen N, Pietiläinen KH (2017) Liver fat and insulin sensitivity define metabolite profiles during a glucose tolerance test in young adult twins. J Clin Endocrinol Metab 102:220–231. https://doi.org/10.1210/jc.2015-3512
23. Ho JE, Larson MG, Vasan RS, Ghorbani A, Cheng S, Rhee EP, Florez JC, Clish CB, Gerszten RE, Wang TJ (2013) Metabolite profiles during oral glucose challenge. Diabetes 62: 2689–2698. https://doi.org/10.2337/db12-0754
24. González-Domínguez R, Mateos RM, Lechuga-Sancho AM, González-Cortés JJ, Corrales-Cuevas M, Rojas-Cots JA, Segundo C, Schwarz M (2017) Synergic effects of sugar and caffeine on insulin-mediated metabolomic alterations after an acute consumption of soft drinks. Electrophoresis 38: 2313–2322. https://doi.org/10.1002/elps.201700044
25. Zhao X, Peter A, Fritsche J, Elcnerova M, Fritsche A, Haring HU, Schleicher ED, Xu G, Lehmann R (2008) Changes of the plasma metabolome during an oral glucose tolerance test: is there more than glucose to look at? AJP Endocrinol Metab 296: E384–E393. https://doi.org/10.1152/ajpendo.90748.2008
26. Gonzalez-Dominguez A, Lechuga-Sancho AM, Gonzalez-Dominguez R (2018) Intervention and observational trials are complementary in metabolomics: diabetes and the oral glucose tolerance test. Curr Top Med Chem 18:896–900. https://doi.org/10.2174/1568026618666180711150525
27. Hernández M, Castellet J, Narvaiza JL, Rincón JM, Ruiz I, Sánchez E, Sobradillo B, Zurimendi A (1988) Curvas y tablas de crecimiento. Instituto de Investigación sobre Crecimiento y Desarrollo, Fundación Faustino Orbegozo. Editorial Garsi, Madrid
28. American Diabetes Association (2019) Standards of Medical Care in Diabetes – 2019 abridged for primary care providers. Clin Diabetes 37:11–34
29. González-Domínguez R, González-Domínguez Á, Sayago A, Fernández-Recamales Á (2020) Recommendations and best practices for standardizing the pre-analytical processing of blood and urine samples in metabolomics. Metabolites 10:229. https://doi.org/10.3390/metabo10060229

Chapter 12

Characterization of the Plasmatic and Erythroid Multielemental Biodistribution in Childhood Obesity Using a High-Throughput Method for Size Fractionation of Metal Species

Álvaro González-Domínguez, María Millán-Martínez, Daniel Sánchez-Rodas, Alfonso María Lechuga-Sancho, and Raúl González-Domínguez

Abstract

In this chapter, we describe a metallomics method based on protein precipitation under non-denaturing conditions and further analysis by inductively coupled plasma mass spectrometry for high-throughput metal speciation in plasma and erythrocyte samples. This methodology enables to study the total multielemental profile of these biological matrices, as well as to quantify the metal fractions conforming the metallometabolome and the metalloproteome. Furthermore, the analytical coverage comprises several essential and toxic metal elements, namely aluminum, arsenic, cadmium, cobalt, chromium, copper, iron, lithium, manganese, molybdenum, nickel, lead, selenium, vanadium, and zinc. Altogether, the metallomics method here proposed represents an excellent approach to comprehensively characterize the metal biodistribution in human peripheral blood, which would enable to decipher the role of metal homeostasis in health and disease, and particularly in childhood obesity.

Key words Metals, Size fractionation, Childhood obesity, Erythrocytes, Plasma, ICP-MS

1 Introduction

Metal elements can be present in living organisms in different chemical forms, mainly as free labile ions, forming complexes with low-molecular-weight metabolites (e.g., organic acids, amino acids), or acting as enzymatic cofactors. Furthermore, metalloids (e.g., selenium, arsenic) can also be covalently incorporated into biomolecules to yield a myriad of metallometabolites and metalloproteins. Accordingly, the comprehensive characterization of the metal species that conform a biological system would require the integration of different analytical approaches, including

Raúl González-Domínguez (ed.), *Mass Spectrometry for Metabolomics*,
Methods in Molecular Biology, vol. 2571, https://doi.org/10.1007/978-1-0716-2699-3_12,
© The Author(s), under exclusive license to Springer Science+Business Media, LLC, part of Springer Nature 2023

metabolomics, proteomics, and ionomics. In this vein, the International Union of Pure and Applied Chemistry (IUPAC) defines metallomics as "the study of the metallome, interactions, and functional connections of metal ions and other metal species with genes, proteins, metabolites, and other biomolecules in biological systems" [1]. Metallomics is a discipline of great relevance in biomedical research because of the pivotal role that metal and metalloid elements play in the correct functioning of essential biological processes. For instance, it should be noted that approximately a half of the existing enzymes are metal-dependent, which participate in a wide range of physiological processes such as electron transportation, structural support, immune response, synthesis of fatty acids, proteins, RNA and DNA, oxygen transport, cell division, antioxidant defense, and many others [2, 3]. Essential metals are also necessary for correct hormonal functioning, being involved in hormone synthesis, maturation, secretion, and receptor binding [4]. Nevertheless, failures in metal homeostasis may also lead to excessive production of reactive oxygen and nitrogen species, which in turn may cause oxidative damage in DNA, proteins, and lipids [5]. This metal-induced oxidative stress has been linked to a higher risk of developing several diseases, including neurological disorders, diabetes, and cardiovascular diseases [6]. Similarly, the exposure to high metal levels and imbalances in the metabolism of essential metals are also closely related to cancer, asthma, cardiovascular, neurodegenerative, and metabolic diseases [7, 8].

In this respect, growing evidence supports a central role of metal homeostasis in the development of obesity and obesity-related comorbidities. Obesity is defined as an excessive fat accumulation with detrimental effects on health. Currently, obesity has reached epidemic importance, with more than 4 million people dying each year because of suffering from it. In particular, the prevalence of obesity and overweight among children and adolescents has grown from a 4% to a 18% between 1975 and 2016, and around 38.2 million children under the age of 5 were overweight or obese according to World Health Organization data from 2019 [9, 10]. However, the etiology of obesity is still not clear today since it is the result of an interplay between a complex set of risk factors, including the environment, genetic, perinatal, dietary, or psychosocial factors. Furthermore, obesity is related to the development of a wide range of comorbidities, such as arterial hypertension, insulin resistance (IR), metabolic syndrome (MetS), polycystic ovarian syndrome, psychological problems, or type 2 diabetes mellitus [11, 12]. Within this tangled meshwork, metals have been demonstrated to participate in various pathological processes underlying obesity and related metabolic disorders. Chromium and vanadium take part in glucose and lipid metabolism, as well as in the regulation of insulin levels [13]. As recently reviewed, iron metabolism is altered at different levels in obesity and metabolic

syndrome, being crucial together with copper for a correct functioning of antioxidant systems, whose malfunction is tightly related to childhood obesity and related comorbidities [14, 15]. Moreover, cadmium and mercury exposures have also been related to an increased risk of diabetes mellitus and insulin resistance [16, 17]. However, it should be noted here that the available literature on the application of metallomics to study obesity relies on the determination of the total metal content in different biological matrices. A deeper characterization of the metallomics profiles would provide a more accurate overview about their involvement in obesity pathogenesis since the chemical form in which metals are present in the organism has a great impact on their toxicological properties, biological activity, and mobility across biological compartments [18].

To identify and quantify different metal species in a biological system, metallomics studies have traditionally been based on the combination of a separation technique (e.g., chromatography, electrophoresis) with sensitive and selective detection methods, mostly inductively coupled plasma mass spectrometry (ICP-MS). As no single tool can effectively resolve the whole metallome from a biological sample, the separation of metal species is typically accomplished by using multidimensional approaches. To this end, various orthogonal separation procedures with complementary retention mechanisms are employed, thereby allowing a substantial expansion of the analytical coverage [19]. Polyacrylamide gel electrophoresis has widely been employed in proteomics, but special care must be taken to ensure non-denaturing analytical conditions to avoid the dissociation of metal–protein coordination complexes. Chromatography has also shown great potential in metallomics due to the diversity of applications that it offers [20, 21]. Size-exclusion chromatography allows for preliminary screenings based on the molecular weight, but its usual low resolution makes mandatory the use of orthogonal separation methods to accomplish more comprehensive metallomics profiling, such as ion-exchange chromatography, normal-phase chromatography, or reversed-phase chromatography [22]. Finally, capillary electrophoresis offers important advantages such as the requirement of small sample volumes and increased resolution [23]. Anyway, chromatographic and electrophoretic approaches present several drawbacks that hinder their implementation in metallomics, including long analysis times and the possibility of occurring losses or transformations of metallo-species during the separation process. To overcome these limitations, some authors have proposed the application of ultra-filtration for rapid and simple size fractionation of high molecular mass (HMM, i.e., metalloproteins) and low molecular mass (LMM, i.e., metallo-metabolites) metal species [24, 25]. Although this method considerably reduces the total analysis time, metal species may be adsorbed onto filter membranes, thus worsening metal–ligand

integrity and leading to the transformation of species. As an alternative, González-Domínguez et al. developed a method for size fractionation of metal species based on the precipitation of proteins under non-denaturing conditions, which showed excellent performance for the characterization of serum metallomics profiles [26]. This method was successfully applied to investigate the involvement of metal homeostasis in Alzheimer's disease [27, 28]. In this chapter, we describe in detail a modification of this high-throughput metallomics method, devised for the analysis of metal species in plasma and erythrocyte samples.

2 Materials

High-purity reagents and solvents (e.g., trace metal grade) must be used to avoid contamination with trace metal impurities.

2.1 Sample Collection and Preparation

1. BD Vacutainer tubes and Advance vacuum system.
2. Saline solution: 9 g L^{-1} of sodium chloride.
3. Protein precipitation solvent: cold acetone (−20 °C).
4. Alkaline solution: 2% 1-butanol (w/v), 0.05% EDTA (w/v), 0.05% Triton X-100 (w/v), 1% NH_4OH (w/v).
5. Internal standard stock solution: 1000 mg L^{-1} of rhodium in 5% nitric acid (v/v).

2.2 Metallomics Analysis

1. Agilent 7900 inductively coupled plasma mass spectrometer (ICP-MS) equipped with a collision/reaction cell system and with nickel sampling and skimmer cones (Agilent Technologies, Tokyo, Japan).
2. Collision gas: high-purity grade helium (>99.999%).
3. Tuning solution: 1 μg L^{-1} of lithium, cobalt, yttrium, and thallium in 2% nitric acid (v/v) (*see* **Note 1**).
4. Multielemental stock solutions: 100 mg L^{-1} of lithium, aluminum, vanadium, chromium, manganese, iron, cobalt, nickel, copper, zinc, arsenic, selenium, molybdenum, cadmium, and lead in 5% nitric acid (v/v) (*see* **Note 1**).
5. Multielemental calibration curve: 0.5, 1, 10, 50, 250, 500, 1000, and 2500 μg L^{-1} for each metal in alkaline solution, containing 1 μg L^{-1} rhodium as the internal standard.

3 Methods

3.1 Sample Collection

1. Collect blood samples by venipuncture of the antecubital region using BD Vacutainer tubes and Advance vacuum system (*see* **Note 2**).
2. Gently invert the tube 8–10 times.
3. Centrifuge at 1500 × *g* for 10 min at 4 °C.
4. Collect the supernatant plasma, divide it into aliquots in cryotubes, and store at −80 °C until analysis (*see* **Note 3**).
5. Add 5 mL of cold saline solution (4 °C) to the Vacutainer tube containing the pelleted blood cells.
6. Gently invert the tube 8–10 times.
7. Centrifuge at 1500 × *g* for 5 min at 4 °C and discard the supernatant.
8. Repeat **steps 5–7** two more times.
9. Aliquot the erythrocyte fraction in cryotubes and store at −80 °C until analysis.

3.2 Size Fractionation of Metal Species by Protein Precipitation Under Non-denaturing Conditions

1. Aliquot 150 μL of plasma, or 50 μl of erythrocytes, in an Eppendorf tube.
2. Place the tubes in an ice bath and add dropwise 300 μl of cold acetone (−20 °C) (*see* **Note 4**).
3. Vortex the samples for 10 min at 4 °C using an orbital rotator mixer.
4. Centrifuge at 10,000 × *g* for 10 min at 4 °C.
5. Transfer the supernatant to a new tube and take it to dryness using a SpeedVac system (*see* **Note 5**).
6. Reconstitute the residue obtained after solvent evaporation in 3 mL of alkaline solution and add rhodium to reach a final concentration of 1 μg L^{-1} (LMM fraction).
7. Reconstitute the protein pellet obtained after centrifugation (*see* **step 4**) in 3 mL of alkaline solution and add rhodium to reach a final concentration of 1 μg L^{-1} (HMM fraction). Use an ultrasound bath to facilitate the reconstitution (*see* **Note 6**).
8. Filter the samples through 0.45-μm pore size PTFE filters and store at 4 °C until analysis.

3.3 Determination of the Total Metal Content

1. Aliquot 150 μL of plasma, or 50 μl of erythrocytes, in a tube.
2. Dilute with alkaline solution to a final volume of 3 mL.
3. Add rhodium to reach a final concentration of 1 μg L^{-1} (TOTAL fraction).
4. Filter the samples through 0.45-μm pore size PTFE filters and store at 4 °C until analysis.

3.4 Multielemental Analysis by ICP-MS

1. Turn the ICP-MS instrument on and let the plasma stabilizes for at least 10 min.
2. Optimize the instrumental conditions in terms of MS resolution and sensitivity using the tuning solution.
3. Aspire blank alkaline solution for at least 10 min to equilibrate the system (*see* **Note 7**).
4. Analyze the samples, calibration curves, blank samples, and quality control samples (*see* **Note 8**) using the following operating conditions:
   (a) Sampling depth: 7 mm.
   (b) Forward power: 1550 W.
   (c) Plasma gas: 15 L min^{-1}.
   (d) Auxiliary gas: 1 L min^{-1}.
   (e) Carrier gas: 1 L min^{-1}.
   (f) Make-up gas: 0.10 L min^{-1}.
   (g) Helium: 5 mL min^{-1}.
   (h) Isotopes monitored: ^{7}Li, ^{27}Al, ^{51}V, ^{52}Cr, ^{53}Cr, ^{55}Mn, ^{56}Fe, ^{57}Fe, ^{59}Co, ^{60}Ni, ^{63}Cu, ^{66}Zn, ^{75}As, ^{77}Se, ^{78}Se, ^{82}Se, ^{95}Mo, ^{98}Mo, ^{103}Rh, ^{11}Cd, and ^{208}Pb.
   (i) Dwell time: 0.3 s per isotope.

3.5 Data Processing and Quality Control

1. Inspect blank samples to corroborate the absence of high-background signals, cross-contaminations, and carryover.
2. Monitor the intensity of the internal standard across all the samples to check the absence of signal drifts along the sequence run (Fig. 1a).
3. Compute the intensity ratios for each target metal element (i.e., ratio between the intensity of the analyte and the intensity of the internal standard) in all the samples. Using the calibration curve, obtain the linear regression equation by plotting the intensity ratio against the concentration, and then calculate the sample concentrations by interpolation.
4. Calculate the relative standard deviation for the concentrations of all the metals detected in quality control samples, and check that it is below 15% (or 20% for low abundance metals).
5. Apply principal component analysis (PCA) to look for outliers and abnormal clustering in the corresponding score plot (Fig. 1b).

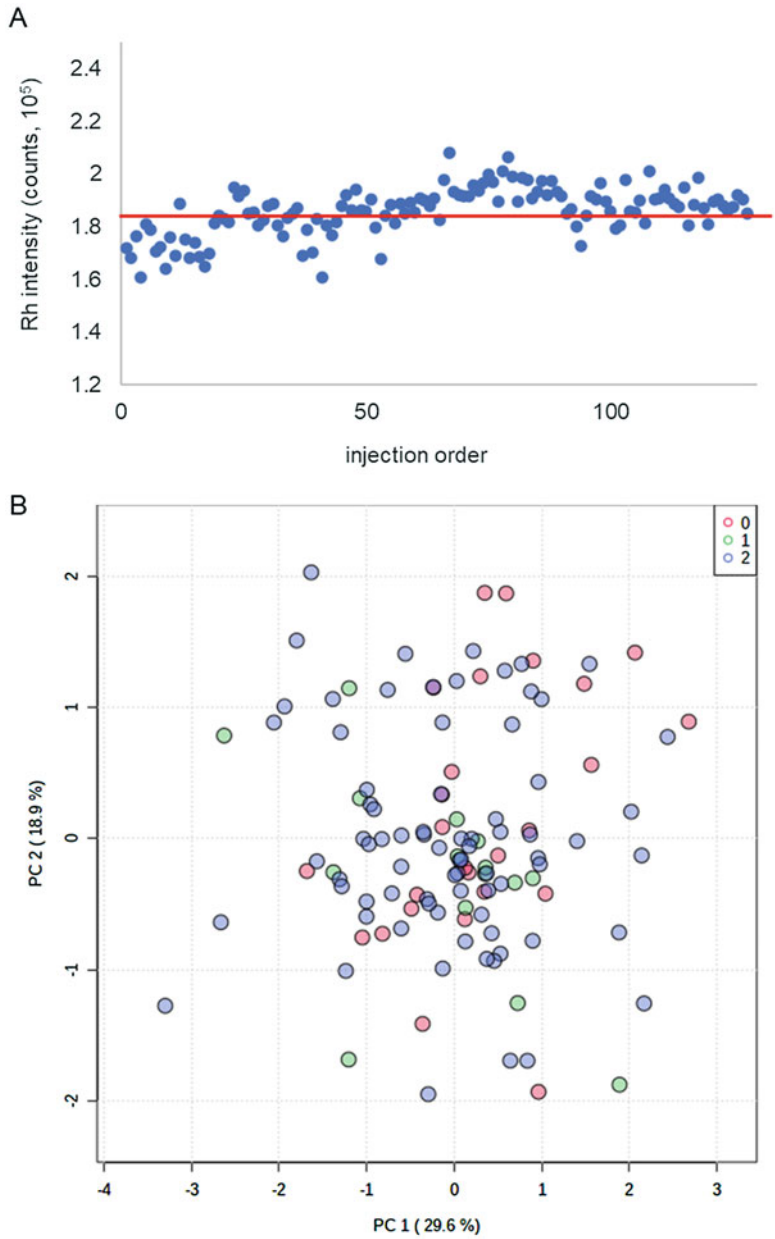

Fig. 1 Representative quality control plots for testing the analytical performance of the metallomics method. (**a**) Control chart of the internal standard intensity along the sequence run; (**b**) principal component analysis (PCA) score plot

4 Notes

1. We use the tuning and multielemental stock solutions that are described under Subheading 2.2, which can be purchased from Agilent Technologies, according to the ICP-MS manufacturer specifications. The reader can opt for other solutions and suppliers depending on the instrumentation to be used.
2. To minimize the influence of the circadian rhythm and diet, it is recommended to collect the samples in the morning and after at least 8 h of fasting.
3. Discard hemolyzed samples, as this may strongly alter the circulating metallomic profile, mainly in terms of iron levels.
4. The organic precipitation solvent must be added slowly and at low temperature to avoid protein denaturation, which would provoke the transformation of the metallo-species present in the sample.
5. A nitrogen stream can also be employed for solvent evaporation if a vacuum evaporator is not available.
6. The reconstitution of the protein pellets must be done as soon as possible after transferring the supernatant to a new tube. Delay times may considerably hinder their complete reconstitution.
7. As the tuning solution is prepared in a nitric acid-containing aqueous solvent, this equilibration step is crucial to ensure reliable analytical reproducibility.
8. First, analyze several blank samples to check the instrumental reproducibility and the absence of background contaminants. Then, inject the calibration curve and, before and after analyzing the set of samples under study, run several blanks to avoid cross-contaminations. Across the entire sequence, intercalate quality control samples to monitor the system stability (e.g., one quality control every 10 samples). Quality control samples can be prepared by mixing equal aliquots of each of the samples under study.

Acknowledgments

AGD is supported by an intramural grant from the Biomedical Research and Innovation Institute of Cádiz (LII19/16IN-CO24), and RGD is a recipient of a “Miguel Servet” fellowship (CP21/00120) funded by Instituto de Salud Carlos III.

References

1. Lobinski R, Becker JS, Haraguchi H, Sarkar B (2010) Metallomics: guidelines for terminology and critical evaluation of analytical chemistry approaches (IUPAC Technical Report). Pure Appl Chem 82:493–504. https://doi.org/10.1351/PAC-REP-09-03-04
2. Wang C, Zhang R, Wei X, Lv M, Jiang Z (2020) Metalloimmunology: the metal ion-controlled immunity. Adv Immunol 145: 187–241. https://doi.org/10.1016/bs.ai.2019.11.007
3. Zoroddu MA, Aaseth J, Crisponi G, Medici S, Peana M, Nurchi VM (2019) The essential metals for humans: a brief overview. J Inorg Biochem 195:120–129. https://doi.org/10.1016/j.jinorgbio.2019.03.013
4. Stevenson MJ, Uyeda KS, Harder NHO, Heffern MC (2019) Metal-dependent hormone function: the emerging interdisciplinary field of metalloendocrinology. Metallomics 11: 85–110. https://doi.org/10.1039/C8MT00221E
5. Valko M, Jomova K, Rhodes CJ, Kuča K, Musílek K (2016) Redox- and non-redox-metal-induced formation of free radicals and their role in human disease. Arch Toxicol 90:1–37. https://doi.org/10.1007/s00204-015-1579-5
6. Paithankar JG, Saini S, Dwivedi S, Sharma A, Chowdhuri DK (2021) Heavy metal associated health hazards: an interplay of oxidative stress and signal transduction. Chemosphere 262: 128350. https://doi.org/10.1016/j.chemosphere.2020.128350
7. Umair M, Alfadhel M (2019) Genetic disorders associated with metal metabolism. Cells 8: 1598. https://doi.org/10.3390/cells8121598
8. Krzywoszyńska K, Witkowska D, Świątek-Kozłowska J, Szebesczyk A, Kozłowski H (2020) General aspects of metal ions as signaling agents in health and disease. Biomolecules 10:1417. https://doi.org/10.3390/biom10101417
9. World Health Organization – Obesity. Available online: https://www.who.int/health-topics/obesity#tab=tab_1. Accessed 3 Mar 2022
10. World Health Organization – Childhood Obesity. Available online: https://www.who.int/dietphysicalactivity/childhood/en/. Accessed 3 Mar 2022
11. Kumar S, Kelly AS (2017) Review of childhood obesity. Mayo Clin Proc 92:251–265. https://doi.org/10.1016/j.mayocp.2016.09.017
12. Morales Camacho WJ, Molina Díaz JM, Plata Ortiz S, Plata Ortiz JE, Morales Camacho MA, Calderón BP (2019) Childhood obesity: aetiology, comorbidities, and treatment. Diabetes Metab Res Rev 35:e3203. https://doi.org/10.1002/dmrr.3203
13. Panchal SK, Wanyonyi S, Brown L (2017) Selenium, vanadium, and chromium as micronutrients to improve metabolic syndrome. Curr Hypertens Rep 19:10. https://doi.org/10.1007/s11906-017-0701-x
14. González-Domínguez Á, Visiedo-García FM, Domínguez-Riscart J, González-Domínguez R, Mateos RM, Lechuga-Sancho AM (2020) Iron metabolism in obesity and metabolic syndrome. Int J Mol Sci 21:5529. https://doi.org/10.3390/ijms21155529
15. Feldman A, Aigner E, Weghuber D, Paulmichl K (2015) The potential role of iron and copper in pediatric obesity and nonalcoholic fatty liver disease. Biomed Res Int 2015:1–7. https://doi.org/10.1155/2015/287401
16. Tinkov AA, Filippini T, Ajsuvakova OP, Aaseth J, Gluhcheva YG, Ivanova JM, Bjørklund G, Skalnaya MG, Gatiatulina ER, Popova EV, Nemereshina ON, Vinceti M, Skalny AV (2017) The role of cadmium in obesity and diabetes. Sci Total Environ 601–602:741–755. https://doi.org/10.1016/j.scitotenv.2017.05.224
17. Rothenberg SE, Korrick SA, Fayad R (2015) The influence of obesity on blood mercury levels for U.S. non-pregnant adults and children: NHANES 2007–2010. Environ Res 138: 173–180. https://doi.org/10.1016/j.envres.2015.01.018
18. Templeton DM, Ariese F, Cornelis R, Danielsson LG, Muntau H, van Leeuwen HP, Lobinski R (2000) Guidelines for terms related to chemical speciation and fractionation of elements. Definitions, structural aspects, and methodological approaches (IUPAC Recommendations 2000). Pure Appl Chem 72: 1453–1470. https://doi.org/10.1351/pac200072081453
19. Issaq HJ (2001) The role of separation science in proteomics research. Electrophoresis 22: 3629–3638. https://doi.org/10.1002/1522-2683(200109)22:17<3629::AID-ELPS3629>3.0.CO;2-O
20. Rabilloud T, Chevallet M, Luche S, Lelong C (2010) Two-dimensional gel electrophoresis in proteomics: past, present and future. J Proteomics 73:2064–2077. https://doi.org/10.1016/j.jprot.2010.05.016

21. Bouyssiere B, Szpunar J, Lobinski R (2002) Gas chromatography with inductively coupled plasma mass spectrometric detection in speciation analysis. Spectrochim Acta Part B At Spectrosc 57:805–828. https://doi.org/10.1016/S0584-8547(02)00024-1
22. Montes-Bayón M, DeNicola K, Caruso JA (2003) Liquid chromatography–inductively coupled plasma mass spectrometry. J Chromatogr A 1000:457–476. https://doi.org/10.1016/S0021-9673(03)00527-2
23. Álvarez-Llamas G, Fernández de la Campa MR, Sanz-Medel A (2005) ICP-MS for specific detection in capillary electrophoresis. TrAC Trends Anal Chem 24:28–36. https://doi.org/10.1016/j.trac.2004.09.007
24. Bantan Polak T, Milačič R, Mitrovič B, Benedik M (2001) Speciation of low molecular weight Al complexes in serum of CAPD patients. J Pharm Biomed Anal 26:189–201. https://doi.org/10.1016/S0731-7085(01)00397-1
25. Nischwitz V, Berthele A, Michalke B (2010) Rapid size fractionation of metal species in paired human serum and cerebrospinal fluid samples using ultrafiltration with off-line element selective detection. J Anal At Spectrom 25:1130. https://doi.org/10.1039/b913468a
26. González-Domínguez R (2017) Size fractionation of metal species from serum samples for studying element biodistribution in Alzheimer's disease. In: White AR (ed) Metals in the brain. Neuromethods, vol 124. Humana Press, New York. https://doi.org/10.1007/978-1-4939-6918-0_8
27. González-Domínguez R, García-Barrera T, Gómez-Ariza JL (2014) Characterization of metal profiles in serum during the progression of Alzheimer's disease. Metallomics 6: 292–300. https://doi.org/10.1039/C3MT00301A
28. González-Domínguez R, García-Barrera T, Gómez-Ariza JL (2014) Homeostasis of metals in the progression of Alzheimer's disease. BioMetals 27:539–549. https://doi.org/10.1007/s10534-014-9728-5

Chapter 13

Analysis of Brain Metabolites Using Two Complementary Ultrahigh-Performance Liquid Chromatography–Mass Spectrometry Methods

Alexa M. Jauregui, Sofia E. Parellada, Emily Neag, and Sanjoy K. Bhattacharya

Abstract

Metabolomics continues to progress, but obstacles remain. The preservation of metabolites in the target tissue and gathering information on the current metabolic state of the organism of interest proves challenging. Robustness, reproducibility, and reliable quantification are necessary for confident metabolite identification and should always be considered for effective biomarker discovery. Recent advancements in analytical platforms, techniques, and data analysis make metabolomics a promising omics for significant research. However, there is no single approach to effectively capturing the metabolome. Coupling separation techniques may improve the power of the analysis and facilitate confident metabolite identification, especially when performing untargeted metabolomics. In this chapter, we will present an untargeted metabolomic analysis of brain tissue from C57BL/6 mice using two UHPLC–MS methods based on reversed-phase and HILIC chromatography.

Key words Untargeted metabolomics, Ultrahigh-performance liquid chromatography, High-resolution mass spectrometry, LC–MS/MS, HILIC chromatography, Reversed-phase chromatography

1 Introduction

An "untargeted" approach reveals the most complete fingerprint of metabolites related to either a normal or diseased state of an organism. The untargeted approach has the potential of covering a wide swath of compounds providing greater insight into all repertoire of metabolites. The goal of untargeted mass spectrometry-based metabolomics is to detect the relative changes in metabolite concentrations and annotate the metabolites based on accurate mass-to-charge ratio and retention time [1]. Untargeted metabolomics methods involve high-resolution instruments such as time-of-flight (TOF), quadrupole time-of-flight (QTOF), and Orbitrap Fourier transform mass spectrometry (FTMS) instruments. The recent

advancements in analytical platforms accounts for the demands of the field, which includes increasing sensitivity for detecting metabolites present in low concentrations in biological samples [2]. Identifying small molecule metabolites is also difficult and poses a challenge to high-confidence analyses with low ppm mass accuracy, specificity, and resolution [2, 3].

Metabolomics analyses are typically performed by gas chromatography–mass spectrometry (GC–MS) or liquid chromatography–mass spectrometry (LC–MS) [1, 4].

To this day, there is no single approach to effectively capture the metabolome. Therefore, coupling separation techniques to improve the power of the analysis is ideal. Multiple methods facilitate higher and more confident metabolite identification, especially when performing untargeted metabolomics. Separation techniques include capillary electrophoresis (CE), gas chromatography (GC), hydrophilic interaction liquid chromatography (HILIC), ion chromatography (IC), or reversed-phase (RP) chromatography. These techniques are coupled with a high-resolution mass spectrometer via either a direct heated GC–MS inlet or by a heated electrospray ionization LC–MS interface. While RP-UHPLC remains the gold standard for untargeted metabolomics, HILIC is becoming an attractive method to use along with RP-UHPLC. HILIC separates samples using a polar stationary phase (bare silica, zwitterionic, NH2), whereas the reversed phase uses a nonpolar stationary phase (C18, polar embedded) [1, 4].

In complex matrices such as plasma, urine, or tissues, hundreds to thousands of metabolites have the potential to be identified. Due to the fact that highly polar compounds (phosphates, amino acids, organic and nucleic bases) are poorly retained on a reversed-phase column, this presents another bottleneck faced by metabolomic scientists. Alternatively, these highly polar compounds can be separated using HILIC. To extend the coverage of measured compounds, an orthogonal method—combining the two types of chromatography—is recommended [5].

LC–MS untargeted metabolomics faces threats from frequent contamination and introduction of artifacts, or background noise. This results from the frequent lack of identification with no actual ability for follow-up. In general, using LC–MS systems will increase the probability of false positives due to mobile phase modifiers and salts giving rise to unwanted adduct formation. Therefore, the acquisition of MS/MS or MS^n spectra is crucial, as well as the implementation of more measures to ensure the instrument's performance, overall sensitivity, and reproducibility [3]. These measures include quality controls (QC), internal standards (IS), extraction blanks, and quality assurance (QA) placed throughout sample analysis. Furthermore, the sample preparation and extraction process must be carried out quickly and efficiently to avoid rapid metabolite turnover and oxidation [6].

In this chapter, we will present two UHPLC–MS techniques, reversed phase and HILIC, to apply untargeted metabolomics analysis on the brain tissue of C57BL/6 mice. This protocol allows for the detection of water-soluble charged metabolites and the investigation of semipolar metabolites.

2 Materials

Prepare all solutions using LC–MS grade solvents and chemicals.

2.1 Chemicals

1. Acetic acid.
2. Acetonitrile.
3. 200 mM ammonium acetate.
4. Dry ice and/or liquid N_2.
5. Formic acid.
6. Extraction solvent (for extraction from tissues): 1:1 MeOH: H_2O (v:v).
7. Phosphate-buffered saline (PBS), pH 7.4.
8. Internal standard: D-Glucose-1,2,3,4,5,6,6-d_7.

2.2 Ultrahigh-Performance Liquid Chromatography Columns and Vials

1. Accucore™ 150 Amide HILIC Column (150 mm × 2.1 mm, 2.6 μm, Thermo Scientific™).
2. Accucore™ Vanquish™ C18+ Column (100 mm × 2.1 mm, 1.5 μm, Thermo Scientific™).
3. HPLC vials.

2.3 Ultrahigh-Performance Liquid Chromatography and Mass Spectrometry Instruments

1. Thermo Scientific™ Vanquish™ Horizon UHPLC.
2. Thermo Scientific™ Q Exactive™ Orbitrap™ Mass Spectrometer coupled to a heated electrospray ionization source (HESI).
3. Thermo Scientific™ Xcalibur™.

3 Methods

3.1 Preparation of Tissue Samples

Carry out extraction procedure in a 4 °C refrigerated room. If unable to do so, pre-chill solutions at −20 °C for 1 h before use. Keep all samples and solutions on ice when carrying out sample preparation. Wear a lab coat throughout the entire procedure and change gloves and pipette tips as needed.

3.1.1 Dissection and Extraction

1. Pre-label and weigh 1.5-mL Eppendorf tubes (*see* **Note 1**).
2. Dissect <15 mg of the tissue sample and rinse in pre-chilled PBS for no more than 10 s prior to transfer.

3. Quench/flash freeze tissue sample in liquid N_2 prior to tissue placement into an Eppendorf tube (*see* **Note 2**). For this, carefully suspend tissue in liquid N_2 using sterilized forceps and let tissue sit for 1 min.
4. Tap the tissue on the side of the Eppendorf tube to release the tissue into the tube.
5. Add 50 μL of chilled extraction solvent to the Eppendorf tube containing the sample.
6. Mince the tissue with the solvent repeatedly for 60 s until it pulverizes completely (*see* **Note 3**).
7. Vortex the sample for 45 s.
8. Centrifuge tissue samples at 14,000 rpm for 20 min at 4 °C (Beckman Microfuge 18) to separate the protein pellet.
9. Transfer supernatant to a new, labeled Eppendorf tube (*see* **Note 4**).
10. *Repeat* **steps 5–9**, *mincing the pellet formed after centrifugation.*
11. Dry samples in the speed vacuum at room temperature until fully dry (*see* **Note 5**).
12. For long-term storage, flush the Eppendorf tubes containing the sample with Argon gas (*see* **Notes 6** and **7**).

3.1.2 Post-extraction

1. Reconstitute sample immediately prior to use by adding 100 μL of water.
2. Add 20 μL of internal standard (D-Glucose-1,2,3,4,5,6,6-d_7).
3. Sonicate samples for 25 min (*see* **Note 8**).
4. Centrifuge tissue samples at 14,000 rpm for 5 min at 4 °C (Beckman Microfuge 18).
5. Prepare and label UHPLC vials for both positive and negative modes.
6. Transfer 10 μL of each sample to the properly labeled, respective UHPLC vial.

3.1.3 Preparation of Pooled Quality Control (QC) Samples

1. Take 10 μL aliquots of each sample tube and transfer them to a new Eppendorf tube labeled "Pooled QC."
2. Vortex for 30 s.
3. Transfer 10 μL into each properly labeled, respective UHPLC vial (*see* **Note 9**).
4. Example of a typical run order can be seen in Table 1. Run order may change depending on the amount of samples.

Table 1
Example of a typical run order for an untargeted UHPLC–MS metabolomics experiment (*see* Notes 11, 12, and 13)

System Suitability Blanks 1–3
QC1
QC2
Extraction Blank 1
Pooled QC1—Full Scan
Sample
Sample
Sample
Sample
Sample
Sample
Sample
Sample
Pooled QC2
Extraction Blank 2
Pooled QC—ddMS2
Pooled QC—ddMS2
System Suitability Blank 4

3.2 Preparation of Extraction Blanks

1. Prepare a total of four extraction blanks following the instructions described in Subheading 3.1.1: two extraction blanks for the reversed phase and two extraction blanks for HILIC (*see* **Note 10**).
2. All blank samples should undergo the same extraction process as the biological sample using only the extraction solvents and no biological tissue.

3.3 Mobile Phase Preparation

Autoclave all solvent bottles prior to use (*see* **Note 11**).

3.3.1 Preparation of Mobile Phases A and B for Reversed-Phase C18+ UHPLC Positive and Negative Ion Modes

1. Mobile Phase A for both ionization modes consists of acetonitrile with 0.1% formic acid (v/v). Take 1 L of acetonitrile, and add 1.0 mL of formic acid and mix thoroughly.
2. Mobile Phase B for both ionization modes consists of water with 0.1% formic acid (v/v). Take 1 L of LC–MS grade water, and add 1.0 mL of formic acid and mix thoroughly.

3. Sonicate both mobile phases for 30 min (*see* **Note 8**).
4. Leave both mobile phases overnight at room temperature to stabilize.

3.3.2 Preparation of Mobile Phases A and B for HILIC UHPLC Positive and Negative Ion Modes

1. Mobile Phase A for both ionization modes consists of 10 mM ammonium acetate in 95% acetonitrile/5% water (v/v) with 0.1% acetic acid (v/v). Mix 450.0 mL of acetonitrile with 50.0 mL of the 200 mM ammonium acetate stock solution. Sonicate for 15 min until the solution is clear followed by the addition of 500.0 mL of acetonitrile to prepare a 1 L mobile phase solution. Add 1.0 mL of acetic acid to the mobile phase solution and mix thoroughly.
2. Mobile Phase B for both ionization modes consists of 10 mM ammonium acetate in 1:1 water/acetonitrile (v/v) with 0.1% acetic acid (v/v). Mix 450.0 mL of water with 500.0 mL of acetonitrile. Add 50.0 mL of the 200 mM ammonium acetate stock solution to prepare a 1-L mobile phase solution. Add 1.0 mL of acetic acid to the mobile phase solution and mix thoroughly.
3. Sonicate both mobile phases for 30 min (*see* **Note 8**).
4. Leave both mobile phases overnight at room temperature to stabilize.

3.4 UHPLC and MS Methods

3.4.1 Reversed-Phase C18+ UHPLC Gradient

1. Set the column temperature to 40 °C.
2. Set injection volume to 5 μL.
3. Details of the flow gradient are presented in Table 2.

Table 2
Reversed-phase method gradient using an Accucore™ Vanquish™ C18+ column

Retention time (min)	Flow (mL/min)	%B
0.000	0.300	0.0
1.000	0.300	0.0
5.000	0.300	50.0
6.000	0.300	98.0
10.000	0.300	98.0
10.100	0.300	0.0
15.000	0.300	0.0

Table 3
HILIC method gradient using an Accucore™ 150 Amide HILIC column

Retention time (min)	Flow (mL/min)	%B
0.000	0.500	1.0
1.000	0.500	1.0
9.000	0.500	95.0
10.000	0.500	95.0
10.500	0.500	1.0
15.000	0.500	1.0

3.4.2 HILIC UHPLC Gradient

1. Set the column temperature to 35 °C.
2. Set injection volume to 5 μL.
3. Details of the flow gradient are presented in Table 3.

3.4.3 HESI-II Ion Source Settings

1. Set the mass range to 67–1000 m/z.
2. Resolution should be set to 140,000 for full scan and 35,000 for $ddMS^2$.
3. The AGC target for full scan = 3×10^6 and $ddMS^2 = 2 \times 10^5$.
4. The max injection time (IT) is 200 s for full scan mode and 50 s for $ddMS^2$.
5. The number of microscans is 2.
6. The normalized collision energy (NCE) should be set to 20, 35, and 50.
7. Details of the Q Exactive HESI-II ion source parameters for the reversed-phase and HILIC method are shown in Tables 4 and 5, respectively.

3.5 Sequence Setup

1. The sequence/analysis run order should be created in XCalibur™ (*see* **Notes 9**, **10**, **12**, and **13**).

4 Notes

1. Laboratory plastic such as Eppendorf tubes and pipette tips should be chosen seeking high-quality plastics and tested for contaminants. Plastics interfere with the metabolic profile.
2. Dry ice may be used in place of liquid N_2. Quenching/flash freezing is essential to prevent the degradation of the metabolites.

Table 4
Q-Exactive HESI-II ion source settings for the reversed-phase C18+ method for both positive and negative acquisition modes

Sheath gas flow rate, *both* modes	40
Aux gas flow rate	
Positive mode	8
Negative mode	10
Sweep gas flow rate	
Positive mode	1
Negative mode	0
Spray voltage (kV)	
Positive mode	3.5
Negative mode	4.0
Capillary temperature (°C)	
Positive mode	275
Negative mode	320
S-lens RF level	
Positive mode	50
Negative mode	60
Aux gas heater temperature (°C)	
Positive mode	300
Negative mode	350

3. We employed this homogenization technique because it yielded the best results. Other possible techniques include pellet mixers, set screw homogenization, freeze/thaw cycles, or the use of ionic detergents.
4. Make sure not to disturb the protein pellet.
5. For best practices, speed vac initially for 20 min and check on samples. If the tube still contains liquid, check on samples every 5 min.
6. For long-term storage, store tissue samples at −80 °C until carrying out the experiment.
7. Samples may be stored for up to 1 week at −20 °C.
8. Sonication allows us to efficiently mix and dissolve all chemicals in solution while also removing air bubbles that might be present.

Table 5
Q-Exactive HESI-II ion source settings for the HILIC method for both positive and negative acquisition modes

In-source CID (kV)	0.5
Sheath gas flow rate, *both* modes	55
Aux gas flow rate, *both* modes	14
Sweep gas flow rate, *both* modes	4
Spray voltage (kV)	
Positive mode	3.5
Negative mode	2.5
Capillary temperature (°C)	
Positive mode	350
Negative mode	380
S-lens RF level, *both* modes	30
Aux gas heater temperature (°C), *both* modes	438

9. Pooled quality controls (QCs) contain the same volume of all experimental samples and, therefore, contain all compounds representative of a batch. "Pooled QCs" are run in separate LC–MS vials to account for reproducibility and analyte stability. Pooled QCs are also used for compound normalization and identification.
10. An extraction blank is prepared exactly the same way as the biological sample minus the biological sample. This includes all reagents, dissection tools, and instruments used from start to finish. Extraction blanks determine and correct for reagent effects, allow for the creation of exclusions lists, mark background components, and filter the background components from the results table.
11. Autoclaving using pressurized steam to sterilize laboratory equipment and decontaminate biohazardous waste ensures mobile phases are clean prior to entering the LC–MS.
12. 1–3 blank injections should be run to ensure sufficient column washing. Exposure of the stationary phase to a matrix sample for several beginning injections prevents retention time shifts.
13. At the beginning of the batch, add QC samples to equilibrate the system for the analyzed matrices. The QC samples also behave as a system suitability check.

References

1. Schrimpe-Rutledge AC, Codreanu SG, Sherrod SD, McLean JA (2016) Untargeted metabolomics strategies-challenges and emerging directions. J Am Soc Mass Spectrom 27(12): 1897–1905. https://doi.org/10.1007/s13361-016-1469-y
2. Ghaste M, Mistrik R, Shulaev V (2016) Applications of Fourier Transform Ion Cyclotron Resonance (FT-ICR) and orbitrap based high resolution mass spectrometry in metabolomics and lipidomics. Int J Mol Sci 17(6): doi:10.3390/ijms17060816
3. Brown M, Dunn WB, Dobson P, Patel Y, Winder CL, Francis-McIntyre S, Begley P, Carroll K, Broadhurst D, Tseng A, Swainston N, Spasic I, Goodacre R, Kell DB (2009) Mass spectrometry tools and metabolite-specific databases for molecular identification in metabolomics. Analyst 134(7): 1322–1332. https://doi.org/10.1039/b901179j
4. Lopes AS, Cruz EC, Sussulini A, Klassen A (2017) Metabolomic strategies involving mass spectrometry combined with liquid and gas chromatography. Adv Exp Med Biol 965:77–98. https://doi.org/10.1007/978-3-319-47656-8_4
5. Van Damme T, Lachova M, Lynen F, Szucs R, Sandra P (2014) Solid-phase extraction based on hydrophilic interaction liquid chromatography with acetone as eluent for eliminating matrix effects in the analysis of biological fluids by LC-MS. Anal Bioanal Chem 406(2):401–407. https://doi.org/10.1007/s00216-013-7281-7
6. Lu W, Su X, Klein MS, Lewis IA, Fiehn O, Rabinowitz JD (2017) Metabolite measurement: pitfalls to avoid and practices to follow. Annu Rev Biochem 86:277–304. https://doi.org/10.1146/annurev-biochem-061516-044952

Chapter 14

Analysis of Cholesterol from the Liver Using Gas Chromatography–Mass Spectrometry

Meher A. Saleem, Betsy Benitez, Charles Yaros, Gabrielle Yamar, and Sanjoy K. Bhattacharya

Abstract

Cholesterol is an essential lipid molecule for several biological functions including the proper functioning of cell membranes, lipoproteins, and lipid rafts, as well as the synthesis of bile acids, vitamin D, and steroid hormones. Cholesterol can be extracted from liver tissue by multiple methods of lipid extraction. Subsequently, gas chromatography–mass spectrometry (GC–MS) can be used to obtain the highest level of sensitivity and selectivity in the analysis of cholesterol. This chapter describes two methods of lipid extraction for liver tissue, Bligh and Dyer and methyl tertiary butyl ether (MTBE), followed by an analysis with GC–MS.

Key words Cholesterol, GC–MS, GC, MS, Gas chromatography, Mass spectrometry, Bligh and Dyer, MTBE, Lipid extraction, Liver

1 Introduction

Cholesterol is a lipid molecule that is essential for many biological functions in humans. Cholesterol allows maintenance of the structural integrity, fluidity, and permeability of cell membranes [1, 2]. It is a component of lipoproteins, which perform the transportation of lipids in the blood [1]. Cholesterol is also found in lipid rafts, which compartmentalize cellular processes in the cell membrane [2]. Cholesterol serves a role in the synthesis of bile acids, vitamin D, and steroid hormones. It has three main components including a tetracyclic carbon ring, a polar hydroxyl group at the proximal end, and a nonpolar carbon chain at the distal end [1, 2].

Cholesterol can be analyzed by several methods, including enzymatic assays, liquid chromatography, gas chromatography, and mass spectrometry. GC–MS is the routine method of analysis for cholesterol. This method can be used to determine cholesterol levels and to separate cholesterol from interfering species such as

Raúl González-Domínguez (ed.), *Mass Spectrometry for Metabolomics*,
Methods in Molecular Biology, vol. 2571, https://doi.org/10.1007/978-1-0716-2699-3_14,
© The Author(s), under exclusive license to Springer Science+Business Media, LLC, part of Springer Nature 2023

sterols. GC–MS is the preferred method for cholesterol analysis because of its accuracy, reproducibility, sensitivity, and selectivity [1, 3]. Another benefit of using GC–MS for cholesterol analysis is the low level of sample required by the instruments [1].

Gas chromatography is used to separate volatile and semi-volatile compounds. In the initial step, the sample containing the analyte of interest is placed in a chamber with an inert gas. The combination of the analyte and the inert gas is then passed through a column, which causes the separation of the mixture into individual compounds. Within the column, compounds are separated based on boiling point. More volatile compounds with a lower boiling point travel through the column more quickly, while compounds with a higher boiling point travel more slowly. Compounds also travel at different speeds based on particle size, with smaller compounds traveling through the column more quickly [4]. These compounds are subsequently analyzed in a detector. In GC–MS, the mass spectrometer is the detector the compounds enter after passing through the column. The mass spectrometer ionizes and fragments the particles and subsequently separates them by their mass to charge ratio [5].

There are multiple techniques available to extract lipids, including Bligh and Dyer, MTBE, Folch, and butanol–methanol (BUME) extractions. Two of the most common techniques are the Bligh and Dyer extraction and the MTBE extraction. The Bligh and Dyer extraction was discovered in 1959. This method of lipid extraction is commonly used for GC–MS and LC–MS analysis [6]. The MTBE extraction was discovered in 2008 and is an effective method for extracting a broad range of lipid classes from a large variety of biological materials. The primary difference between Bligh and Dyer and MTBE extraction is the constitution of the upper and lower phase found in the extraction process. Separation with the Bligh and Dyer method results in an aqueous upper layer and organic lower layer, while the MTBE method results in an organic upper layer and aqueous lower layer. When lipids are extracted from the lower organic layer in the Bligh and Dyer method, there is a risk of contamination from the upper aqueous layer. In the MTBE method, this risk of contamination is reduced because the lipids are found in the upper layer [7]. Both the Bligh and Dyer and MTBE methods can be used with liver tissue to achieve successful lipid extractions with comparable efficacy [8].

In this chapter, we describe a method to analyze cholesterol metabolites from liver tissue. We provide a modified protocol for both a Bligh and Dyer extraction as well as an MTBE extraction. The end user can choose to use either method or both methods of lipid extraction. We subsequently describe a protocol for analysis of the lipids with GC–MS.

2 Materials

2.1 Tissue Preparation

1. Liver tissue from mice.
2. Electric balance.
3. Scissors.

2.2 Lipid Extraction

2.2.1 Bligh and Dyer Lipid Extraction

1. Argon gas.
2. 1:1 methanol/chloroform solution (v:v).
3. Chloroform.
4. Vortexer.
5. Centrifuge.
6. Syringe.
7. Speed vacuum.

2.2.2 Methyl Tertiary Butyl Ether (MTBE) Lipid Extraction

1. 0.15 M ammonium acetate.
2. Vortexer.
3. 0.516 mg/mL of butylhydroxytoluene (BHT) in methanol.
4. Methyl t-butyl ether (MTBE).
5. Rocker.
6. Centrifuge.
7. Speed vacuum.
8. Ultrasonic cleaner.

2.3 GC–MS

1. Agilent 7980 A GC system with Agilent 7693 autosampler and Agilent 5975C mass spectrometer.
2. Agilent CP-Sil 8 CB Column.
3. Reconstitution solution for Bligh and Dyer extraction: chloroform.
4. Reconstitution solution for MTBE extraction: 1:1 methanol/chloroform solution (v:v).
5. GC–MS lipid standards: cholesterol and cholesterol-d7.
   (a) Prepare a 1 mg/mL solution of standards in chloroform. Dilute to 100 μg/mL in chloroform.

3 Methods

3.1 Tissue Preparation and Lipid Extraction

3.1.1 Bligh and Dyer Lipid Extraction Method

1. Place 1–5 mg of the liver tissue sample in a cryovial, and determine the weight on an electric balance.
2. Perform freezing and thawing of the tissue sample (*see* **Note 1**).
   (a) Place the sample in a −80 °C freezer (or liquid nitrogen) for 5–10 min.

(b) Place the sample in a 40 °C water bath for 5–10 min.

(c) Repeat **steps 2a–2b** for 4 additional times.

3. Mince the tissue sample using scissors for 60 s (*see* **Notes 2** and **3**).
4. Flush the tissue sample with argon gas (*see* **Note 4**).
5. Add 400 μL of methanol/chloroform solution to the tissue sample.
6. Add 350 μL of chloroform to the sample.
7. Vortex the sample for 20 s.
8. Centrifuge the sample twice at 4 °C at 10,000 × *g* for 15 min. Centrifugation will result in an upper (aqueous) layer containing protein and lower (organic) layer containing lipids (*see* **Note 5**).
9. Place a syringe in the lower (organic) layer of lipids and remove the entire layer in a single attempt. Place the lipids in a 2-mL glass vial (*see* **Note 6**).
10. Flush the lipids with argon gas (*see* **Note 4**).
11. Place vial in a speed vacuum for 90 min.
12. If any solvent remains after 90 min, keep the vial in the speed vacuum for an additional time until the sample is completely dry.
13. Flush the lipids with argon gas (*see* **Note 4**).
14. Store the lipids at −80 °C until GC–MS analysis is performed.

3.1.2 Methyl Tertiary Butyl Ether (MTBE) Lipid Extraction Method

1. Place the liver tissue sample in a cryovial, and determine the weight on an electric balance.
2. Add 100 μL of ammonium acetate solution to the tissue sample.
3. Mince the tissue sample using scissors for 60 s (*see* **Note 3**).
4. Vortex the sample for 2–3 min.
5. Add 1.2 mL of BHT solution to the sample.
6. Transfer the sample to a 12-mL glass tube for lipid extraction.
7. Add 4 mL of MTBE to the sample, maintaining a ratio of 10:3 MTBE/BHT solution.
8. Incubate the sample overnight at 4 °C in the dark on a rocker (*see* **Note 7**).
9. Add 1.05 mL of ammonium acetate solution to the sample, maintaining a ratio of 20:5:6 MTBE/BHT solution/ammonium acetate solution.

10. Centrifuge the sample at 4 °C for 10 min at 2000 × *g* at 4 °C. Centrifugation will result in an upper (organic) layer and lower (aqueous) layer.
11. Collect the upper (organic) layer in a 2-mL glass vial.
12. Place the vial in a speed vacuum until the sample is completely dry.
13. Store the lipids at −20 °C until GC–MS analysis is performed.

3.2 Gas Chromatography–Mass Spectrometry Analysis

3.2.1 Preparation of Instrument Settings

1. Prepare the gas chromatograph as follows:
   (a) Column: Agilent CP-Sil 8 CB.
   (b) Carrier gas: ultrahigh purity helium.
   (c) Carrier gas flow rate: 1.2 mL/min.
   (d) Oven temperature:
      (i) Set the oven temperature to 70 °C for 4 min.
      (ii) Increase the oven temperature to 318 °C at a rate of 10 °C/min.
      (iii) Maintain the oven temperature at 318 °C for 2 min.
   (e) Heater temperature: 250 °C.
   (f) Inlet pressure: 9.4 psi.
   (g) Septum purge flow: 3 mL/min.
   (h) Injection volume: 1 μL.
   (i) Injection mode: Splitless.
2. Prepare the mass spectrometer as follows:
   (a) Ion polarity: positive.
   (b) Mass range: 30–400 m/z.
   (c) Amu gain: 1840.
   (d) Amu offset: 118.69.
   (e) EM Volts: 1576.

3.2.2 Sample Preparation

1. Resuspend the dried lipid samples (from Subheading 3.1).
   (a) For lipids extracted by the Bligh and Dyer method (Subheading 3.1.1), resuspend the lipids in 100 μL of chloroform.
   (b) For lipids extracted by the MTBE method (Subheading 3.1.2), resuspend the lipids in 50 μL of methanol/chloroform solution, sonicate the lipids, and resuspend the pellet.
2. Add internal standard to each lipid sample.
   (a) For lipids extracted by the Bligh and Dyer method (Subheading 3.1.1), add 2 μL of internal standard solution.

(b) For lipids extracted by the MTBE method (Subheading 3.1.2), add 1 μL of internal standard solution.

(c) Compare the peak areas of the spectra to the internal standard peak area to determine quantification.

4 Notes

1. Freezing and thawing the tissue causes the phospholipid cell membrane to break, allowing the lipids to be solubilized more easily in the organic phase.
2. Using a spatula, check that no large pieces of tissue remain after mincing the tissue.
3. Keep the tissue sample on ice during the procedure.
4. Argon gas is used to prevent lipid oxidation from reactive species in the atmospheric air.
5. If separation is not visible, vortex the sample for 1 min and repeat the centrifugation. If separation is not visible after the second centrifugation, add 100 μL of double-distilled water (DD H_2O), vortex the sample for 1 min, and repeat centrifugation.
6. If you do not extract the lipids in a single use of the syringe and use the syringe more than once, protein may be present in your lower (organic) layer. If so, repeat centrifugation and separation (Subheading 3.1.1, **steps 7–9**).
7. The tubes should be laid down parallel to the rocker's surface to allow the liquid to move back and forth inside the tube.

References

1. Li L, Dutkiewicz EP, Huang Y et al (2019) Analytical methods for cholesterol quantification. Yàowu shípǐn fenxi 27(2):375–386. https://doi.org/10.1016/j.jfda.2018.09.001
2. Schade DS, Shey L, Eaton RP (2020) Cholesterol review: a metabolically important molecule. Endocr Pract 26(12):1514–1523. https://doi.org/10.4158/EP-2020-0347
3. Beale DJ, Pinu FR, Kouremenos KA et al (2018) Review of recent developments in GC–MS approaches to metabolomics-based research. Metabolomics 14(11):1–31. https://doi.org/10.1007/s11306-018-1449-2
4. Dutta M, Cai J, Gui W et al (2019) A review of analytical platforms for accurate bile acid measurement. Anal Bioanal Chem 411(19):4541–4549. https://doi.org/10.1007/s00216-019-01890-3
5. Sparkman OD, Penton Z, Kitson FG (2011) Gas chromatography and mass spectrometry, 2nd edn. Elsevier Science, San Diego
6. Sundermann A, Eggers LF, Schwudke D (2016) Liquid extraction: Bligh and dyer. Springer, Dordrecht
7. Eggers LF, Dominik S (2016) Lipid extraction: basics of the methyl-tert-butyl ether extraction. Springer, Dordrecht
8. Ostermann AI, Müller M, Willenberg I et al (2014) Determining the fatty acid composition in plasma and tissues as fatty acid methyl esters using gas chromatography – a comparison of different derivatization and extraction procedures. Prostaglandins Leukot Essent Fat Acids 91(6):235–241. https://doi.org/10.1016/j.plefa.2014.10.002

Chapter 15

Metabolomics Analysis of Aqueous Humor Based on Liquid Chromatography–Mass Spectrometry

Karolina Pietrowska, Diana Anna Dmuchowska, Adam Kretowski, and Michal Ciborowski

Abstract

Aqueous humor (AH) is a transparent fluid that fills the anterior segment of the eye. The composition and level of metabolites in AH are important for understanding its physiology and changes caused by the occurrence of eye disease. A simple method for the preparation and analysis of AH samples was developed using the liquid chromatography–quadrupole time-of-flight mass spectrometry (LC–QTOF-MS) technique. The analyses were performed using two types of chromatography: reversed-phase liquid chromatography–mass spectrometry (LC-RP–MS) and hydrophilic interaction liquid chromatography–mass spectrometry (LC-HILIC–MS), in the sample prepared by one protocol.

Key words Aqueous humor, Untargeted metabolomics, LC–MS, Reversed-phase analysis, HILIC analysis

1 Introduction

Aqueous humor (AH) is a transparent fluid that fills the anterior and posterior chambers of the eye. It is secreted by the ciliary epithelium in the eye's posterior chamber and flows around the lens and through the pupil to the anterior chamber [1]. It provides nutrients and removes metabolic waste from avascular tissues in the eye and maintains intraocular pressure [2]. AH is a mixture of proteins and small molecules, e.g., electrolytes, oxygen, carbon dioxide, glucose, urea, antioxidants, amino acids, fatty acids, acylcarnitines, lipids, and others [3, 4]. The composition and level of small molecules can be considered a chemical reflection of the current phenotype of a particular biological system [5]. In that case, analysis of AH metabolites is important for understanding its physiology and changes caused by the occurrence of eye disease.

From the chemical analysis point of view, the determination of compounds in AH is difficult. In the case of humans, an additional

Raúl González-Domínguez (ed.), *Mass Spectrometry for Metabolomics*,
Methods in Molecular Biology, vol. 2571, https://doi.org/10.1007/978-1-0716-2699-3_15,
© The Author(s), under exclusive license to Springer Science+Business Media, LLC, part of Springer Nature 2023

problem is a relatively small sample volume (about 50–150 μL) and the invasiveness of the collection procedure (during ophthalmic surgery) [6]. Therefore, the determination of small molecules in AH (especially those with low concentration) requires the development of sensitive and robust analytical methods [7].

Untargeted metabolomics is an advanced high-throughput approach that allows the detection of many small molecules even at very low concentrations. Among the analytical platforms used in metabolomics, liquid chromatography–mass spectrometry (LC–MS) allows for the highest metabolome coverage. A method for the preparation and analysis of AH samples was developed using the liquid chromatography–quadrupole time-of-flight mass spectrometry (LC–QTOF-MS) technique, which enables the simultaneous determination of metabolites from different classes. The analyses were performed using two types of chromatography: reversed-phase liquid chromatography–mass spectrometry (LC-RP–MS) to measure nonpolar or low-polar metabolites and hydrophilic interaction liquid chromatography–mass spectrometry (LC-HILIC–MS) to measure compounds with high polarity. Due to the small sample volume, the AH preparation procedure has been designed to allow RP and HILIC analysis from the same sample.

2 Materials

Prepare all solutions using ultrapure water (prepared by purifying deionized water) and LC–MS-grade solvents.

2.1 Samples and Solvents

1. AH samples
2. Extraction mixture: 1 ppm Zomepirac sodium salt in methanol: ethanol (1:1, v:v). Mix 100 μL of methanol and ethanol. Weigh 0.2 mg of Zomepirac sodium salt (used as internal standard) and add to the methanol/ethanol mixture to get a final concentration of 1 ppm. Store at −20 °C (*see* **Note 1**).
3. Needle wash mixture: isopropanol:water (1:1, v:v) with 0.1% formic acid. Mix 500 μL of isopropanol and water, and add 1 mL of formic acid.
4. Mobile phase A (LC-RP analysis): water with 0.1% formic acid.
5. Mobile phase A (LC-HILIC analysis): 5 mM ammonium formate in water. To the graduated 250-mL flask, add 78.8 μL of 25% ammonia and 47.5 μL of formic acid. Make up to 250 mL with water to get 5 mM ammonium formate at pH = 4.
6. Mobile phase B (LC-RP and LC-HILIC analysis): acetonitrile with 0.1% formic acid.
7. Reference mass mixture: acetonitrile:water (95:5, v:v) with 5 mM purine solution, and 2.5 mM HP-0921 solution. To

prepare 500 mL, mix 475 mL of acetonitrile, 25 mL of water, 250 μL of 5 mM purine solution, and 600 μL of 2.5 mM HP-0921 solution (*see* **Note 2**). Mix precisely after adding each reagent (*see* **Note 3**). Store at 4 °C. In the case of using instruments from other vendors, prepare adequate reference mass mixture.

8. Calibration solution: acetonitrile:water (95:5, v:v) with ESI–L Undiluted Tuning Mix, and 0.1 mM HP-0321 solution. To prepare 100 mL, mix 10 mL ESI-L Undiluted Tuning Mix, 88.5 mL of acetonitrile, 4.5 mL of water, and 3 μL of 0.1 mM HP-0321. Mix precisely after adding each reagent (*see* **Notes 3** and **4**). Store at 4 °C. In the case of using instruments from other vendors, prepare an adequate calibration solution.

2.2 Equipment

1. LC–MS system: 1260/1290 Infinity UHPLC (consisted of a degasser, two binary pumps, isocratic pump, column oven, and a thermostated autosampler) coupled to a 6545 QTOF-MS with electrospray ionization (ESI) type of ion source (all Agilent Technologies, Santa Clara, CA, USA).
2. LC-RP guard column: Poroshell 120 EC-C18, 3.0 × 5 mm, 2.7 μm column (Agilent Technologies, Santa Clara, California, USA)
3. LC-RP column: Poroshell 120 EC-C18, 3.0 × 100 mm, 2.7 μm column (Agilent Technologies, Santa Clara, California, USA)
4. LC-HILIC guard column: SeQuant® ZIC®-HILIC Guard PEEK, 20 × 2.1 mm, 5 μm column (Merk KGaA, Darmstadt, Germany).
5. LC-HILIC column: SeQuant® PEEK ZIC®-HILIC, 2.1 × 100 mm, 3.5 μm column (Merk KGaA, Darmstadt, Germany).

3 Methods

3.1 AH Sample Collection

1. To a patient undergoing cataract surgery, puncture the anterior chamber of the eye using a 30G needle and aspirate approximately 50–150 μL of AH.
2. Immediately transfer AH to Eppendorf tubes and place them at −20 °C in the operating theatre.
3. As soon as possible, transfer AH samples to −80 °C and store them until the day of analysis.

3.2 AH Sample Preparation

1. Thaw samples on ice.
2. Vortex the samples and transfer 50 μL of AH to a fresh tube (*see* **Note 5**).
3. Add 50 μL of precooled at −20 °C extraction mixture to 50 μL of AH sample.
4. Vortex-mix samples vigorously for 1 min.
5. Incubate samples on ice for 10 min.
6. Centrifuge samples at 21,000 × *g* for 10 min at 4 °C (*see* **Note 6**).
7. Filter the supernatant through a 0.22-μm nylon filter directly into the vials' inserts.

3.3 Quality Assurance

1. Prepare quality control (QC) sample by pooling 10 μL of each sample and following the same protocol as sample preparation.
2. Prepare a blank extraction sample using 50 μL of extraction mixture instead of AH and follow the same protocol as sample preparation.
3. Prepare a worklist:
   (a) Start the sequence from the analysis of a few blank samples.
   (b) Before AH samples, inject 8–10 times QC sample (*see* **Note 7**).
   (c) Then start to inject experimental samples (*see* **Note 8**).
   (d) After each 5–10 samples, inject one QC sample.
   (e) Continue AH and QC samples injection until all samples are analyzed (*see* **Note 9**).
   (f) At the end, inject another few blank samples.

3.4 LC-RP–MS Method

1. Set the autosampler temperature at 4 °C.
2. Set the following LC gradient with a flow rate of 0.5 mL/min:
   (a) Start at 1% of phase B for the first minute.
   (b) Increase to 100% of phase B in 9 min.
   (c) Decrease to 1% of phase B in 0.1 min.
   (d) Re-equilibrate the column with 1% of phase B for 4.9 min.
3. Set the injection cleaning and switch the valve at 2.0, 10.2, and 14.0 min of the method.
4. Set the column oven temperature at 30 °C.
5. Set the flow rate of the reference mass mixture (*see* **Notes 10** and **11**) with a 1:100 split.

6. Set the reference masses: *m/z* 121.0509 (protonated purine) and *m/z* 922.0098 (protonated hexakis (1H,1H,3H-tetrafluoropropoxy)phosphazine or HP-921) in positive ion mode and *m/z* 119.0363 (proton abstracted purine) and *m/z* 966.0007 (formate adduct of HP-921) in negative ion mode (*see* **Note 12**).
7. Operate the spectrometer in separate runs, in both positive and negative ion modes in full scan mode from *m/z* 50–1000.
8. Set collecting data type to centroid mode at a scan rate of 2 spectra per second.
9. Set the following Dual AJS ESI source parameters (*see* **Note 13**):
   (a) Capillary voltage: 3000 V.
   (b) Drying gas flow rate: 12 L/min at 250 °C.
   (c) Gas nebulizer: 45 psig.
   (d) Sheath gas flow rate: 11 L/min at 370 °C.
   (e) Fragmentor voltage: 175 V and 225 V for positive and negative ionization modes, respectively.
10. Inject 1 μL of the sample (*see* **Note 14**).

3.5 LC-HILIC–MS Method

1. Set the autosampler temperature at 4 °C.
2. Set the following LC gradient with a flow rate of 0.1 mL/min:
   (a) Start at 80% of phase B and decrease to 15% of phase B in 13 min.
   (b) Return to starting conditions (80% of phase B) in 1 min.
   (c) Re-equilibrate the column with 80% of phase B for 5 min.
3. Set the injection cleaning and switch the valve at 2.0, 13.5, and 18.0 min of the method.
4. Set the column oven temperature at 30 °C.
5. Set the flow rate of the reference mass mixture (*see* **Notes 10** and **11**) with a 1:100 split. Use the same reference masses solution as for the LC-RP method.
6. Operate the spectrometer in separate runs, in both positive and negative ion modes in full scan mode from *m/z* 50–1000.
7. Set collecting data type to centroid mode at a scan rate of 1.5 spectra per second.
8. Set the following Dual AJS ESI source parameters (*see* **Note 13**):
   (a) Capillary voltage: 3500 V.
   (b) Drying gas flow rate: 13 L/min at 200 °C.
   (c) Gas nebulizer: 30 psig.

(d) Sheath gas flow rate: 11 L/min at 350 °C.

(e) Fragmentor voltage: 175 V for both ionization modes.

9. Inject 1 μL of the sample (*see* **Note 14**).

4 Notes

1. It is better to prepare the extraction mixture in advance to get the mixture appropriately frozen. The prepared solution can be stored at −20 °C for a few months.
2. 5 mM purine solution and 2.5 mM HP-0921 solution were purchased from Agilent Technologies (API-TOF Reference Mass Solution Kit, G1969-85001).
3. The reagents should be added in the given order and mixed precisely to prevent precipitation of the components of the mixture.
4. ESI-L Undiluted Tuning Mix (G1969-85000) and 0.1 mM HP-0321 (ESI-TOF Biopolymer Analysis Ref Mass Std, G1969-85003) were purchased from Agilent Technologies.
5. Samples must be prepared in randomized order to avoid a correlation between biological factors and preparation order.
6. Precool the centrifuge before use.
7. QC samples must be analyzed at the beginning for system equilibration and conditioning to achieve fully reproducible conditions.
8. Samples must be analyzed in a randomized order (as they should be prepared) to avoid correlation between samples' worklist order, source dirtiness during analysis, and systemic bias.
9. QC samples must be analyzed before, at regular intervals, and at the end of the worklist to monitor the stability and reproducibility of the analytical process.
10. The flow rate should be set to get the level of each reference mass between 50K and 100K counts.
11. Check the accuracy of reference masses. In case they have more than 1 ppm error, perform a calibration procedure.
12. Controlled m/z values depend on the reference masses used.
13. It may be needed to adjust ion source parameters to mass spectrometer from other vendors or with a different types of ESI ion source.
14. After drawing the sample to the loop and before injection, wash the needle in the Flush Port for 10 s to avoid contamination and transfer of samples.

References

1. Barbas-Bernardos C, Armitage EG, Garcia A et al (2016) Looking into aqueous humor through metabolomics spectacles – exploring its metabolic characteristics in relation to myopia. J Pharm Biomed Anal 127:18–25
2. Pietrowska K, Dmuchowska DA, Krasnicki P et al (2018) Analysis of pharmaceuticals and small molecules in aqueous humor. J Pharm Biomed Anal 159:23–36. https://doi.org/10.1016/j.jpba.2018.06.049
3. Snytnikova O, Khlichkina A, Yanshole L et al (2016) Metabolomics of the human aqueous humor. Metabolomics 13(1):1–9
4. Pietrowska K, Dmuchowska DA, Samczuk P et al (2017) LC-MS-based metabolic fingerprinting of aqueous humor. J Anal Methods in Chem 2017:1–13. https://doi.org/10.1155/2017/6745932
5. Pietrowska K, Dmuchowska DA, Krasnicki P et al (2018) An exploratory LC-MS-based metabolomics study reveals differences in aqueous humor composition between diabetic and non-diabetic patients with cataract. Electrophoresis 39(9–10):1233–1240. https://doi.org/10.1002/elps.201700411
6. Kliuchnikova AA, Samokhina NI, Ilina IY et al (2016) Human aqueous humor proteome in cataract, glaucoma, and pseudoexfoliation syndrome. Proteomics 16(13):1938–1946
7. Hassib ST, Elkady EF, Sayed RM (2016) Simultaneous determination of timolol maleate in combination with some other anti-glaucoma drugs in rabbit aqueous humor by high performance liquid chromatography-tandem mass spectroscopy. J Chromatogr B 1022:109–117

Chapter 16

Analyses and Localization of Serotonin and L-DOPA in Ocular Tissues by Imaging Mass Spectrometry

Varun Krishnan, Sean Meehan, Colin Hayter, and Sanjoy K. Bhattacharya

Abstract

Imaging mass spectrometry (IMS) allows for visualization of the spatial distribution of proteins, lipids, and other metabolites in a targeted or untargeted approach. The identification of compounds through mass spectrometry combined with the mapping of compound distribution in the sample establishes IMS as a powerful tool for metabolomics. IMS analysis for serotonin will allow researchers to pinpoint areas of deficiencies or accumulations associated with ocular disorders such as serotonin selective reuptake inhibitor optic neuropathy. Furthermore, L-DOPA has shown great promise as a therapeutic approach for disorders such as age-related macular degeneration, and IMS allows for localization, and relative magnitudes, of L-DOPA in the eye. We describe here an end-to-end approach of IMS from sample preparation to data analysis for serotonin and L-DOPA analysis.

Key words High-resolution mass spectrometry, Imaging mass spectrometry, MALDI, Localization, Metabolite visualization, Metabolite localization, Mass spectrometry, Metabolomics, Serotonin, L-DOPA

1 Introduction

Mass spectrometry has been established as an industry standard for molecular identification with high sensitivity and specificity [1, 2]. The use of mass spectrometry in tissue analysis, however, requires tissue homogenization and subsequent separation steps, resulting in the loss of spatial information of metabolites [3]. Imaging mass spectrometry (IMS) addresses this by enabling the analysis of intact tissue, thereby preserving the localization of compounds in the sample. IMS instruments differ from traditional MS setups. They consist of a fast-moving stage alongside an ionizing probe with a high firing rate [4]. The ionizing probe is run over the sample surface and at each contact point ionized a mass spectrum is generated [5]. The number of points, or pixels, that are ionized is dictated by the user and is what determines the spatial resolution [6]. Ionizing probes vary from ion beams to lasers, and here we use

Raúl González-Domínguez (ed.), *Mass Spectrometry for Metabolomics*,
Methods in Molecular Biology, vol. 2571, https://doi.org/10.1007/978-1-0716-2699-3_16,
© The Author(s), under exclusive license to Springer Science+Business Media, LLC, part of Springer Nature 2023

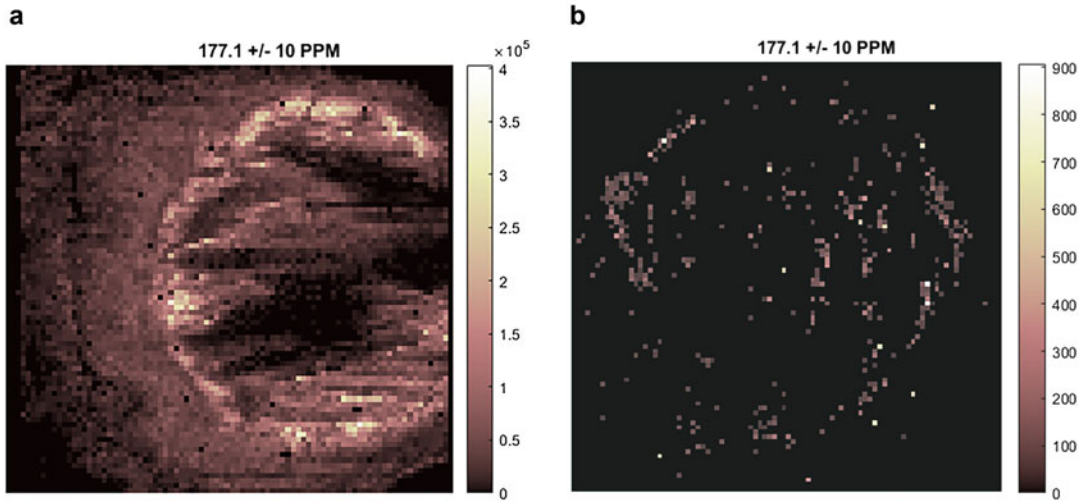

Fig. 1 Serotonin MSI heatmaps at 6 months and 8 months for knockout mice model. (**a**) Gene knockout model at 6 months. (**b**) Gene knockout model at 8 months

a laser-based approach through matrix-assisted laser desorption/ionization (MALDI) [5]. MALDI confers several advantages compared to electrospray ionization, including production of only singly charged ions, high throughput, and speed. MALDI requires matrix-facilitated sample crystallization for efficient ionization, with the matrix chosen being dependent on the compound of interest [7]. For L-DOPA and serotonin, due to their chemical nature with low molecular mass, 2,5-dihydroxybenzoic acid (DHB) is a suitable matrix option. 2,5-DHB results in less background noise of matrix clusters and hence, higher resolutions in the lower m/z range [8].

We used atmospheric pressure-MALDI (AP-MALDI) as our ionization instrument, which has the additional benefit of ionizing samples at normal atmospheric pressures and does not require that ionization take place in a vacuum. Used in conjunction with a Thermo Q-Exactive orbitrap mass spectrometer, we were able to successfully generate a heat map of serotonin (MW = 176.1) and L-DOPA (MW = 180.1) from ocular tissue of a gene knockout model at two different ages (Figs. 1 and 2). Similar methods using MALDI MS have resulted in comparable visualizations of other metabolites [9–11].

2 Materials

Ensure use of only HPLC-grade solvents. Please exercise caution when handling solutions respective to their materials safety data sheets (MSDS). Always wear PPE when needed and for transferring

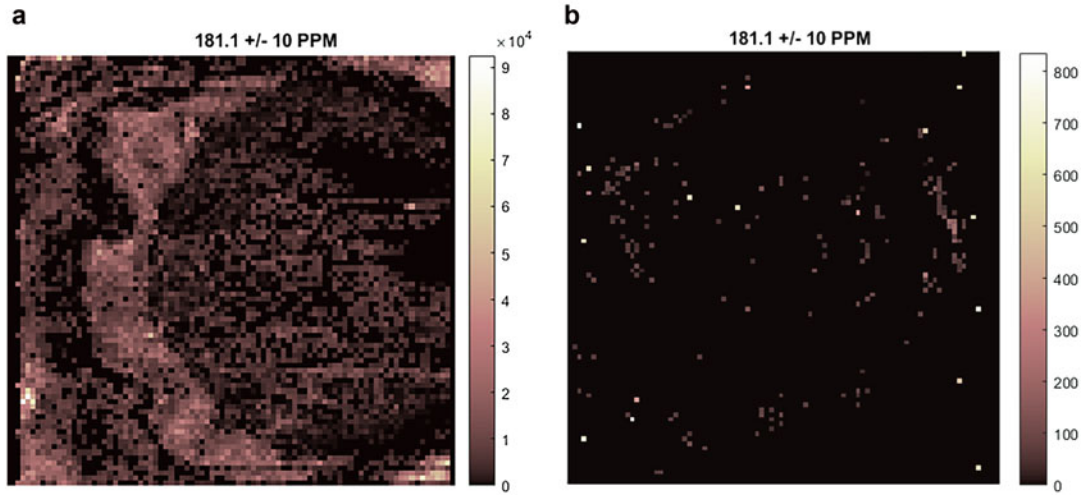

Fig. 2 L-DOPA MSI heatmaps at 6 months and 8 months for knockout mice model. (**a**) Gene knockout model at 6 months. (**b**) Gene knockout model at 8 months

solutions use a fume hood. Waste disposal should be properly followed per EPA/EHS guidelines.

2.1 Solvents

1. Solvent A: 80% HPLC-grade water and 20% acetonitrile (v:v). Add 800 μl of HPLC-grade (or 18 MΩ•cm) water into a 1.7-mL microcentrifuge tube, add 200 μl of acetonitrile to the tube, and then mix well.
2. Solvent B: 0.003% TFA, 13% ethanol, and 84% acetonitrile (v:v:v). Mix 100 μl of 1% TFA solution with 900 μl of HPLC-grade (or 18 MΩ·cm) water in a 1.7-mL microcentrifuge tube, mix 30 μl of this 0.1% TFA solution with 130 μl of 200 proof ethanol, and add 840 μl of acetonitrile.

2.2 MALDI/MS Instrument or Accessories

1. MALDI Source-AP/MALDI (ng) UHR (MassTech Inc.).
2. Alignment Jig.
3. 0.3-mm tip airbrush (*see* **Note 4**).
4. Steel well for MALDI source.
5. Slide adapter (*see* **Note 5**).
6. Indium tin oxide (ITO) slides (25 mm × 50 mm, resistivity 20 W, standard thickness (1.1 mm)) (APMaldi, Model Number: MT-ITO) (*see* **Note 2**).
7. Q-Exactive™ Mass spectrometer (Thermo Fisher Scientific).

2.3 Software

1. TUNE software was used to set mass spectrometer parameters. Any analogous software that allows changing of your mass spectrometer parameters is viable.

2. TARGET ng software for control of MALDI source. Version 8.9.0 was used for this protocol and is available at https://www.apmaldi.com/main/downloads/.
3. MTIMZL Converter (MassTech Inc.) available at https://www.apmaldi.com/main/downloads/.
4. MSIReader for data analysis and heatmap generation (*see* **Note 1**).

2.4 Other Materials

1. Multimeter.
2. Cryostat.
3. OCT-embedded Eye Globe in cryostat mold.
4. Pipette.
5. 1X phosphate-buffered solution.
6. Vacuum desiccator.
7. Methanol.
8. Permanent lab marker.
9. Capillary extender (*see* **Note 3**).

2.5 Matrix Solutions for Calibration

1. 10 mg/mL stock α-cyano-4-hydroxycinnamic acid (α-CHCA) matrix solution: Dissolve 5 mg of α-cyano-4-hydroxycinnamic acid with 0.5 mL of Solvent B. Vortex or sonicate the mixture until the matrix is dissolved completely.
2. 3.5 mg/mL matrix solution: Mix 35 μl of the prepared 10 mg/mL Stock Matrix Solution (**step 1**) with 65 μl of Solvent B. Briefly vortex.

2.6 Calibration Mix Solutions

1. Reconstitute one vial of Normal Mass Calibration Mix (Sigma Aldrich) with 58 μl of Solvent A.
2. Incubate this vial for 30 min at room temperature.
3. Add 45 μl of the 3.5 mg/mL Matrix Solution to 5 μl of the reconstituted mass calibration mix. Vortex each sample briefly.

2.7 Matrix Solution for Tissue

1. 35 mg/mL DHB Matrix: Add 7.5 mL of methanol, 2.5 mL of water, and 100 μL of TFA into a glass amber bottle. Then, dissolve 350 mg of 2,5-dihydroxy benzoic acid into the solution (*see* **Note 6**).

3 Methods

3.1 Tissue Preparation Using Fresh Frozen Tissue Samples

1. Use the multimeter to determine and label the conductive side of the ITO-coated slide (*see* **Note 7** if you would like to overlay serial sections).
2. Equilibrate ITO-coated slides in cryomicrotome for 15 min.

3. Pipette 100 μL of 1X PBS onto tissue sample when reaching the section of choice and then carefully take an 8–10 μm section onto the *conductive* side of the ITO-coated slide.
4. Repeat **step 2** so that you have two sections on the ITO-coated slide (*see* **Note 8** for long-term storage).
5. Place slide in a vacuum desiccator for 30 min.

3.2 Matrix Application

1. Add a small amount of methanol to the airbrush container.
2. Clean the airbrush nozzle by spraying methanol for a couple of seconds.
3. Fill airbrush with 35 mg/mL DHB matrix (do *not* use the calibration matrix) in airbrush and spray onto ITO-coated slide with tissue evenly in a fume hood (*see* **Note 9**).
4. Let the slide dry for 5 min at room temperature.
5. Orient the slide into the slide adapter.
6. Screw slide into slide adapter.
7. Draw brackets with a permanent lab marker around the section to give a reference point of the sample during ionization (*see* **Note 10**) (*see* Fig. 3).

3.3 MALDI Source Preparation

1. Set the capillary temperature to 60 °C and wait for the capillary to cool down.
2. After the temperature has cooled, screw the capillary extender 1 mm from the base of the sheath cone (*see* **Note 11**).
3. Place MALDI source on a mass spectrometer and plug all necessary cables from MALDI into the mass spectrometer.

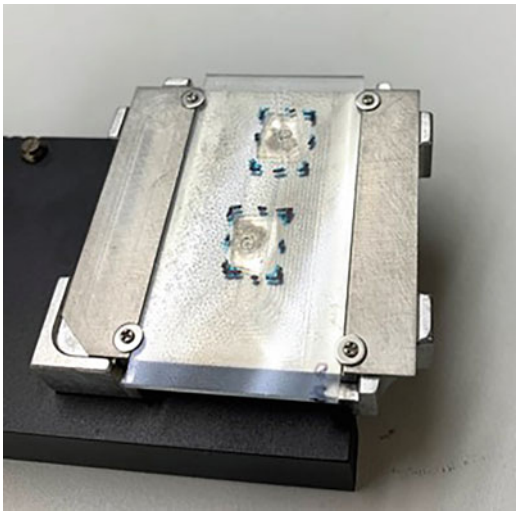

Fig. 3 Slide with tissue samples screwed on slide adapter after matrix application

4. Confirm the MALDI source is seen by TUNE software.
5. Turn MALDI power on.
6. Open TARGET analysis software and wait for motor plate xy initialization.
7. Remove MALDI source (but do not turn power off).
8. Put alignment jig into the ion transfer tube with spring side facing outward.
9. Put MALDI source back on.
10. Turn the laser on and see if the laser aligns with the jig. If misaligned, turn the knob on the side of the MALDI source so that the laser and the alignment jig become matched up.
11. Once completed, click and drag the cross (using control + click) on the TARGET software to align with the laser point.
12. Turn laser off, remove the source, and carefully remove the jig.
13. Put source back on. The laser is now aligned and MALDI is initialized for sample analysis.

3.4 Calibration of Mass Spectrometer

1. Spot 0.5 μL of calibration mix 3 times onto one well of the steel plate (1.5 μL total in well) (*see* **Note 12**). Wait 30 s for the mix to dry.
2. Insert steel plate into MALDI.
3. Set the TUNE parameters to the values in Table 1.

Table 1
Mass spectrometry parameters for calibration

Scan range (m/z)	500–2000
Fragmentation	None
Resolution	70K (*see* **Note 13**)
Polarity	Positive
Micro scans	1
Injection time (IT)	50
Lock masses	Off
AGC target	Fixed
Spray gas flow rate	0
Spray voltage	3.5 kV
Spray current	0
Capillary temperature	450°C
S lens RF	100

Table 2
TARGET parameters for calibration

Settings > Set Parameters > Laser Rep Rate (Hz)	1000
Settings > UHR Attenuator > Energy	4%
Settings > Zoom Parameters > Number of Rows	10
Settings > Zoom Parameters > Number of Columns	1
Settings > Zoom Parameters > Column Spacing	0.01
Settings > Zoom Parameters > Row Spacing	0.01
Settings > Zoom Parameters > CSR > Direction	Vertical
Settings > Zoom Parameters > CSR > Velocity	2.5

4. Set the acquisition parameters in TUNE to the following:
   (a) Acquisition time—Continuously
   (b) Destination File—[File of your choice]
   (c) On Start—Wait for contact closure
5. On TARGET software, set the parameters to the values in Table 2.
6. Select the well containing the spotted calibration mix in the TARGET software.
7. In TUNE, turn the mass spectrometer on and then click "Start on Acquisition."
8. Press "Play" in TARGET software.
9. Wait for line to complete and then in TUNE hit "Stop" under Acquisition (which should also turn the mass spectrometer to standby). Hit "Stop" on TARGET software as well.
10. Open the saved mass spectrometer file in Qual (or another mass spectrometry viewer), integrate, and confirm the m/z values in Table 3 are present.
11. In TUNE, under calibrate, click the customized calibration checkbox and add the masses in Table 3.
12. In TARGET, select the well containing the calibration mix.
13. In TUNE, turn the mass spectrometer on.
14. In TARGET, make sure the Magnifying Glass button in the toolbar is toggled. Then hit "Play" to begin laser scanning.
15. In TUNE, press "Calibrate."
16. Once calibration finished, ensure to stop the mass spectrometer in TUNE and the MALDI source in Target.

Table 3
ProteoMass normal mass calibration mix

Peptide	Amount per vial (pmol)	Mass $(M+H)^+$
MRFA	1217	524.27
Bradykinin 1–7	435	757.40
Bradykinin	680	1060.57
Angiotensin I	924	1296.69
Neurotensin	667	1672.92
Renin substrate	868	1758.93

Table 4
TARGET parameters for tissue

Settings > Set Parameters > Laser Rep Rate (Hz)	5000
Settings > UHR Attenuator > Energy	100%
Settings > Zoom Parameters > Column Spacing	0.03 (adjust as needed)
Settings > Zoom Parameters > Row Spacing	0.03 (adjust as needed)
Settings > Zoom Parameters > CSR > Direction	Horizontal, flyback

3.5 Data Acquisition from Sample

1. Use the same TUNE parameters in Table 1.
2. Set the TARGET parameters according to the values in Table 4. Modify the row and column spacing and use the preview to ensure that the entire area of interest is encompassed.
3. In TUNE, set the destination file.
4. In TUNE, set acquisition mode to "by time" and after acquisition to "standby" (*see* **Note 14**).
5. In TARGET, look under the magnifying glass tool to find the total time the scans will take. Add 5 min to this time and input this into TUNEs "by time" box.
6. In TUNE, turn on the mass spectrometer.
7. In TUNE, press "Start" under Acquisition.
8. In TARGET, make sure magnifying glass is toggled. Then, press "Play."
9. After Acquisition, you can store used slides at −80 °C if you want to re-run the tissue at a future date (*see* **Note 8** for long-term storage).

3.6 Data Analysis

1. The data file must be converted to .imzl prior to analysis. Begin by obtaining the raw mass spectrometer data file from the destination file set in TUNE. This extension should be .raw.

2. Obtain the .xml file containing position data of the mass spectrums by opening up Target Log Folder, which is found under Help > Log Folder in TARGET software.
3. Place both the .raw file and the .xml file in the same empty folder. Rename files so that both have the same name (preserving their current extension).
4. Open MT IMZL Converter and input raw file and target xml file in the inputs.
5. Click "Start" and wait for the generation of the .imzl file. Upon completion, the .imzl file will be in the same folder as the raw and xml files.
6. Open MSIReader.
7. In MSIReader, click the .imzl file under dropdown menu if not already selected and then click "Load Data." Do not adjust spots per line and number of lines as that is generated automatically from the .imzl file.
8. To generate heat maps for L-DOPA, type 181.1 into the mass navigation window on the left. Set tolerance to an acceptable value (10 ppm for our heatmaps). For serotonin, use the m/z of 177.1 (*see* **Notes 15** and **16**).

4 Notes

1. Many other MSI imaging software can be used, some free and some paid. A list is available here: http://ms-imaging.org/imzml/software-tools/.
2. A standard glass slide is 25 mm × 75 mm, meaning a 25 mm× 50 mm slide is hard to come by, especially when looking for an ITO-coated slide. One alternative is to buy a 25 mm × 75 mm ITO-coated slide and using a glass cutter to cut it to 50 mm. If this is done, please be sure to follow all safety measures associated with glass-cutting and handling of sharp glass. Note that the size of your required slide will depend on your specific MALDI source. Please check with your manufacturer.
3. A capillary extender is not needed, but we found it to greatly improve results. A capillary extender is used so that more ions generated from the sample can enter the ion transfer tube due to the closer proximity of the capillary and the sample.
4. There are many ways to apply matrices to a sample. We found the airbrush to be the most cost-effective way without compromising uniformity of the matrix application. The airbrush used here from Amazon costs $30, which is well cheaper than the thousands of dollars an automatic sprayer would cost.

When looking for an airbrush, we recommend obtaining a tip that is approximately 0.3 mm so the droplet size is optimally small, at least a 1/3 oz fluid cap, so constant refill is not required, and an air hose sufficiently long to operate the airbrush in a fume hood.

5. Most MALDI sources do not come with the capability to use a slide directly in the source. Hence, a slide adapter to your MALDI source is usually needed and ensure the adapter fits the correct dimensions of the slide used.
6. Many matrices can be used for MALDI depending on your experimental purpose. We used an ionic liquid matrix called 2,5-dihydroxybenzoic acid. Different matrices depending on your masses of interest, class of compound desired (e.g., peptides vs lipids), and many other factors, can be selected. For example, in the use of the calibration matrix, alpha-cyano-4-hydroxycinnamic acid is used due to its high resolution with smaller mass peptides such as those present in the mix [12, 13].
7. While not done in this protocol, one can overlay IMS-generated molecular distribution heatmaps concurrent with histologically stained (e.g., H&E) serial sections. To do this, one simply needs to take a serial section—a section right before or right after the section used on the ITO-coated slide. This section can be histologically stained as normal and then digitally overlayed with the IMS heatmap [14–16].
8. Slides can be stored at −80 C for approximately 1 year. To use the slide after freezing, bring it up to room temperature inside a vacuumed desiccator [17].
9. Applying a matrix onto a slide with an airbrush takes time and patience. After each spray coat, you should wait before the next coat to minimize droplet size. Our setup involved propping the slide against a cardboard cutout inside of the fume hood to catch any spray not falling on the ITO slide. We also recommend covering all surfaces around the spray area to ensure any spilling of the matrix does not result in contamination. It is advisable to practice spraying with DI water until comfort is reached. We ran the airbrush left to right on the slide holding the airbrush a constant 8 inches from the sample to ensure small droplet size.
10. This step is optional but greatly improves orientation when trying to find the sample prior to ionization. By placing a small mark with a marker near the sample, this mark can be easily distinguished with the MALDI camera allowing you to locate the sample faster and more predictably (see Fig. 3).
11. The distance you must screw the capillary extender may be different depending on the length of your ion transfer and

your MALDI source. The important point is to ensure the extender gets very close to, but not touching, the sample.

12. Do not simply pipette 1.5 μL of mix onto a well because this will cause the mix to overflow past the well. Instead, by spotting in 0.5-μL increments, the surface tension should allow all of the mix to stay confined in the well.
13. If you are looking for masses <500 Da, this resolution can be increased to 140K when running your sample.
14. By setting the acquisition mode to "standby," it will allow for the sample preparation to be run without supervision and the mass spectrometer not left turned on after sample acquisition is complete. Because higher resolution requires more data samples to be taken and hence, higher acquisition time, many samples will most likely be run overnight, allowing the use of this option.
15. Many more advanced processing methods are available. One common one is hotspot removal allowing you to normalize extremely high values in your heatmap and hence prevent "washing out" of lower range values. More on hotspot removal can be found in the MSIReader User Manual.
16. The mass to charge ratio of serotonin was determined by simply adding H^+ (1 g/mol) to the mass of serotonin (176.21 g/mol) as it has undergone single ionization. The mass to charge ratio of L-DOPA was determined by accounting for a methyl (16.04 g/mol) fragmentation of L-DOPA (197.18 g/mol), resulting in an m/z of 181.1.

References

1. Rivera ES, Djambazova KV, Neumann EK, Caprioli RM, Spraggins JM (2020) Integrating ion mobility and imaging mass spectrometry for comprehensive analysis of biological tissues: a brief review and perspective. J Mass Spectrom. https://doi.org/10.1002/jms.4614
2. Lössl P, van de Waterbeemd M, Heck AJ (2016) The diverse and expanding role of mass spectrometry in structural and molecular biology. EMBO J 35:2634–2657. https://doi.org/10.15252/embj.201694818
3. Seeley EH, Caprioli RM (2008) Molecular imaging of proteins in tissues by mass spectrometry. Proc Natl Acad Sci 105: 18126–18131
4. Spraggins JM, Caprioli RM (2011) High-speed MALDI-TOF imaging mass spectrometry: rapid ion image acquisition and considerations for next generation instrumentation. J Am Soc Mass Spectrom 22:1022–1031. https://doi.org/10.1007/s13361-011-0121-0
5. Piehowski PD, Zhu Y, Bramer LM et al (2020) Automated mass spectrometry imaging of over 2000 proteins from tissue sections at 100-μm spatial resolution. Nat Commun. https://doi.org/10.1038/s41467-019-13858-z
6. Buchberger AR, DeLaney K, Johnson J, Li L (2018) Mass spectrometry imaging: a review of emerging advancements and future insights. Anal Chem 90:240–265. https://doi.org/10.1021/acs.analchem.7b04733
7. Singhal N, Kumar M, Kanaujia PK, Virdi JS (2015) MALDI-TOF mass spectrometry: an emerging technology for microbial identification and diagnosis. Front Microbiol 6:791. https://doi.org/10.3389/fmicb.2015.00791
8. Fukuyama Y (2015) MALDI matrix research for biopolymers. Mass Spectrom. https://doi.org/10.5702/massspectrometry

9. Schulz S, Michael Becker M, Groseclose R, Schadt S, Hopf C (2019) Advanced MALDI mass spectrometry imaging in pharmaceutical research and drug development. Curr Opin Biotechnol 55:51–59. https://doi.org/10.1016/j.copbio.2018.08.003
10. Calligaris D, Feldman DR, Norton I et al (2015) MALDI mass spectrometry imaging analysis of pituitary adenomas for near-real-time tumor delineation. Proc Natl Acad Sci U S A 112:9978–9983. https://doi.org/10.1073/pnas.1423101112
11. Castellino S, Groseclose MR, Wagner D (2011) MALDI imaging mass spectrometry: bridging biology and chemistry in drug development. Bioanalysis 3(21):2427–2441. https://doi.org/10.4155/bio.11.232
12. Leopold J, Popkova Y, Engel KM, Schiller J (2018) Recent developments of useful MALDI matrices for the mass spectrometric characterization of lipids. Biomolecules 8:173. https://doi.org/10.3390/biom8040173
13. Gobom J, Schuerenberg M et al (2001) α-Cyano-4-hydroxycinnamic acid affinity sample preparation. Anal Chem 73:434–438. https://doi.org/10.1021/ac001241s
14. Karlsson O, Hanrieder J (2017) Imaging mass spectrometry in drug development and toxicology. Arch Toxicol 91:2283–2294. https://doi.org/10.1007/s00204-016-1905-6
15. Deutskens F, Yang J, Caprioli RM (2011) High spatial resolution imaging mass spectrometry and classical histology on a single tissue section. J Mass Spectrom 46:568–571. https://doi.org/10.1002/jms.1926
16. Patterson NH, Tuck M, Lewis A, Kaushansky A et al (2018) Next generation histology-directed imaging mass spectrometry driven by autofluorescence microscopy. Anal Chem 90: 12404–12413. https://doi.org/10.1021/acs.analchem.8b02885
17. Yalcin EB, de la Monte SM (2015) Review of matrix-assisted laser desorption ionization-imaging mass spectrometry for lipid biochemical histopathology. J Histochem Cytochem 63: 762–771. https://doi.org/10.1369/0022155415596202

Chapter 17

Isobaric Incorporation of C13-Histidine for the Assessment of Remyelination

Faith Christine Harvey, Anddre Osmar Valdivia, Colin Hayter, and Sanjoy K. Bhattacharya

Abstract

Multiple sclerosis is a demyelinating disease of the central nervous system characterized by the loss of the myelin sheath—the nonconductive membrane surrounding neuronal axons. Demyelination interrupts neuronal transmission, which can impair neurological pathways and present a variety of neurological deficits. Prolonged demyelination can damage neuronal axons resulting in irreversible neuronal damage. Efforts have been made to identify agents that can promote remyelination. However, the assessment of remyelination that new therapies promote can be challenging. The method described in this chapter addresses this challenge by using isobaric C13-histidine as a tag for monitoring its incorporation into myelin proteins and thus monitoring the remyelination process.

Key words Remyelination, Demyelinating disease, C13-histidine, Isobaric labeling, EAE mouse model, Mass spectrometry, Proteomics

1 Introduction

Demyelinating diseases, such as multiple sclerosis, can cause progressive loss of visual function due to continuous demyelination of the optic nerve [1–3]. Remyelinating agents may be used to counteract the disease progression. This chapter outlines a method to assess the amount of remyelination that occurs by injecting a C13-histidine tag with a remyelinating agent into the optic nerve. Resulting incorporation of the C13-histidine tag into newly synthesized myelinating proteins (such as myelin basic protein, MBP) indicates the amount of remyelination [4]. This chapter utilizes the experimental autoimmune encephalomyelitis (EAE) model, a well-established model of demyelination that has been documented to undergo demyelination in the optic nerve [5]. In brief, EAE optic nerves injected with a remyelinating agent and C13-histidine tag are separated by sucrose density ultracentrifugation, which will

Raúl González-Domínguez (ed.), *Mass Spectrometry for Metabolomics*,
Methods in Molecular Biology, vol. 2571, https://doi.org/10.1007/978-1-0716-2699-3_17,
© The Author(s), under exclusive license to Springer Science+Business Media, LLC, part of Springer Nature 2023

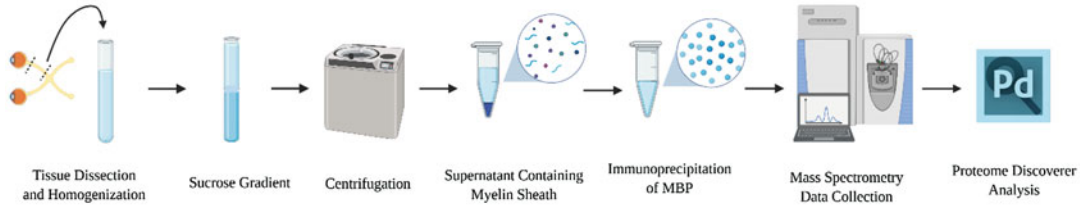

Fig. 1 General workflow of sample preparation and enrichment for mass spectrometry. Created with BioRender.com

isolate the contents of the myelin sheath, followed by myelin protein enrichment and mass spectrometry analysis (Fig. 1). The method outlined is specifically used to assess remyelination in the optic nerve but can be adapted to assess remyelination in other neurological regions.

2 Materials

2.1 Experimental Autoimmune Encephalomyelitis (EAE) Models

1. 2-month-old female C57Bl6/J mice strain (*see* **Note 1**).
2. Immunizing neuroantigen: emulsify 2 mg of myelin oligodendrocyte glycoprotein (MOG_{35-55}) and 6 mg of *Mycobacterium tuberculosis* (H37RA) in 800 μl of incomplete Freund's adjuvant and 800 μl of sterile phosphate-buffered saline (PBS) at pH 7.4.
3. 50 μg/mL pertussis toxin in sterile PBS, pH 7.4.

2.2 Optic Nerve Injections

1. 10 μM L-histidine-$^{13}C_6$ hydrochloride monohydrate (C13-histidine) in PBS.
2. Remyelinating agent: 150 μM LPC 18:1, 10 μM C13-histidine.
3. Ketamine and xylazine cocktail: 15 mg/mL of ketamine and 3 mg/mL of xylazine in sterile water.
4. Balanced Saline Solution (BSS).
5. Forceps and spring scissors.
6. 26G and 33G needles.
7. Neomycin/polymyxin B sulfate/dexamethasone ophthalmic ointment.
8. Heating pad and hydration packs.

2.3 Immunoprecipitation

1. Antibody for Myelin Basic Protein (MBP).
2. Centrifuge (Beckman L7, SW 28 rotor at 24,000 rpm).
3. Sepharose G beads.
4. 50 mM sodium borate buffer, pH 9.0.

5. Dimethyl pimelimidate dihydrochloride (DMP).
6. 200 mM ethanolamine.
7. Phosphate-buffered saline (PBS), pH 7.4.
8. 100 mM glycine–HCl buffer, pH 3.0.

2.4 Sample Preparation for Mass Spectrometry

1. Acetone.
2. Centrifuge (Thermo Fisher Megafuge 8R).
3. 6 M urea.
4. 50 mM ammonium bicarbonate.
5. 10 mM dithiothreitol (DTT) in 50 mM ammonium bicarbonate.
6. 20 mM dithiothreitol (DTT) in 50 mM ammonium bicarbonate.
7. 15 mM iodoacetamide in 50 mM ammonium bicarbonate.
8. 0.1 μg/μL of trypsin.
9. 50% formic acid (v/v).
10. Pierce Graphite Spin Columns (Thermo, 88302).
11. CentriVap Concentrator system (Labconco, Kansas City, MO).
12. Protein resuspension solution: 2% (v/v) acetonitrile in mass spectrometry grade water with 0.1% (v/v) formic acid.

2.5 Mass Spectrometry Analysis

1. Liquid chromatography system coupled to a high-resolution mass spectrometer: Easy nLC system coupled to a QExactive orbitrap mass spectrometer equipped with an Easy-Spray ionization source and controlled with Xcalibur software (Version 4.1.31.9, Released 2017) (Thermo Fisher Scientific).
2. Easy-Spray HPLC Column (Thermo Fisher Scientific).
3. Mobile phase A buffer: water with 0.1% formic acid.
4. Mobile phase B buffer: acetonitrile with 0.1% formic acid.

3 Methods

3.1 Generation of Experimental Autoimmune Encephalomyelitis (EAE) Models

1. Immunize animals with the neuroantigen solution s.c. at day 0. Each animal receives 0.333 mg of $MOG_{35\text{-}55}$ and 1 mg of *Mycobacterium tuberculosis*.
2. Inject 200 ng of the pertussis toxin solution by intraperitoneal injection at day −1 and day 1.
3. Monitor animals for the development of EAE symptoms (*see* **Note 2**).
4. Inject the remyelinating agent in the optic nerve at an EAE clinical score of 2 (*see* **Note 3**) [5].

3.2 Optic Nerve Injections

1. Anesthetize the animals using an intraperitoneal (IP) injection of 100 μL per 20 g of body weight of the ketamine and xylazine cocktail.
2. Place a drop of balanced saline solution (BSS) in the animal's eye to prevent dryness.
3. Use forceps to lift the conjunctiva and use spring scissors to make a small incision.
4. Insert forceps between the superior rectus and lateral rectus muscles. Make an opening by allowing the forceps to gradually return to the open position.
5. While the optic nerve is fully exposed, make a small incision in the meninges with a 26G needle.
6. Insert the blunt end of a 33G needle into the small space and deliver 0.5 μL of the remyelinating agent [4].
7. Apply ophthalmic ointment to the eye.
8. Keep animals on a 37 °C warm pad after the procedure and monitor for recovery from anesthesia.

3.3 Immunoprecipitation and Enrichment of MBP

1. Perform euthanasia with CO_2 chamber immersion followed by cervical dislocation.
2. Dissect optic nerve and place in a microcentrifuge tube (store at −80 °C, until needed).
3. Homogenize and suspend homogenate in sucrose for a final concentration of 1.25 M.
4. Layer a 1 M sucrose solution on top of 1.25 M sucrose layer to create a sucrose gradient.
5. Centrifuge samples at 85,000 × *g* for 2 h.
6. Collect the upper fraction containing the myelin sheath.
7. Wash myelin sucrose fraction by adding 3–4 times the volume of PBS followed by ultracentrifugation at 4 °C for 3 h at 140,000 g using SW 28 rotor. Repeat this step 2 times (*see* **Note 4**).
8. Prepare immunoprecipitation mixture by adding approximately 67 μg of Sepharose G beads suspended overnight in 200 μL of 50 mM sodium borate buffer, pH 9.0.
9. Incubate beads with 10 μg of MBP antibodies at room temperature for an hour.
10. Cross-link beads and antibodies by adding 10 μg of DMP, with 2-h incubation at room temperature. Repeat three times (total of 6-h incubation).
11. Incubate at 4 °C overnight.

12. Neutralize antibody-coupled beads with 200 μL of 200 mM ethanolamine and wash with 1 mL of PBS twice.
13. Incubate the antibody-coupled beads with 200 μg of lysate for an hour at room temperature.
14. Wash the beads twice with 500 μL of PBS and elute with two 20 μL volumes of 100 mM glycine–HCl buffer, pH 3.0.
15. Combine eluents and divide them into two equal amounts. Prepare samples for mass spectrometry analysis.

3.4 Sample Preparation for Mass Spectrometry

1. Take 15 μg of total protein from the sample (measured using BCA assay) and add 4 times the volume of acetone at −20 °C. Incubate overnight at −20 °C.
2. Centrifuge samples at 21,000 *g* at 4 °C for 30 min.
3. Discard the supernatant and air-dry pellet for 10 min.
4. Re-suspend pellet in 8 μL of 50 mM ammonium bicarbonate.
5. Denature by adding 15 μL of 6 M urea.
6. Reduce with 2 μL of 10 mM DTT.
7. Incubate for 1 h at room temperature.
8. Alkylate with 5 μL of 15 mM iodoacetamide for 30 min in the dark at room temperature.
9. Quench alkylation reaction with 3.33 μL of 20 mM DTT solution for 1 h at room temperature in the dark.
10. Dilute sample using 50 mM ammonium bicarbonate until it contains 1 M urea.
11. Digest sample with trypsin at a 1:30 (w/w) ratio of enzyme to protein. Incubate overnight at 37 °C.
12. Terminate the reaction using 50% formic acid at a 5:100 (v/v) ratio of formic acid to sample volume.
13. Desalt samples with Pierce Graphite Spin Columns following the manufacturer's recommendations.
14. Evaporate samples using the CentriVap Concentrator system.
15. Resuspend in 30 μL of protein resuspension solution and proceed to mass spectrometry analysis.

3.5 Mass Spectrometry and Proteomics Analysis

1. Set Easy-Spray source parameters to 2.0 kV spray voltage, 250 °C capillary temperature, and 70.00 S-Lens RF Level.
2. Set up the LC gradient as shown in Table 1.
3. Use an injection volume of 5.00 μL and a column temperature of 55 °C.
4. Set QExactive parameters to run a full scan at 70,000 resolution, AGC target of 1×10^6, maximum IT of 100 ms, and a scan range of 150–1600 m/z.

Table 1
nLC gradient for proteomics analysis

Time [mm:ss]	Duration [mm:ss]	Flow [nl/min]	Mixture [%B]
00:00	00:00	350	2
20:00	20:00	350	30
35:00	15:00	350	40
55:00	20:00	350	70
69:00	14:00	350	98
84:00	15:00	350	98

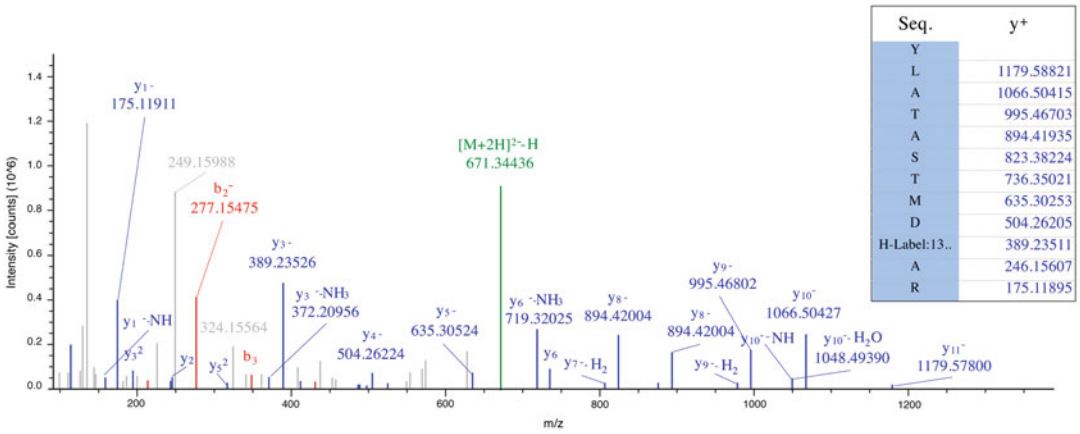

Fig. 2 Mass spectrometry spectrum showing the incorporation of C13-histidine into the MBP sequence

5. Set dd-MS^2 parameters at 17,500 resolution, AGC target of 5×10^5, maximum IT of 50 ms, and a scan range of 200–2000 m/z. Set the N(CE) to 28. Scans should be run in positive mode.
6. Identify proteins using Proteome Discoverer 2.2 (Version 2.2.0.388, Released 2017) and UniProt sequence database (Downloaded July 2021). The search parameters used in Proteome Discoverer for trypsin digested enzymes were as follows: Max missed cleavage sites: 2, Min. peptide length: 6, Max peptide length: 144, Modification: C13-His (+6.02 Da (H)), Max modification per peptide: 3, Precursor mass tolerance: 10 ppm, Fragment mass tolerance: 0.02 Da, Signal/Noise Threshold for spectra: 1.5, False discovery rate: Strict for PSMs 0.01, Strict for peptides 0.01.
7. Inspect mass spectra to corroborate the incorporation of the C13-histidine tag into the myelin basic protein (Fig. 2) [6].

4 Notes

1. Two-month-old female mice present the best age to induce EAE, as induction of EAE at later time points presents with immune resistance against EAE agents. Furthermore, male mice have been documented to have inherited resistance to EAE agents, therefore increasing the difficulty of inducing EAE.
2. EAE symptoms typically begin to manifest at about 12 days post-injection. Score 0 = normal mouse, no signs of disease; score 1 = tail paralysis; score 2 = partial hind limb paralysis; score 3 = compete hind limb paralysis; score 4 = fore limb paralysis; and score 5 = moribund state—suggesting euthanizing animal for humane reasons.
3. Optic nerve function begins to decline at EAE score 3. It is suggested to administer a remyelinating agent at EAE score 2 to allow time for the agent to take effect when the animal advances to EAE score 3.
4. Several repeats might be needed to ensure a minimal amount of sucrose, as sucrose can interfere with the immunoprecipitation procedure.

References

1. Lee GI, Park KA, Oh SY, Min JH, Kim BJ (2021) Differential patterns of parafoveal and peripapillary vessel density in multiple sclerosis and neuromyelitis optica spectrum disorder. Mult Scler Relat Disord 49:102780. https://doi.org/10.1016/j.msard.2021.102780
2. Murphy OC, Kalaitzidis G, Vasileiou E, Filippatou AG, Lambe J, Ehrhardt H, Pellegrini N, Sotirchos ES, Luciano NJ, Liu Y, Fitzgerald KC, Prince JL, Calabresi PA, Saidha S (2020) Optical coherence tomography and optical coherence tomography angiography findings after optic neuritis in multiple sclerosis. Front Neurol 11:618879. https://doi.org/10.3389/fneur.2020.618879
3. Park SH, Park CY, Shin YJ, Jeong KS, Kim NH (2020) Low contrast visual acuity might help to detect previous optic neuritis. Front Neurol 11: 602193. https://doi.org/10.3389/fneur.2020.602193
4. Valdivia AO, Bhattacharya SK (2021) Lyso-lipid induced oligodendrocytes maturation underlie restoration of optic nerve function. eNeuro
5. Valdivia AO, Agarwal PK, Bhattacharya SK (2020) Myelin basic protein phospholipid complexation likely competes with deimination in experimental autoimmune encephalomyelitis mouse model. ACS Omega 5(25): 15454–15467. https://doi.org/10.1021/acsomega.0c01590
6. Chauhan MZ, Valencia AK, Piqueras MC, Enriquez-Algeciras M, Bhattacharya SK (2019) Optic nerve lipidomics reveal impaired glucosylsphingosine lipids pathway in glaucoma. Invest Ophthalmol Vis Sci 60(5):1789–1798. https://doi.org/10.1167/iovs.18-25802

Chapter 18

Isolation and Lipidomic Screening of Human Milk Extracellular Vesicles

Victoria Ramos-Garcia, Isabel Ten-Doménech, Abel Albiach-Delgado, Marta Gómez-Ferrer, Pilar Sepúlveda, Anna Parra-Llorca, Laura Campos-Berga, Alba Moreno-Giménez, Guillermo Quintás, and Julia Kuligowski

Abstract

Extracellular vesicles (EVs) are secreted by cells and can be found in biological fluids (e.g., blood, saliva, urine, cerebrospinal fluid, and milk). EV isolation needs to be optimized carefully depending on the type of biofluid and tissue. Human milk (HM) is known to be a rich source of EVs, and they are thought to be partially responsible for the benefits associated with breastfeeding. Here, a workflow for the isolation and lipidomic analysis of HM-EVs is described. The procedure encompasses initial steps such as sample collection and storage, a detailed description for HM-EV isolation by multistage ultracentrifugation, metabolite extraction, and analysis by liquid chromatography coupled to mass spectrometry, as well as data analysis and curation.

Key words Extracellular vesicles, Exosomes, Human milk (HM), Lipidomics, Liquid chromatography–mass spectrometry (LC–MS)

1 Introduction

Eukaryotic and prokaryotic cells release a variety of nano- and micron-sized membrane-containing vesicles into their extracellular environment, which are collectively referred to as extracellular vesicles (EVs). EVs can be harvested from cell culture supernatants and from all body fluids including plasma, saliva, urine, cerebrospinal fluid, and human milk (HM). HM is known to be a rich source of EVs, with early milk containing a greater EV concentration compared to mature milk.

The physiological purpose of generating EVs remains largely unknown, and questions surrounding the function of EVs are mostly focused on understanding the fate of their constituents

Raúl González-Domínguez (ed.), *Mass Spectrometry for Metabolomics*,
Methods in Molecular Biology, vol. 2571, https://doi.org/10.1007/978-1-0716-2699-3_18,
© The Author(s), under exclusive license to Springer Science+Business Media, LLC, part of Springer Nature 2023

and the phenotypic and molecular alterations that they induce in recipient cells. The effects of EVs on recipient cells can vary due to the expression of different cell surface receptors, resulting in an array of possible biological functions including the induction of cell survival, apoptosis, and immunomodulation. In addition, recent studies indicate a functional, targeted, mechanism-driven accumulation of specific cellular components such as RNAs, small RNAs, nucleic acids, lipids, proteins, and metabolites in EVs, suggesting that they have a role in regulating intercellular communication [1]. Furthermore, encapsulation in EVs confers protection against enzymatic and nonenzymatic degradation of cargos and provides a pathway for cellular uptake of cargos by endocytosis of EVs [2, 3].

The composition of HM-EVs is still an open question. The richness of microRNAs (miRNAs) within HM-EVs was first discovered, and the dynamics during lactation stages was surveyed [4, 5]. Proteins have been studied in HM-EVs [6, 7], showing an enrichment of pathways involved in early-life immunity, as well as intestinal cell proliferation and migration. Although metabolites are known to be part of the EV cargo [8–10], literature reports providing insight into the metabolite cargo or lipid composition of HM-EVs are scarce. Hence, we developed a pipeline for the isolation and liquid chromatography–mass spectrometry (LC–MS)-based lipidomic screening of HM-EVs [11].

2 Materials

Prepare all solutions using ultrapure water (Q-POD® system, Merck KGaA, Darmstadt, Germany) and analytical grade reagents. Prepare all reagents at room temperature.

1. Phosphate-buffered saline (PBS) solution: 10 mM phosphate buffer, 2.7 mM potassium chloride, and 137 mM sodium chloride solution, pH 7.4. Dissolve one commercially available phosphate-buffered saline tablet in 200 mL of ultrapure water at 25 °C (*see* **Note 1**). Filter this solution with a 0.40 μm syringe filter. Store at 4 °C.
2. Internal standard (IS) mixture: Prepare individual 5 mM stock solutions of internal standards by weighing, e.g., 1.46 mg of oleic acid-D_9 and 1.49 mg of prostaglandin $F_{2\alpha}$-D_4 and dissolving them in 1 mL of ultrapure water. Put 976 μL of ultrapure water into a microcentrifuge tube. Add 16 and 7.8 μL of oleic acid-D_9 and prostaglandin $F_{2\alpha}$-D_4 5 mM stock solutions, respectively (resulting concentrations: 80 and 39 μM, respectively) (*see* **Note 2**). Aliquot to avoid freeze–thaw cycles and store at −20 °C.
3. Methanol (MeOH). Store at room temperature.

4. Methyl tert-butyl ether (MTBE). Store at room temperature.
5. Solution A: Isopropanol (IPA):MeOH:H_2O (5:1:4, v:v:v), 5 mM CH_3COONH_4, 0.1% formic acid (FA). Add 500 mL of IPA and 100 mL of MeOH into a 1-L volumetric flask. Do not fill to the calibration mark yet. Weigh 0.39 g of CH_3COONH_4, dissolve it in 20 mL of ultrapure water, and add this solution into the volumetric flask. Add 1 mL of pure FA into the volumetric flask. Fill to the calibration mark with ultrapure water. Store at 4 °C.
6. Solution B: IPA:H_2O (99:1, v:v), 5 mM CH_3COONH_4, 0.1% FA. Add 900 mL of IPA into a 1-L volumetric flask. Do not fill to the calibration mark yet. Weigh 0.39 g of CH_3COONH_4, dissolve it in 9 mL of ultrapure water, and add this solution into the volumetric flask. Add 1 mL of pure FA into the volumetric flask. Fill to the calibration mark with IPA. Store at 4 °C.
7. Solution A:B (1:1, v:v): mix 10 mL of solution A with 10 mL of solution B. Store at 4 °C.

For the performance of untargeted screening experiments, an LC (preferably: ultra-performance LC)-high-resolution MS instrument and a reversed-phase chromatographic column should be employed. The following protocol was developed using a 1290 Infinity HPLC system from Agilent Technologies (CA, USA) equipped with a UPLC BEH C18 column (50 × 2.1 mm, 1.7 μm, Waters, Wexford, Ireland) coupled to an Agilent 6550 Spectrometer iFunnel quadrupole time-of-flight (QTOF) MS.

3 Methods

Carry out all procedures at room temperature unless otherwise specified.

3.1 HM Sample Collection and Storage (See Note 3)

1. Clean the hands with water and soap for at least 15 s followed by drying with a clean towel.
2. Sterilize hands with hand sanitizer.
3. Clean and sterilize the milk pump (*see* **Note 4**).
4. Clean the skin area that gets into contact with the milk pump with water and dry it with sterile gauzes (*see* **Note 5**).
5. Extract the milk with the milk pump into clean sterile bottles following the manufacturer's instructions (*see* **Note 6**).
6. Manual shake gently during 30 s for sample homogenization.
7. Put the sample into 50-mL plastic Falcon tubes using a 25-mL plastic pipette (*see* Fig. 1a).

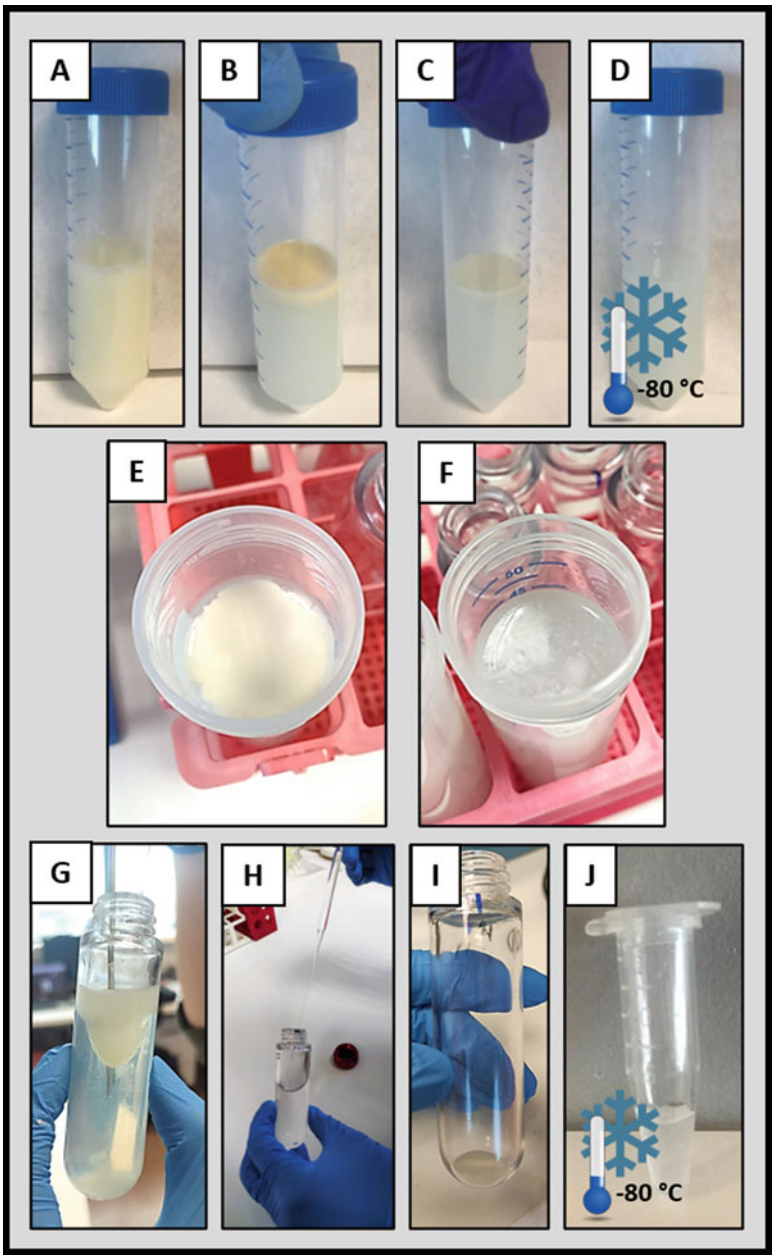

Fig. 1 Isolation of extracellular vesicles (EVs) from human milk (HM). (**a**) Raw HM sample. (**b** and **c**) HM sample after consecutive centrifugation steps with fat layer on the top. (**d**) Partially defatted HM sample for storage at −80 °C prior to EV isolation. (**e**) HM sample after first centrifugation step with remaining fat layer on the top. (**f**) Skimmed HM sample. (**g**) Supernatant aspiration after the first ultracentrifugation (10,000 rpm, 1 h, 4 °C) (protein pellet is discarded). (**h**) Supernatant removal after the second ultracentrifugation (30,000 rpm, 2 h, 4°) with the pellet containing EVs at the bottom. (**i**) Pellet containing EVs. (**j**) Reconstituted EVs in phosphate-buffered saline for storage at −80 °C until use

8. Centrifuge at 3000 × *g* for 10 min at 22 °C for cream, cell, and platelet removal [12, 13].
9. Remove and discard the cream layer with a spatula (*see* **Note 7** and Fig. 1b).
10. Transfer the liquid into a new 50-mL plastic Falcon tube.
11. Repeat **steps 8** and **9** (*see* Fig. 1c) and store milk at −80 °C until further processing (*see* Fig. 1d).

3.2 Isolation of EVs from HM

1. Thaw HM samples at 4 °C overnight by putting them in the refrigerator.
2. Centrifuge at 3000 × *g* for 10 min at 4 °C to remove the remaining milk fat and milk fat globules.
3. Remove the upper fat layer with a spatula (*see* **Note 7**) and discard (*see* Fig. 1e and f).
4. Transfer the liquid into a 25-mL polycarbonate bottle appropriate for ultracentrifugation.
5. Continue filling the ultracentrifuge tube with PBS until an estimated volume of 25 mL is reached and make sure all tubes have the same weight (including the caps) (*see* **Note 8**). This step is crucial as any ultracentrifuge can easily become unbalanced if equal masses are not located opposite to each other in the rotor.
6. Ultracentrifuge at 10,000 rpm for 1 h at 4 °C to pellet cellular debris, large size vesicles, and protein aggregates using a Hitachi CP100NX centrifuge with a Beckman Coulter 50.2 Ti rotor (Indianapolis, United States) or similar.
7. Collect the supernatant with a 25-mL plastic pipette or a syringe with a needle (*see* **Note 9**) and transfer it to a new 25-mL ultracentrifuge tube (*see* Fig. 1g).
8. Repeat **steps 5** and **6**.
9. Collect the supernatant into a different 25-mL ultracentrifuge tube, filtering it with a 0.45-μm syringe filter (*see* **Note 10**).
10. Repeat **step 5**.
11. Ultracentrifuge at 30,000 rpm for 2 h at 4 °C to pellet HM EVs using the same centrifuge rotor as in **step 6**.
12. Slowly aspirate and discard the supernatant using, e.g., a pipette (*see* Fig. 1h).
13. Wash and reconstitute the remaining pellet with 25 mL of PBS (*see* **Note 11**).
14. Repeat **steps 5** and **11–13** for a second ultracentrifugation.
15. Repeat **steps 5** and **11** for a third ultracentrifugation.
16. Slowly aspirate the PBS supernatant and save it to be later used as a blank. Store at −80 °C.

17. Reconstitute the pellet containing the isolated EVs (*see* Fig. 1i) with 200 μL of PBS (*see* **Note 12**).
18. Aliquot the isolated EVs in microtubes for quality control (QC) assays such as bicinchoninic acid (BCA) assay for protein quantification, nanoparticle tracking analysis (NTA) for vesicle concentration and size determination, etc., and for LC–MS analysis. Store at −80 °C (*see* Fig. 1j).

3.3 Extraction of the HM-EV Lipid Fraction

Lipids and other polar metabolites are extracted from HM-EVs using a single-phase extraction procedure [14–16].

1. Thaw the isolated HM-EV suspension in PBS obtained in Subheading 3.2 on ice and vortex for sample homogenization.
2. Mix 45 μL of isolated EVs with 5 μL of IS mixture.
3. Sonicate for 2 min.
4. Add 175 μL of MeOH followed by 175 μL of MTBE (*see* **Note 13**).
5. Vortex for 30 s for protein precipitation and compound extraction.
6. Sonicate for 2 min to assist the release of metabolites from EVs during extraction.
7. Centrifuge at 4000 × *g* for 15 min at 4 °C.
8. Transfer 100 μL of supernatant containing the extracted lipids and metabolites to a different microtube.
9. Dry at 35 °C using a centrifugal vacuum concentrator (*see* **Note 14**).
10. Reconstitute in 100 μL of solution A:B (1:1).
11. Prepare a pooled QC sample by mixing 5 μL of each reconstituted sample extract.
12. Prepare a calibration blank by repeating **steps 1–10**, replacing the initial 45 μL of the sample with 45 μL of ultrapure water.
13. Prepare a procedural blank by repeating **steps 1–10**, replacing the initial 45 μL of the sample with 45 μL of aspirated PBS supernatant from **step 18** in Subheading 3.2.
14. Analyze by LC–MS or, alternatively, store at −80 °C for further analysis.

3.4 LC–MS Lipidomics Method (See Note 15)

1. Perform a system suitability check of the MS (e.g., resolution, accuracy, and sensitivity) and UPLC performances (e.g., retention time (RT) of standards, resolution, lack of contamination of the analytical system) using blanks, standard mixtures, and QC samples (*see* **Note 15**).
2. Thaw the samples and place them in the autosampler.

3. Equilibrate the system by injecting the QC sample repeatedly. MS/MS data for peak annotation is typically acquired at the beginning or end of the batch [16] (*see* **Note 16**).
4. Run sample sequence including QC samples, study samples, and blanks, acquiring data using a suitable ionization interface (e.g., positive and/or negative electrospray ionization) (*see* **Notes 17** and **18**).

3.5 Data Processing and Analysis

1. Perform an initial data quality check through the manual integration of IS and several (e.g., 5–10) endogenous compounds with different intensities and RTs, which are expected to be detected in the sample under study. Assess the stability of peak areas, *m/z* accuracy, RT, and chromatographic parameters (e.g., peak width, resolution) throughout the analytical sequence.
2. Convert raw data into mzXML format using, e.g., ProteoWizard (http://proteowizard.sourceforge.net/) (optional) and generate a peak table (*see* **Notes 19** and **20**).
3. Annotate lipids using MS/MS information and available spectral and/or in-house LC–MS libraries (*see* **Note 21** and Figs. 2a–c).
4. Identify and correct the within-batch effect (*see* **Note 22** and Fig. 2d).
5. Perform data cleanup: (i) filtering of features found in blanks and (ii) filtering of features with high relative standard deviation (RSD) detected in QC samples (*see* **Note 23**).
6. Check data quality (*see* **Note 24** and Fig. 2e).

4 Notes

1. Prepare this solution fresh each time.
2. It is convenient to immerse the pipette tip in the water when adding small volumes.
3. HM samples are usually collected by lactating mothers and not by the hospital staff. Hence, the staff must provide detailed instructions to the mothers to avoid sampling bias and procedural differences between the different HM samples collected within a study. We recommend establishing a Standard Operation Procedure for the collection of milk samples and their processing to avoid experimental bias between samples. Details regarding the sample collection procedure need to be indicated, including, e.g., extraction of foremilk, hindmilk, or full expression of one or both breasts; time of the day; time since the last feeding; and extraction method (i.e., manual expression, milk pump).

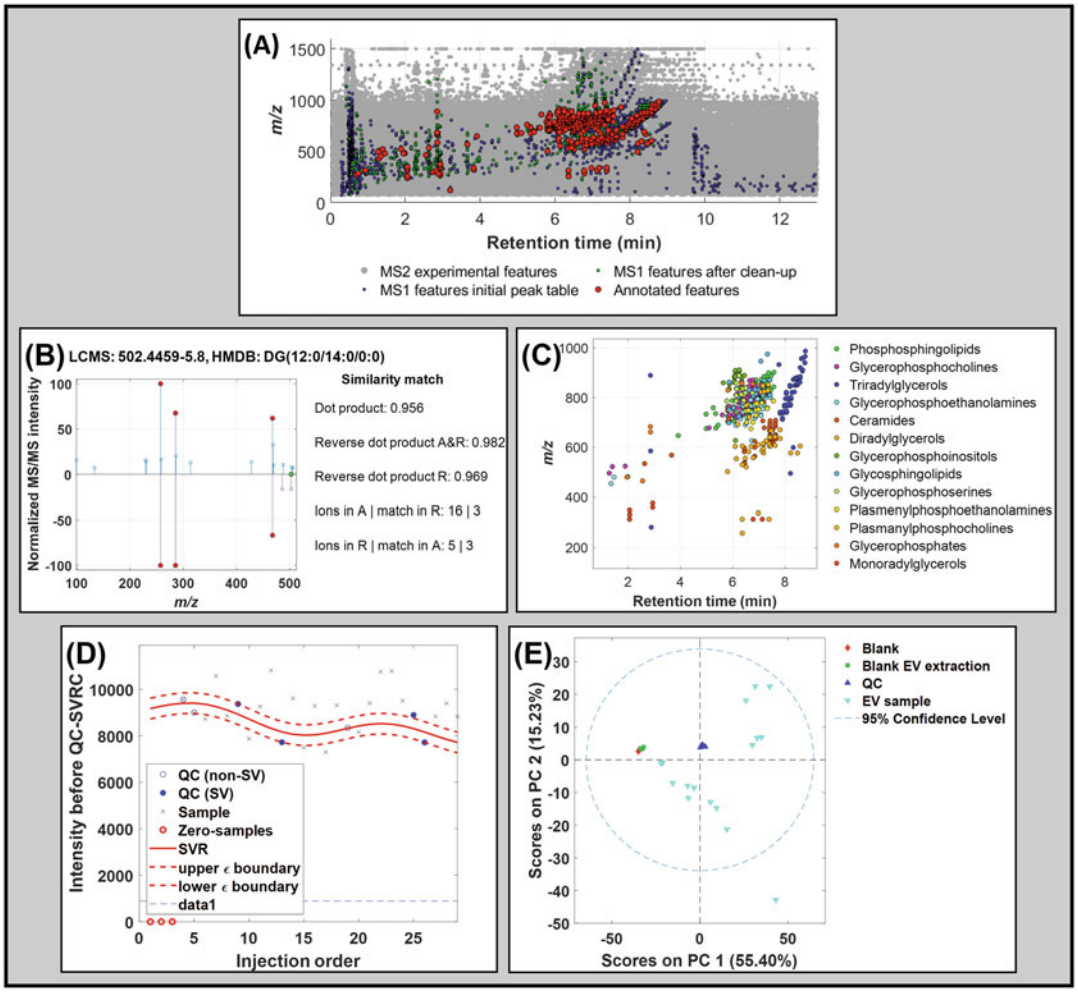

Fig. 2 LC–MS lipidomic data processing. (**a**) Distribution of MS1 (blue dots) and MS2 (gray dots) experimental features measured in extracellular vesicles (EVs) from human milk (HM) samples in the *m/z*-retention time space and MS1 experimental features and annotated features after applying cleanup steps (green and red dots, respectively). (**b**) Similarity match of a feature (*m/z*-retention time: 502.4489–5.8) annotated as DG (12:0/14:0/0:0) from the HMDB database. MS2 experimental spectrum (up). MS2 database spectrum (down). (**c**) Distribution in the *m/z*-retention time space of annotated features by sub-classes in EVs from HM samples. Only sub-classes containing at least five features are represented. (**d**) Intensity of a selected feature as a function of the injection order before intra-batch effect correction with the quality control-supported vector regression correction (QC-SVRC) approach. (**e**) Score plot of the principal component analysis model using the preprocessed data set. Note: *DG* diradylglycerol, *QC* quality control

4. We recommend the use of electric milk pumps as indicated in this protocol; however, if participants prefer manual expression, this could also be accommodated within the protocol. Milk pumps can be sterilized in the microwave by using dedicated plastic bags or in a water bath (15 min at 100 °C).

All parts need to be dried with sterile gauzes. The milk pump should be cleaned with hot water and soap immediately after every use.

5. No ointments should be applied to the skin before HM extraction. If they had been applied, clean the skin carefully with water and soap.
6. Take note of date, time, and extracted HM volume. When HM samples are manipulated, gloves must be used in order to avoid sample contamination.
7. Remove the upper fat layer with a spatula and aspirate the supernatant with a pipette. Be careful not to aspirate the remaining fat.
8. If the tubes do not have the same weight, compensate by adding PBS.
9. Pellet should not be aspirated.
10. More than one filter may be needed due to obturation.
11. Reconstitute the pellet in 1 mL of PBS and add the remaining 24 mL when the pellet is dissolved.
12. Do it gently to avoid foam formation.
13. The order of the solvents' addition is important.
14. This step usually takes 3–4 h.
15. Basic considerations for untargeted, LC–MS-based lipidomics and metabolomics studies should be implemented throughout the pipeline. For further information, the reader is referred to [17].
16. A set of approximately 5–10 QCs should be injected at the beginning of each batch for system conditioning and MS/MS data acquisition. MS/MS spectra need to be acquired for subsequent annotation purposes as described elsewhere [16]. Depending on the employed MS system, parameters need to be adjusted. For this example, a rate of 5 spectra/s in the extended dynamic range mode (2 GHz), a collision energy set to 20 V, an automated selection of five precursor ions per cycle, and an exclusion window of 0.15 min after two consecutive selections of the same precursor were used.
17. Samples should be injected in randomized order and a QC sample should be intercalated every 5–10 samples for subsequent correction of the batch effect [17–19]. In our pipeline, we use a binary mobile phase gradient starting at 98% of solution A (5:1:4 IPA:CH_3OH:H_2O 5 mM CH_3COONH_4, 0.1% v/v FA) for 0.5 min followed by a linear gradient from 2 to 20% of solution B (99:1 IPA:H_2O 5 mM CH_3COONH_4, 0.1% v/v FA) for 3.5 min and from 20 to 95% v/v of solution B in 4 min; 95% v/v of solution B are

maintained for 1 min; return to initial conditions in 0.25 min and maintain for a total run time of 13 min. Column and autosampler are kept at 55 and 4 °C, respectively, an injection volume of 2 μL is used, and the flow rate is set to 400 μL/min. Full scan MS data in the range between 70 and 1500 *m/z* are acquired at a scan frequency of 5 Hz using the following parameters: gas T, 200 °C; drying gas, 14 L/min; nebulizer, 37 psi; sheath gas T, 350 °C; sheath gas flow, 11/min. Mass reference standards are introduced into the source for on-the-fly automatic MS spectra recalibration during analysis via a reference sprayer valve using the 149.02332 (phthalic anhydride), 121.050873 (purine), and 922.009798 (HP-0921) *m/z* in ESI+, and 119.036 (purine) and 980.0163 (HP-0921, [M-H + CH_3COOH]-) *m/z* in ESI-, as references. ESI+ and ESI- analyses were carried out in independent batches, and between them, the instrument was cleaned and calibrated according to the manufacturer's guidelines.

18. Blank injections can affect the performance of the LC–MS system and hence, the position of the blanks within the analytical sequence should be chosen carefully [20].
19. We use MassHunter Workstation (version B.07.00) from Agilent for UPLC-TOFMS data acquisition and manual integration. In our workflow, peak detection, integration, deconvolution, alignment, and pseudospectra identification are carried out using XCMS [21] and CAMERA [22] in R 3.6.1. See reference [11] for an example of XCMS and CAMERA settings. Employed parameters depend on the LC–MS system and the method's performance and might need to be optimized in advance using, e.g., Isotopologue Parameter Optimization (IPO) [23]. Alternatively, peak tables can be generated using vendor and other available software such as MZmine2 or MS-DIAL. We recommend comparing manual vs. automatic integration for IS and endogenous metabolites considered in **step 1** to assess the accuracy of the automated integration.
20. If ESI+ and ESI- data were acquired in separate injections, perform data cleanup and filtering independently.
21. Metabolite annotation (Level ID: 2, putatively annotated compounds without matching to data for chemical standards acquired under the same experimental conditions) can be performed by matching experimentally acquired MS/MS spectra with the experimental MS/MS databases (e.g., HMDB, METLIN) in accordance with the Metabolomics Standards Initiative (MSI) reporting standards [24]. Metabolite annotation can also exploit in-silico databases (e.g., LipidBlast [25]). Annotation and subsequent correction and cleanup steps can

be programmed in different programming languages for data science such as MATLAB (Mathworks Inc., Natick, MA, USA), R, or Python. Besides, metabolite annotation can also be performed using vendor software and other widely used tools such as MZMine2, MS-DIAL, GNPS (Global Natural Products Social Molecular Networking), or LipidDex [16, 26].

22. Features detected in QC samples can efficiently be used to monitor the instrument performance and correct within-batch effects [17]. We developed a nonparametric approach based on the QC-Supported vector regression correction (SVRC) approach employing a Radial Basis Function kernel for within batch effect removal [18, 19]. The selection of the tolerance threshold (ε), the penalty term applied to margin slack values (C), and the kernel width (γ) can be carried out using a grid search. QC-SVRC is robust to hyperparameter selection and a pre-selection of C and optimization of ε and γ using a grid search by leave-one-out cross-validation, and using the root mean square error of cross-validation (RMSECV) as target function allows a significant reduction in computation time. C can be selected for each LC–MS feature as the median value of the intensities observed in QC replicates. The ε search range is chosen based on the expected instrumental precision (e.g., 5–10% of the signal intensity). The γ search interval can be set to $[1, 10^5]$.
23. It is important to remove variables that are likely originating from background contaminants or cross-contamination, as well as those from unstable features to enhance performance of subsequent data analysis steps. We typically remove features with more than 20% missing values in QCs, those with RSD (QC)>20% after within-batch effect correction, and those for which the ratio between the median peak area values in QCs and blanks is lower than 6.
24. Different qualitative and quantitative tools are available for checking data quality [27], including principal component analysis, which is the most widely employed tool for this purpose. We recommend comparing the data quality of raw data and preprocessed data.

References

1. Kalluri R, LeBleu VS (2020) The biology, function, and biomedical applications of exosomes. Science 367
2. Altadill T, Campoy I, Lanau L et al (2016) Enabling metabolomics based biomarker discovery studies using molecular phenotyping of exosome-like vesicles. PLoS One 11: e0151339
3. Zempleni J, Aguilar-Lozano A, Sadri M et al (2017) Biological activities of extracellular vesicles and their cargos from bovine and human milk in humans and implications for infants. J Nutr 147:3–10
4. Liao Y, Du X, Li J et al (2017) Human milk exosomes and their microRNAs survive

digestion in vitro and are taken up by human intestinal cells. Mol Nutr Food Res 61: 1700082

5. Kahn S, Liao Y, Du X et al (2018) Exosomal MicroRNAs in milk from mothers delivering preterm infants survive in vitro digestion and are taken up by human intestinal cells. Mol Nutr Food Res 62:1701050
6. Yang M, Song D, Cao X et al (2017) Comparative proteomic analysis of milk-derived exosomes in human and bovine colostrum and mature milk samples by iTRAQ-coupled LC-MS/MS. Food Res Int 92:17–25
7. de la Torre Gomez C, Goreham RV, Bech Serra JJ et al (2018) "Exosomics"—a review of biophysics, biology and biochemistry of exosomes with a focus on human breast Milk. Front Genet 9:92
8. Royo F, Gil-Carton D, Gonzalez E et al (2019) Differences in the metabolite composition and mechanical properties of extracellular vesicles secreted by hepatic cellular models. J Extracell Vesicles 8:1575678
9. Clos-Garcia M, Loizaga-Iriarte A, Zuñiga-Garcia P et al (2018) Metabolic alterations in urine extracellular vesicles are associated to prostate cancer pathogenesis and progression. J Extracell Vesicles 7:1470442
10. Haraszti RA, Didiot M-C, Sapp E et al (2016) High-resolution proteomic and lipidomic analysis of exosomes and microvesicles from different cell sources. J Extracell Vesicles 5:32570
11. Ramos-Garcia V, Ten-Doménech I, Moreno-Giménez A et al (2021) ATR-FTIR spectroscopy for the routine quality control of exosome isolations. Chemom Intell Lab Syst 104401
12. Zonneveld MI, Herwijnen MJC, Fernandez-Gutierrez MM et al (2021) Human milk extracellular vesicles target nodes in interconnected signalling pathways that enhance oral epithelial barrier function and dampen immune responses. J Extracell Vesicles:10
13. Witwer KW, Buzás EI, Bemis LT et al (2013) Standardization of sample collection, isolation and analysis methods in extracellular vesicle research. J Extracell Vesicles 2:20360
14. Ten-Doménech I, Ramos-Garcia V, Piñeiro-Ramos JD et al (2020) Current practice in untargeted human milk metabolomics. Meta 10:43
15. Piñeiro-Ramos JD, Parra-Llorca A, Ten-Doménech I et al (2020) Effect of donor human milk on host-gut microbiota and metabolic interactions in preterm infants. Clin Nutr
16. Ten-Doménech I, Martínez-Sena T, Moreno-Torres M et al (2020) Comparing targeted vs. untargeted MS2 data-dependent acquisition for peak annotation in LC-MS metabolomics. Metabolites 10:126
17. Broadhurst D, Goodacre R, Reinke SN et al (2018) Guidelines and considerations for the use of system suitability and quality control samples in mass spectrometry assays applied in untargeted clinical metabolomic studies. Metab Off J Metab Soc 14:72
18. Kuligowski J, Sánchez-Illana Á, Sanjuán-Herráez D et al (2015) Intra-batch effect correction in liquid chromatography-mass spectrometry using quality control samples and support vector regression (QC-SVRC). Analyst 140:7810–7817
19. Sánchez-Illana Á, Pérez-Guaita D, Cuesta-García D et al (2018) Model selection for within-batch effect correction in UPLC-MS metabolomics using quality control – support vector regression. Anal Chim Acta 1026:62–68
20. Martínez-Sena T, Luongo G, Sanjuan-Herráez D et al (2019) Monitoring of system conditioning after blank injections in untargeted UPLC-MS metabolomic analysis. Sci Rep 9: 1–9
21. Smith CA, Want EJ, O'Maille G et al (2006) XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. Anal Chem 78:779–787
22. Kuhl C, Tautenhahn R, Böttcher C et al (2012) CAMERA: an integrated strategy for compound spectra extraction and annotation of liquid chromatography/mass spectrometry data sets. Anal Chem 84:283–289
23. Libiseller G, Dvorzak M, Kleb U et al (2015) IPO: a tool for automated optimization of XCMS parameters. BMC Bioinf 16:118
24. Salek RM, Steinbeck C, Viant MR et al (2013) The role of reporting standards for metabolite annotation and identification in metabolomic studies. GigaScience 2
25. Kind T, Liu K-H, Lee DY et al (2013) LipidBlast in silico tandem mass spectrometry database for lipid identification. Nat Methods 10: 755–758
26. Hutchins PD, Russell JD, Coon JJ (2018) LipiDex: an integrated software package for high-confidence lipid identification. Cell Syst 6:621-625.e5
27. Sánchez-Illana Á, Piñeiro-Ramos JD, Sanjuan-Herráez JD et al (2018) Evaluation of batch effect elimination using quality control replicates in LC-MS metabolite profiling. Anal Chim Acta 1019:38–48

Chapter 19

Mass Spectrometry-Based Untargeted Metabolomics and Lipidomics Platforms to Analyze Cell Culture Extracts

Elias Iturrospe, Katyeny Manuela da Silva, Maria van de Lavoir, Rani Robeyns, Matthias Cuykx, Tamara Vanhaecke, Alexander L. N. van Nuijs, and Adrian Covaci

Abstract

Metabolites represent the most downstream level of the cellular organization. Hence, an in vitro untargeted metabolomics approach is extremely valuable to deepen the understanding of how endogenous metabolites in cells are altered under a given biological condition. This chapter describes a robust liquid chromatography–high-resolution mass spectrometry-based metabolomics and lipidomics platform applied to cell culture extracts. The analytical workflow includes an optimized sample preparation procedure to cover a wide range of metabolites using liquid–liquid extraction and validated instrumental operation procedures with the implementation of comprehensive quality assurance and quality control measures to ensure high reproducibility. The lipidomics platform is based on reversed-phase liquid chromatography for the separation of slightly polar to apolar metabolites and covers a broad range of lipid classes, while the metabolomics platform makes use of two hydrophilic interaction liquid chromatography methods for the separation of polar metabolites, such as organic acids, amino acids, and sugars. The chapter focuses on the analysis of cultured HepaRG cells that are derived from a human hepatocellular carcinoma; however, the sample preparation and analytical platforms can easily be adapted for other types of cells.

Key words Untargeted metabolomics, Untargeted lipidomics, High-resolution mass spectrometry, Cell culture extracts, HepaRG cells, Hydrophilic interaction liquid chromatography, Reversed-phase liquid chromatography, Quality assurance and quality control

1 Introduction

The metabolome consists of a heterogeneous variety of molecules, ranging from small polar organic acids to apolar triglycerides. The complexity of the metabolome is reflected in the 220,945 metabolite entries listed in the Human Metabolome Database (HMDB) as of October 2021 [1]. In addition, Alseekh et al. estimated that up to 1,000,000 different metabolites occur across living organisms, with approximately up to 40,000 metabolites in a single species

Raúl González-Domínguez (ed.), *Mass Spectrometry for Metabolomics*,
Methods in Molecular Biology, vol. 2571, https://doi.org/10.1007/978-1-0716-2699-3_19,
© The Author(s), under exclusive license to Springer Science+Business Media, LLC, part of Springer Nature 2023

[2]. Untargeted metabolomics aims to detect a wide range of endogenous metabolites in a reproducible and unbiased manner to create hypotheses on how metabolic pathways are altered under a given biological condition [3]. To cover this vast array of metabolites, comprehensive sample preparation combined with highly sensitive and specific analytical methods using complementary techniques are often required [4]. To increase metabolite coverage, samples can be fractionated during sample preparation in order to analyze the fractions by different analytical platforms, increasing specificity [5, 6].

Liquid chromatography (LC) coupled to high-resolution mass spectrometry (HRMS) remains one of the most widely used hyphenated techniques in metabolomics, with a constantly increasing number of applications [7]. This latter is due to, for instance, the evolution of the development of chromatographic stationary phases (e.g., improved particle size, stability, and selectivity), the increase in resolution of mass spectrometry instrumentation (e.g., resolving power at full width at half maximum of an Orbitrap can reach up to 240,000 at *m/z* 400), and the LC–MS versatility in being able to be coupled to additional techniques (e.g., ion mobility spectrometry) [5, 6, 8]. Moreover, hyphenating LC to MS reduces co-elution and ion suppression, hence improving metabolite coverage and the quality of detected features (i.e., an entity with assigned mass/charge ratio (*m/z)* and retention time (RT) associated with a response signal (peak intensity or area)). Next to MS/MS fragmentation spectra, the retention time behavior of lipids obtained by reversed-phase separation offers additional information for annotation [9].

The analytical platforms presented in this chapter are a combination of hydrophilic interaction liquid chromatography high-resolution mass spectrometry (HILIC-HRMS) and reversed-phase liquid chromatography high-resolution mass spectrometry (RPLC-HRMS) combined with electrospray ionization in positive and negative modes (ESI (+) and ESI (−), respectively) (Fig. 1). The lipidomics platform uses RPLC to separate slightly polar to apolar metabolites and covers a broad range of lipid classes [6, 10]. The metabolomics platform consists of two HILIC methods, which can separate polar metabolites, such as organic acids, amino acids, and sugars that show almost no retention or co-elute on reversed-phase LC columns [5, 10]. The implementation of a quality management system (QMS) is imperative to ensure the collection of high-quality data and, subsequently, reliable data interpretation [11]. Quality assurance and quality control (QA/QC) measures are performed before, during, and after the analytical workflow, such as the use of a chromatographic column inventory, system suitability samples with documented performance over time, isotopically labeled internal standards (IS), pooled QC samples, and the use of extraction blanks.

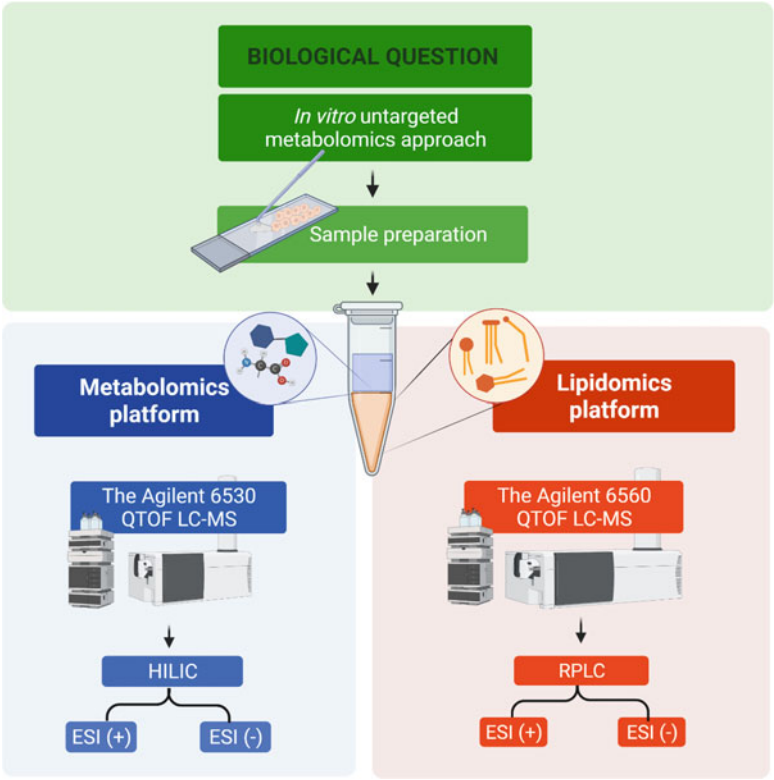

Fig. 1 Overview of the metabolomics and lipidomics platform

Herein, this chapter describes an LC–MS-based metabolomics and lipidomics platform applied to cell culture extracts. First, the major steps in the analytical workflow are introduced, including sample preparation, instrumental operation, and QA/QC. Compared to clinical in vivo studies, in vitro cell metabolomics enable mechanistic elucidation at the cellular level, show less ethical restrictions, and enable easy access to the samples (e.g., human hepatic cell lines vs patient liver biopsies). In addition, usage of cell lines reduces unwanted variations introduced by, e.g., inter-individual differences in diet or age. The chapter is dedicated to the analysis of cultured HepaRG cells (a cell line derived from a human hepatocellular carcinoma); however, the platforms described in detail here can be adapted for the use of other cell types.

2 Materials

2.1 Solutions and Mixtures for Sample Preparation and Instrumental Analysis

1. Quenching solution: 10 mM ammonium acetate in MeOH/H_2O (80/20, v/v). Store at −80 °C.
2. Solution for the polar fraction of the liquid–liquid extraction (LLE) (*see* **Note 1**): 0.5 mM L-ascorbic acid and 1 mM EDTA diammonium salt in 5 mM ammonium acetate with 0.1% (v/v) acetic acid.

3. Solution for the apolar fraction of the LLE (*see* **Note 1**): 1 mM butylated hydroxytoluene in $CHCl_3$ (*see* **Note 2**).
4. Internal standard solution for the polar fraction: 7 μg/mL of L-lysine-$^{13}C_6$-$^{15}N_2$, L-phenylalanine-$^{13}C_9$-^{15}N, caffeine-$^{13}C_3$ and hippuric acid-(phenyl-$^{13}C_6$) in H_2O/MeOH (1/1, v/v) to obtain a final concentration of 1 μg/mL in the extracts of HepaRG cell samples (*see* **Note 3**).
5. Internal standard solution for the apolar fraction: 11 μg/mL of L-octanoyl-carnitine N-methyl-D_3, cholic acid-2,2,4,4-D_4, lysophosphatidylethanolamine 18:1-D_7, and ceramide (d18:1/18:1(9Z)-$^{13}C_{18}$) in $CHCl_3$ to obtain a final concentration of 1 μg/mL in the extracts of HepaRG cell samples (*see* **Note 3**).
6. Reconstitution solvent HILIC-method: MeCN/H_2O (65/35, v/v). Store at 4 °C (*see* **Note 4**).
7. Reconstitution solvent RPLC-method: MeOH/IPA (65/35, v/v). Store at −20 °C.
8. Blank solvent samples: Transfer reconstitution solvent to a glass HPLC screw-top vial with a PTFE silicone cap.
9. System suitability (SS) solution HILIC-methods: 1 μg/mL 2-octanoyl-L-carnitine, nicotinic acid, L-tryptophan, L-lysine, L-arginine, and thymine in MeCN/H_2O (65/35, v/v) (*see* **Note 5**).
10. SS solution RPLC-methods: 1 μg/mL 2-octenoyl-L-carnitine, glycerol tripalmitate-[$^{13}C_3$], 1,2-dipalmitoyl-sn-glycero-3-phosphoethanolamine, 1-stearoyl-phosphatidylethenolamine, lithocholic acid, and 1,2-dipalmitoyl-rac-glycero-3-phosphoinositol in MeOH/IPA (65/35, v/v) (*see* **Note 5**).
11. Mobile phases (*see* **Note 6**):
    (a) HILIC-ESI (+) mobile phase A: 10 mM ammonium formate and 0.1% (v/v) formic acid in H_2O/MeOH (90/10, v/v) (pH aqueous fraction: 3.5).
    (b) HILIC-ESI (+) mobile phase B: MeCN.
    (c) HILIC-ESI (−) mobile phase A: 2 mM ammonium carbonate with 2 mM ammonium acetate in H_2O (pH 8.20).
    (d) HILIC-ESI (−) mobile phase B: MeCN/MeOH (90/10, v/v).
    (e) RPLC- ESI (+) mobile phase A: MeCN/5 mM ammonium acetate and 0.1% (v/v) acetic acid in H_2O (30/70, v/v) (pH aqueous fraction: 4.2).
    (f) RPLC-ESI (+) mobile phase B: 5 mM ammonium acetate and 0.1% (v/v) acetic acid in H_2O/MeCN/IPA (2/10/88, v/v/v) (pH aqueous fraction: 4.2).

(g) RPLC-ESI (−) mobile phase A: MeCN/5 mM ammonium acetate in H_2O (30/70, v/v) (pH aqueous fraction: 6.5).

(h) RPLC-ESI (−) mobile phase B: 5 mM ammonium acetate in H_2O/MeCN/IPA (2/10/88, v/v/v) (pH aqueous fraction: 6.5).

2.2 Solutions and Mixtures for Cell Culturing

1. Collagen solution for coating of chamber slides.
   (a) 0.02 M acetic acid solution in sterile H_2O. Filter the solution through a 0.22-μm filter (*see* **Note 7**).
   (b) 0.1 mg/mL solution of rat tail collagen (type I) in 0.02 M acetic acid (*see* **Notes 7** and **8**).
2. HepaRG Thaw, Seed, and General-Purpose Medium (*see* **Note 9**).
   (a) Prewarm Basal Hepatic Medium (BHM) and HepaRG Thaw, Seed, and General-Purpose supplement (Biopredic International, Rennes, France) to 37 °C.
   (b) Vortex the supplement and pipette the supplement to the BHM. Ensure a homogeneous solution by pipetting up and down (*see* **Note 7**).
3. HepaRG Maintenance and Metabolism Medium (*see* **Note 9**).
   (a) Prewarm BHM and HepaRG Maintenance and Metabolism supplement (Biopredic International) to 37 °C.
   (b) Vortex the supplement and pipette the supplement to the BHM. Ensure a homogeneous solution by pipetting up and down (*see* **Note 7**).

2.3 LC–MS System

1. LC–MS system for the HILIC methods: Agilent 1290 Infinity UPLC system coupled to an Agilent 6530 QTOF MS system with Agilent Jet Stream Electrospray Ionization (ESI) (Agilent Technologies, Santa Clara, USA).
2. Column HILIC ESI (+) method: HILICON iHILIC-Fusion (SS) (100 × 2.1 mm, 1.8 μm) with HILICON iHILIC-Fusion (SS) (20 × 2.1 mm, 1.8 μm) guard column (Hilicon AB, Sweden).
3. Column HILIC ESI (−) method: HILICON iHILIC-Fusion (P) (PEEK) (100 × 2.1 mm, 5 μm) with HILICON iHILIC-Fusion (P) (PEEK) (20 × 2.1 mm, 5 μm) guard column.
4. LC–MS System RPLC method: Agilent 1290 Infinity II UPLC system coupled to an Agilent 6560 drift tube-ion mobility-QTOF MS using Agilent Dual Jet Stream ESI.
5. Column RPLC ESI (+) and ESI (−) method: ACQUITY Premier BEH C18 (150 × 2.1 mm, 1.7 μm) with VanGuard FIT cartridge (5 × 2.1 mm, 1.7 μm) (Waters, Massachusetts, USA).

2.4 Additional Materials and Equipment for Sample Preparation

1. Polypropylene slide mailers for storage and/or transportation of chamber slides after cultivation.
2. Adjustable single-channel pipettes and pipette tips (200 μL, 1000 μL, 10 mL).
3. Benchtop centrifuge.
4. Cell scrapers.
5. Wet ice.
6. Dry ice.
7. Liquid nitrogen (N_2) in a Nalgene container.
8. Wooden pincers.
9. Mobile phase filtration assembly.
10. 0.2-μm cellulose acetate membrane filters to filter the aqueous fraction of the mobile phases.
11. Ultrasonic bath.
12. Vortex mixer.
13. 1.5-mL Safe-Lock tubes.
14. 0.2-μm centrifugal nylon filters.
15. 0.3-mL Reacti-Vials with open-top screw caps and PTFE/rubber discs.
16. 2-mL amber HPLC screw-top vials with PTFE silicone caps and inserts.
17. 2-mL clear HPLC screw-top vials with PTFE silicone caps as LLE vials.
18. 384-well polypropylene plates.
19. Aluminum sealing tape for microplates.
20. Glass syringes for chloroform handling (airtight, 500 μL and 5 mL).
21. Pure, dry nitrogen for solvent evaporation.
22. Nitrogen evaporator.
23. Fridge (4 °C) and freezer (−20 °C and −80 °C).
24. Amber flasks of 50 mL or 100 mL.

2.5 Additional Materials and Equipment for Cell Culturing

1. Differentiated HepaRG cells (1×10^6 cells per sample, Biopredic International). Usage of other cell types requires a pilot study to optimize the number of cells and/or dilution factor used during sample preparation (*see* **Note 10**).
2. Chamber slides for cell cultivation (e.g., Permanox 2-well Lab-Tek).
3. Petri dishes that fit one or more chamber slides.
4. LAF cabinet (class II or higher).
5. Laboratory water bath.

6. Cell incubator.
7. Electronic pipette aid and compatible graduated sterile transfer pipets (5 mL, 10 mL, 25 mL, 50 mL).
8. Adjustable single-channel pipettes and pipette tips (200 μl, 1000 μL, 10 mL).
9. Sterile phosphate-buffered saline (PBS).
10. Falcon tubes (50 mL).
11. Benchtop centrifuge.
12. Vacuum aspirator.
13. Vortex mixer.
14. Fridge (4 °C).
15. Wet ice.
16. Light microscope.
17. Sterile syringe with 0.2-μm filter adapter.

3 Methods

3.1 Incubation and Exposure

1. Coating of chamber slides.
   (a) When working with Permanox 2-well Lab-Tek chamber slides, pipette 750 μL of the 0.1 mg/mL collagen solution to one well of each chamber slide (*see* **Notes 7** and **11**).
   (b) Incubate the chamber slides with collagen within the LAF cabinet for at least 1.5 h.
   (c) Remove the uncoated collagen and wash the coated wells three times using sterile PBS (*see* **Note 7**).
2. Seeding of HepaRG cells.
   (a) Transfer 9 mL of the HepaRG Thaw, Seed, and General Purpose Medium to a Falcon tube and prewarm to 37 °C (*see* **Note 7**).
   (b) Thaw a cryovial of differentiated HepaRG cells to 37 °C and pipette the cell suspension gently up and down ten times before transferring the cell suspension to the Falcon tube with the medium (*see* **Notes 7**, **12**, and **13**).
   (c) Rinse the cryovial with 1 mL of HepaRG Thaw, Seed, and General Purpose Medium (37 °C) and transfer the medium to the same Falcon tube (*see* **Note 7**).
   (d) Centrifuge the Falcon tube containing the HepaRG cell suspension for 2 min at 357 *g* and aspirate the supernatant to obtain a HepaRG cell pellet (*see* **Note 7**).
   (e) Resuspend the HepaRG cell pellet in 1 mL of HepaRG Thaw, Seed, and General Purpose Medium (37 °C) (*see* **Note 7** and **13**).

(f) Calculate the volume of HepaRG Thaw, Seed, and General Purpose Medium that needs to be added to 1 mL of the HepaRG cell suspension to obtain a concentration of 0.86×10^6 cells/mL (*see* **Note 14**).

(g) Add the calculated volume of HepaRG Thaw, Seed, and General Purpose Medium (37 °C) to 1 mL of the HepaRG cell suspension (*see* **Notes 7** and **13**).

(h) Gently pipette 1.2 mL of the obtained cell suspension to the coated wells of the chamber slides. Make sure to distribute the cells equally along the surface by gently swirling (*see* **Notes 7** and **15**).

(i) To obtain extraction blanks, pipette 1.2 mL of HepaRG Thaw, Seed, and General Purpose Medium (37 °C) without cells to the coated wells of the chamber slides (*see* **Note 16**).

(j) Put the chamber slides inside Petri dishes (*see* **Note 7**).

3. Incubation of HepaRG cells.

(a) Incubate the samples at 37 °C with 5% CO_2 and saturated humidity.

(b) After 24 h of incubation, aspirate the old media from the cells and extraction blanks and add 1.2 mL of HepaRG Maintenance and Metabolism Medium (37 °C) to each well containing cells or extraction blanks (*see* **Note 7**).

(c) Incubate the samples at 37 °C with 5% CO_2 and saturated humidity.

(d) Replace the old medium with new HepaRG Maintenance and Metabolism Medium (37 °C) after 48 h and 120 h (*see* **Note 7**).

(e) Monitor the HepaRG cell cultures using microscopic imaging. Discard contaminated cultures.

4. Exposure of HepaRG cells.

(a) Dissolve the compound(s) of interest in HepaRG Maintenance and Metabolism Medium (37 °C) at the concentration of interest. Filter the solution through a 0.2-μm filter (*see* **Notes 7** and **17**).

(b) Aspirate the old media from the cells and extraction blanks and add 1.2 mL of HepaRG Maintenance and Metabolism Medium containing the compound(s) of interest (37 °C) to each well containing cells or extraction blanks (*see* **Notes 7** and

(c) **18**).

(d) Incubate the samples at 37 °C with 5% CO_2 and saturated humidity for the desired exposure time.

3.2 Sample Preparation

1. Prepare all solutions needed for the sample preparation as described in Subheading 2.1.
2. Calculated for ten samples, transfer 5 mL of organic extraction mixture to an amber flask and add 238 μL of the internal standard solution for the apolar fraction. Vortex the mixture for 30 s.
3. Transfer 6 mL of the aqueous extraction mixture to a second amber flask and add 240 μL of the internal standard solution for the polar fraction. Vortex the mixture for 30 s.
4. Add 440 μL of the organic extraction mixture (with IS) and 520 μL of the aqueous extraction mixture (with IS) to each LLE vial.
5. Store at −20 °C until the moment of extraction.

3.2.1 Extraction

1. Obtain microscopic images of the different exposure groups.
2. After incubation of HepaRG cells in collagen-coated chamber slides (1×10^6 cells/well), wash the chamber slides twice using PBS (37 °C) (*see* **Note 7**) before snap-freezing with liquid N_2 (−196 °C). Use a wooden pincer to hold the chamber slide in the liquid N_2 for at least 20 s (*see* **Note 19**). Collect extraction blanks the same way as the samples (*see* **Note 16**).
3. Put each chamber slide in a slide mailer and dig the slide mailer in dry ice until all samples are collected.
4. Add 300 μL of the quenching solution to the frozen culture on the chamber slide on dry ice. Dispense the solution evenly over the cultured zone (*see* **Note 20**).
5. After 2 min, scrape the cells and transfer them to the LLE vial.
6. Add another 300 μL of the quenching solution to rinse the chamber slide and collect the solution in the same LLE vial.
7. Vortex the LLE vial for 90 s, equilibrate for 10 min on wet ice, centrifuge at 2200 *g* for 7 min, and equilibrate again for 10 min on wet ice.
8. Transfer 900 μL of the polar fraction (upper phase) to a safe-lock tube without transferring solid particles from the protein layer. Vortex for 20 s and transfer 450 μL from this safe-lock tube to another safe-lock tube, creating two aliquots (*see* **Note 21**).
9. Transfer 240 μL of the apolar fraction (lower phase) to a glass Reacti-Vial using a glass syringe. Vortex for 20 s and transfer 120 μL from this Reacti-Vial to another Reacti-Vial (*see* **Notes 21** and **22**).
10. Evaporate the samples using a stream of nitrogen at room temperature.
11. Store the dried extracts at −80 °C and reconstitute them immediately before analysis.

3.2.2 Reconstitution of Dried Extracts

1. Polar fraction: Reconstitute each sample on wet ice using 60 μL of MeCN/H_2O (65/35, v/v) and vortex for 40 s.
2. Apolar fraction: Reconstitute each sample on wet ice using 60 μL of MeOH/IPA (65/35, v/v) and vortex for 40 s.
3. Filter samples using 0.2-μm nylon centrifugal filters and centrifugation at 14,000 *g* for 2 min. Vortex all filtered samples for 40 s.
4. Transfer 10 μL of each sample (except for extraction blanks) to an amber HPLC screw top vial and vortex for 40 s to obtain a QC pool (*see* **Note 23**).
5. Transfer 20 μL of each reconstituted sample to a 384-well plate and seal the plate using aluminum adhesive.

3.3 LC–MS

1. Specific LC parameters are described in Table 1 (*see* **Note 24**).
2. Specific MS and MS/MS parameters are described in Table 2.
3. The reference masses 922.0098 *m/z* and 121.0508 *m/z* for ESI (+) and 966.0007 *m/z* and 119.0363 *m/z* for ESI (−) are constantly infused with an additional isocratic pump, used to correct the mass axis during acquisition (*see* **Note 25**).
4. Prepare an acquisition worklist in a format similar to the sequence table shown in Fig. 2. Randomize samples to prevent bias related to injection order.
5. Equilibrate the column using the mobile phase start composition (preferably without modifiers, *see* **Note 26**). Inject blank solvent samples after the column pressure baseline is stable for at least 10 min. The pressure profiles of these blanks should be aligned when the column is sufficiently equilibrated.
6. Before data acquisition of the test and QC samples, inject the system suitability sample (Table 3 and Fig. 3) and compare mass accuracy, peak height, peak area, and RT with the acceptance criteria (*see* **Note 5**).
7. Inject pooled QC samples to condition the LC–MS system (*see* **Note 27**). During conditioning injections, acquire MS/MS data of the pooled QC sample (Table 2).
8. Inject one or more extraction blanks after conditioning the system. Acquire data in MS1 mode during these injections (see Table 1). After injection of extraction blanks, the system needs to be conditioned again using the QC pooled sample (*see* **Note 27**).
9. Inject the samples according to the sequence table.
   (a) Inject a pooled QC sample minimum six times at regular intervals across the worklist. Acquire data in MS1 mode (Table 1).

Table 1
Liquid chromatography parameters of the HILIC and RPLC-methods

	HILIC ESI (+)	HILIC ESI (−)	RPLC ESI (+)	RPLC ESI (−)
Chromatographic column	HILICON iHILIC-Fusion	HILICON iHILIC-Fusion(P)	ACQUITY UPLC BEH C18	ACQUITY UPLC BEH C18
Column material	Stainless steel	Polyether ether ketone	Stainless steel	Stainless steel
Column dimensions (mm)	100 × 2.1	100 × 2.1	150 × 2.1	150 × 2.1
Column particle size (μm)	1.8	5	1.7	1.7
Mobile phase A	H_2O/MeOH (90/10, v/v)	H_2O	MeCN/H_2O (30/70, v/v)	MeCN/H_2O (30/70, v/v)
Mobile phase B	MeCN	MeCN/MeOH (90/10, v/v)	H_2O/MeCN/IPA (2/10/88, v/v/v)	H_2O/MeCN/IPA (2/10/88, v/v/v)
Modifier	10 mM ammonium formate +0.1% (v/v) formic acid	2 mM ammonium carbonate +2 mM ammonium acetate	5 mM ammonium acetate +0.1% (v/v) acetic acid	5 mM ammonium acetate
Flow rate (mL/min)	0.25	0.20	0.20	0.20
Temperature (°C)	60	Room temperature[a]	60	60
Gradient	Min–%B 0–95 4–95 12.5–60 20–60 21–95 26–95	Min–%B 0–95 1–95 10–20 14–20 15–95 20–95	Min–%B 0–15 2–15 3–30 5–60 8–60 20–100 30–100 35–15 40–15	Min–%B 0–15 2–15 3–30 5–60 8–60 20–100 30–100 35–15 40–15
Total run time (min)	26	20	40	40
Time segments (min)	0–0.5: Waste 0.5–22: MS 22–26: Waste	0–0.5: Waste 0.5–19.5: MS 19.5–20: Waste	0–0.5: Waste 0.5–30: MS 30–40: Waste	0–0.5: Waste 0.5–30: MS 30–40: Waste
Void time (t_0) (min)	0.7[b]	1.4[b]	1.1[c]	1.1[c]
Injection volume (μL)	3	3	2	2

[a]The heat exchanger is bypassed for the HILIC ESI (−) method to reduce chelating interactions of anionic metabolites with trace metals from the respective hardware (improved peak shape and peak intensities)
[b]Void time (t_0) was estimated using the elution time (min) of naphthalene
[c]Void time (t_0) was estimated using the elution time (min) of uracil

Table 2
Electrospray ionization and mass spectrometry data acquisition parameters per the LC method

	HILIC ESI (+)	HILIC ESI (−)	RPLC ESI (+)	RPLC ESI (−)
Nozzle voltage (V)	0	0	500	500
Capillary voltage (V)	2000	2000	3500	3750
Fragmentor voltage (V)	150	100	200	200
Drying gas	Nitrogen	Nitrogen	Nitrogen	Nitrogen
Sheath gas	Nitrogen	Nitrogen	Nitrogen	Nitrogen
Drying gas temperature (°C)	250	250	325	350
Sheath gas temperature (°C)	350	350	325	350
Drying gas flow (L/min)	8	10	8	8
Sheath gas flow (L/min)	11	10	8	8
Nebulizer gas pressure (psig)	45	45	30	30
MS1 range (*m/z*)	60–1200	60–1200	100–1500	100–1500
MS1 acquisition mode	Profile	Profile	Profile	Profile
MS1 scan rate (spectra/s)	2	2	4	4
MS2 mass range (*m/z*)	40–1000	40–1000	60–1200	60–1200
MS2 acquisition mode	Auto MSMS, profile	Auto MSMS, profile	Iterative exclusion MSMS, profile	Iterative exclusion MSMS, profile
MS2 scan rate (spectra/s)	6	6	6	6
Maximum number of precursors per scan cycle	4	4	4	4
Collision energy (eV)[a]	10–20–40	10–20–40	10–20–40	10–20–40
Quadrupole isolation width	Small (1.3 amu)	Small (1.3 amu)	Small (1.3 amu)	Small (1.3 amu)

[a]For HILIC ESI (+) and HILIC ESI (−) methods, additional MS2 runs are acquired using one collision energy at a time (10, 20, or 40 eV) with a maximum of 12 precursors per scan cycle

(b) Inject the second (set of) blank extraction sample(s) at the end of the worklist, followed by the system suitability sample (*see* **Note 5**). Acquire data in MS1 mode (Table 1).

10. Monitor column pressure, retention times, signal intensities, and mass accuracy of the internal standards during the acquisition.

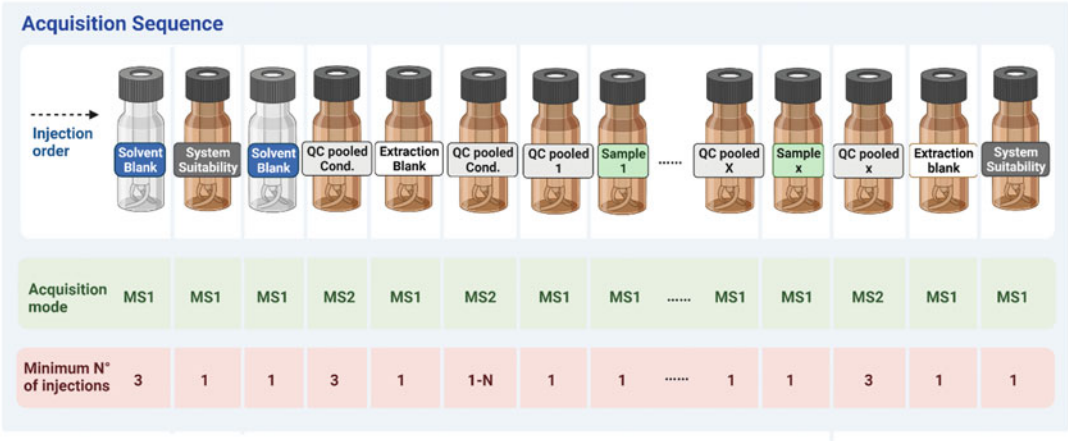

Fig. 2 General overview of an acquisition sequence for untargeted metabolomics analysis. The minimum number of injections is a suggestion based on the analytical platforms and sample matrix used in this work; these values should be experimentally re-evaluated for different systems. A system suitability (SS) sample is injected at the start and end of the analysis (MS1 data acquisition mode). Based on set acceptation criteria, the SS sample at the start of the analysis will be used to determine whether the instrumental performance is sufficient to start the worklist. When acceptance criteria are met for the SS sample, pooled QC samples are analyzed at the start to condition the system. If the test samples are not run in MS2 acquisition mode, the pooled QC conditioning samples at the beginning of the run and/or an additional set of pooled QC samples at the end of the run can be applied for MS2 data acquisition. In addition, pooled QC samples are run periodically, every 5–10 test samples, for QC processes (MS1 data acquisition mode)

4 Notes

1. Protect the antioxidant and chelator solutions from light and air. Prepare the solutions not earlier than 24 h before sample preparation.
2. When working with $CHCl_3$, use glass recipients and glass syringes instead of plastic recipients to avoid plasticizer leakage, leading to contamination during LC–MS acquisition. Perform all manipulations inside a fume hood.
3. Isotopically labeled internal standards are spiked in all analyzed samples to assess system stability for each injection (evaluation based on RT, mass accuracy, and peak intensity/area). In addition, other isotopically labeled internal standards may be used. However, in each method (HILIC and RPLC, ESI (+) and ESI (−)), at least three internal standards should be detected and spread throughout the retention time dimension, ideally with large differences in *m/z* value.
4. Store the HILIC reconstitution solvent at a temperature not lower than 4 °C to avoid phase separation.

Table 3
System suitability mixture of standards (1 μg/mL) for RPLC and HILIC methods in ESI (+) and ESI (−) modes

Compound name	Formula	Ionization species	*m/z*	Retention time (min)				Acceptance criteria		
				RPLC ESI (+)	RPLC ESI (−)	HILIC ESI (+)	HILIC ESI (−)	Peak height (counts)	Mass error (ppm)	RT deviation (min)
Lithocholic acid	$C_{24}H_{40}O_3$	[M−H]−	375.2904		6.95			> 5000 < saturation	< 10	< 0.5
1,2-Dipalmitoyl-rac-glycero-3-phosphoinositol (PI 16:0/16:0)	$C_{41}H_{79}O_{13}P$	[M−H]−	809.5185		13.69			> 5000 < saturation	< 10	< 0.5
1,2-Dipalmitoyl-sn-glycero-3-phosphoethanolamine (PE 16:0/16:0)	$C_{37}H_{74}NO_8P$	[M−H]−	690.5079		17.49			> 5000 < saturation	< 10	< 0.5
2-Octenoyl-L-carnitine (CAR 8:1)	$C_{15}H_{27}NO_4$	[M+H]+	286.2013	3.12		10.44		> 5000 < saturation	< 10	< 0.5 HILIC < 0.2 RPLC
1-Stearoyl-phosphatidylethenolamine (LPE 18:1)	$C_{23}H_{46}NO_7P$	[M+H]+	480.3085	9.40				> 5000 < saturation	< 10	< 0.5 HILIC < 0.2 RPLC
Glycerol tripalmitate-[$^{13}C_3$] (TG 16:0/16:0/16:0-[13C3])	$C_{48}[^{13}C]_3H_{98}O_6$	[M+NH_4]+	827.7802	21.87				> 5000 < saturation	< 10	< 0.5
Thymine	$C_5H_6N_2O_2$	[M−H]−	125.0355				2.59	> 5000 < saturation	< 10	< 0.2
L-tryptophan	$C_{11}H_{12}N_2O_2$	[M−H]−	203.0826				6.09	> 5000 < saturation	< 10	< 0.2
L-lysine	$C_6H_{14}N_2O_2$	[M−H]−	145.0990				14.44	> 5000 < saturation	< 10	< 0.2
Nicotinic acid	$C_6H_5NO_2$	[M+H]+	124.0393			5.71		> 5000 < saturation	< 10	< 0.2
L-arginine	$C_6H_{14}N_4O_2$	[M+H]+	175.1190			17.50		> 5000 < saturation	< 10	< 0.2

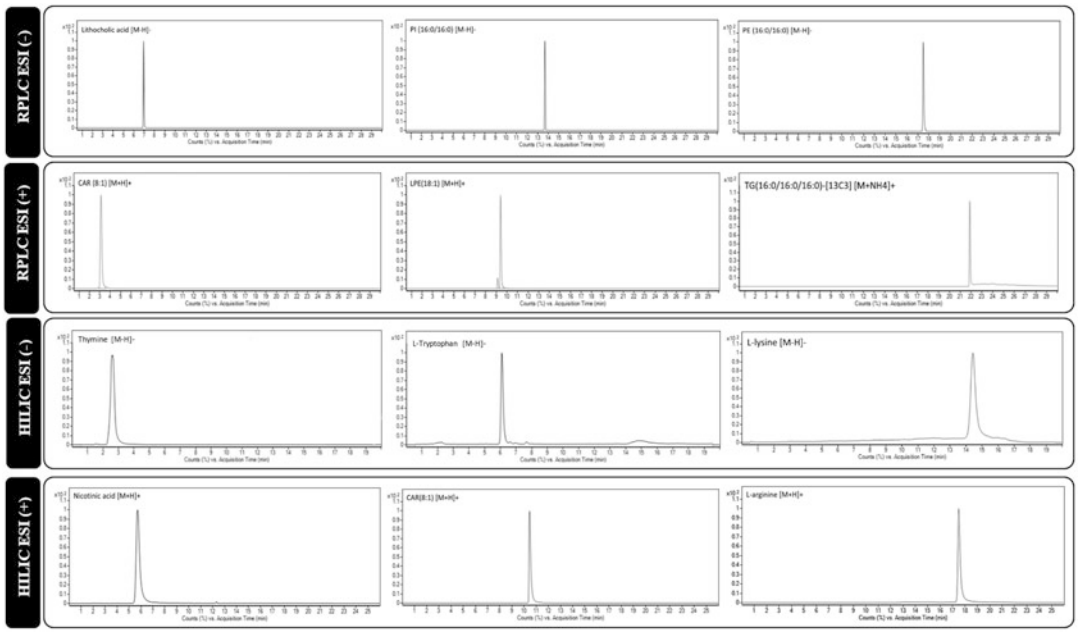

Fig. 3 Chromatograms of the compounds used for system suitability in the different platforms

5. The system suitability sample is applied to verify that the system will perform according to the pre-defined acceptance criteria (e.g., concerning peak height, area, shape, retention time, and mass accuracy). Other analytical standards can be used if distributed across the *m/z* and retention time range (i.e., to assess the complete analysis window) in positive and negative ionization modes. Acceptance criteria depend on the instrumental performance and can be set based on the variation of repeated injections over time (e.g., every week, three injections per day for 1 month). Acceptance criteria used for LC-HRMS methods described in this chapter were (1) peak height above 5000 counts and without saturation, (2) maximal mass error of 10 ppm, and (3) RT deviation of maximum ±0.5 min for HILIC methods and ±0.2 min for RPLC methods. The system suitability sample is injected before each sample batch analysis, and when the acceptance criteria are fulfilled, data acquisition can be initiated. In addition, the system suitability sample is injected at the end of the worklist and results are again compared to the acceptance criteria. This latter is important to estimate the maintenance of the instrumental performance during the run. Prepare a sufficient volume of the system suitability sample to enable long-term usage.
6. Filter the aqueous fraction of the mobile phases (0.2-μm cellulose acetate membrane filters) after dissolving the modifier(s) to remove particulates in order to keep the system and column

from clogging. After the preparation, sonicate the mobile phases for 30 min to degas them. During sonication, put the cap on the mobile phase bottle, but leave it unscrewed. The aqueous mobile phases have a maximum shelf life of 1 week. The aqueous/organic mobile phases have a shelf life of up to 3 weeks and the organic mobile phases up to 3 months.

7. When preparing solutions for incubation or exposure of HepaRG cells, always work in a LAF cabinet of class II or higher. After turning the LAF cabinet on, wait at least 15 min for stabilization and disinfect all surfaces. Only use disinfected materials inside the environment of the LAF cabinet, and make sure to disinfect your hands each time before entering the laminar flow. Use only sterile equipment when contact with the culture media is anticipated.
8. Do not vortex collagen solutions, but gently swirl instead. The collagen solution should only be prepared just before the coating of the chamber slides.
9. HepaRG Thaw, Seed, and General Purpose Medium and HepaRG Maintenance and Metabolism Medium should be stored at 4 °C for a maximum of 1 month.
10. In order to use the sample preparation method for other types of cells, the number of cells and/or the dilution factor need to be optimized since the cell volume, nucleus/cytoplasm ratio, and metabolic content vary between different types of cells. For this latter optimization, the recommendations of Wu et al. can be used [12].
11. When using chamber slides for cultivation, only use one well per chamber slide to scrape the cells during extraction efficiently. When using multiple wells of the same chamber slide, the risk of cross-contamination increases considerably.
12. Before thawing the HepaRG cryovial, twist the cap a quarter turn in the LAF cabinet and close the cap again. This twist will ensure the release of internal pressure after storage of the cryovial in liquid N_2.
13. Cell suspensions need to be pipetted up and down several times to minimize the clustering of cells. Cell suspensions should not be vortexed.
14. Differentiated HepaRG cells can be purchased in cryovials containing 8, 10, or 12×10^6 cells.
15. It is advisable to check the homogeneity of the cell distribution in the well using a light microscope.
16. A minimum of two extraction blanks should be acquired by performing incubation without cells. These extraction blanks should be handled the same as other samples. Signals in the blank extraction samples are dealt with appropriately during blank subtraction procedures.

17. Required concentrations depend on the design of the experiment. Relevant concentrations may include physiological relevant concentrations, IC5 or IC10 values. Higher concentrations should be avoided for metabolomics experiments, as high mortality rates result in material bias. A stock solution of the investigated compound(s) may be prepared and diluted to the required concentrations.
18. Prevent incubation bias through randomization of medium exchange order and incubation position.
19. When working with liquid nitrogen, wear a protective apron, protective gloves, a full-face shield, or chemical splash goggles, and ensure the area is well-ventilated. Handle liquid nitrogen slowly to minimize boiling and splashing. Only use approved containers and do not fill them over 80% of their capacity.
20. The processing order of chamber slides should be randomized to prevent bias related to processing order.
21. Each fraction of the liquid–liquid extraction (i.e., polar and apolar) is divided into two aliquots right before the evaporation step. Since each fraction is analyzed in both ESI (+) and ESI (−), dividing the polar and apolar fractions into two sub-fractions avoids freeze–thaw cycles during analyses.
22. Rinse the syringe three times using $CHCl_3$ after each transfer.
23. Pooled QC samples are applied to condition the analytical system and perform precision measurements. The precision of the dataset can be estimated by calculating the relative standard deviation (RSD) of the intensity of the features for each ionization mode after blank subtraction.
24. For the HILIC ESI (−) method, the heat exchanger and the inline filter located after the injector are bypassed to avoid the chelating interaction of anionic metabolites, such as carboxylic acids and phosphorylated anions with trace metals from the concerned hardware, which would result in an impaired peak shape and reduced peak intensities [5].
25. Reference masses used for continuous mass axis calibration can vary among instrument providers
26. For equilibration of the LC column, usage of mobile phases without modifiers is preferred. Depending on the solvent used to store the LC columns, buffer salts can precipitate and cause backpressure build-up inside the column.
27. The number of pooled QC injections necessary for conditioning the system is dependent on the samples under investigation. The number can be optimized by iterative injections of the QC pooled sample. When the total ion current chromatogram, the RT, and peak intensity/area of the internal standards become stable, acquisition of samples can be initiated.

References

1. Wishart DS, Tzur D, Knox C et al (2007) HMDB: the human metabolome database. Nucleic Acids Res 35:D521–D526. https://doi.org/10.1093/nar/gkl923
2. Alseekh S, Aharoni A, Brotman Y et al (2021) Mass spectrometry-based metabolomics: a guide for annotation, quantification and best reporting practices. Nat Methods 18:747–756. https://doi.org/10.1038/s41592-021-01197-1
3. Gertsman I, Barshop BA (2018) Promises and pitfalls of untargeted metabolomics. J Inherit Metab Dis 41:355–366
4. Patti GJ (2011) Separation strategies for untargeted metabolomics. J Sep Sci 34:3460–3469. https://doi.org/10.1002/jssc.201100532
5. Iturrospe E, Da Silva KM, Talavera Andújar B et al (2021) An exploratory approach for an oriented development of an untargeted hydrophilic interaction liquid chromatography-mass spectrometry platform for polar metabolites in biological matrices. J Chromatogr A 1637: 461807. https://doi.org/10.1016/j.chroma.2020.461807
6. da Silva KM, Iturrospe E, Heyrman J et al (2021) Optimization of a liquid chromatography-ion mobility-high resolution mass spectrometry platform for untargeted lipidomic and application to HepaRG cell extracts. Talanta 122808. https://doi.org/10.1016/j.talanta.2021.122808
7. da Silva KM, Iturrospe E, Bars C et al (2021) Mass spectrometry-based zebrafish toxicometabolomics: a review of analytical and data quality challenges. Meta 11:635. https://doi.org/10.3390/metabo11090635
8. Ghaste M, Mistrik R, Shulaev V (2016) Applications of fourier transform ion cyclotron resonance (FT-ICR) and orbitrap based high resolution mass spectrometry in metabolomics and Lipidomics. Int J Mol Sci 17. https://doi.org/10.3390/IJMS17060816
9. Xu T, Hu C, Xuan Q, Xu G (2020) Recent advances in analytical strategies for mass spectrometry-based lipidomics. Anal Chim Acta 1137:156–169. https://doi.org/10.1016/j.aca.2020.09.060
10. Cuykx M, Negreira N, Beirnaert C et al (2017) Tailored liquid chromatography-mass spectrometry analysis improves the coverage of the intracellular metabolome of HepaRG cells. J Chromatogr A 1487:168–178. https://doi.org/10.1016/j.chroma.2017.01.050
11. Broadhurst D, Goodacre R, Reinke SN et al (2018) Guidelines and considerations for the use of system suitability and quality control samples in mass spectrometry assays applied in untargeted clinical metabolomic studies. Metabolomics 14:1–17. https://doi.org/10.1007/s11306-018-1367-3
12. Wu ZE, Kruger MC, Cooper GJS et al (2019) Tissue-specific sample dilution: an important parameter to optimise prior to untargeted LC-MS metabolomics. Meta 9:1–19. https://doi.org/10.3390/metabo9070124

Chapter 20

Data Processing and Analysis in Mass Spectrometry-Based Metabolomics

Ángela Peralbo-Molina, Pol Solà-Santos, Alexandre Perera-Lluna, and Eduardo Chicano-Gálvez

Abstract

Metabolomics is the latest of the omics sciences. It attempts to measure and characterize metabolites—small chemical compounds <1500 Da—on cells, tissue, or biofluids, which are usually products of biological reactions. As metabolic reactions are closer to the phenotype, metabolomics has emerged as an attractive science for various areas of research, including personalized medicine. However, due to the complexity of data obtained and the absence of curated databases for metabolite identification, data processing is the major bottleneck in this area since most technicians lack the required bioinformatics expertise to process datasets in a reliable and fast manner. The aim of this chapter is to describe the available tools for data processing that makes an inexperienced researcher capable of obtaining reliable results without having to undergo through huge parametrization steps.

Key words Mass spectrometry, Untargeted metabolomics, Data processing, Liquid chromatography

Abbreviations

ANOVA	Analysis of Variance
BMI	Body Mass Index
CAWG	Chemical Analysis Working Group
CC	Component Correction
CE	Capillary Electrophoresis
ChEBI	Chemical Entities of Biological Interest
CITAC	Cooperation on International Traceability in Analytical Chemistry
CPCA	Common Principal Component Analysis
CROP	Correlation-Based Removal of Multiplicities
CUMSUM	Cumulative Sum
CWT	Continuous Wavelet Transform
DI	Direct Infusion
EDA	Exploratory Data Analysis
EIC	Extracted Ion Chromatography
ESI	Electrospray Ionization

Raúl González-Domínguez (ed.), *Mass Spectrometry for Metabolomics*,
Methods in Molecular Biology, vol. 2571, https://doi.org/10.1007/978-1-0716-2699-3_20,
© The Author(s), under exclusive license to Springer Science+Business Media, LLC, part of Springer Nature 2023

EWMA	Exponential Weighted Moving Average
FAIMS	High-Field Asymmetric Waveform Ion Mobility Spectrometry
FI	Flow Injection
GB	Gigabytes
GC	Gas Chromatography
GNPS	Global Natural Products Social Molecular Networking Database
GNU GPL	General Public License
GSEA	Gene Set Enrichment Analysis
HCA	Hierarchical Clustering Analysis
HMDB	Human Metabolome Database
HPLC	High-Performance Liquid Chromatography
hRUV	Remove Unwanted Variation in a Hierarchical Structure
IMS	Ion Mobility Separation
IP4M	Integrated Mass Spectrometry-Based Untargeted Data Mining
IPO	Isotopologue Parameter Optimization
IS	Internal Standard
KEGG	Kyoto Encyclopedia of Genes and Genomes
KNN	K-Nearest Neighbor Imputation
LC	Liquid Chromatography
LOESS	Locally Estimated Scatterplot Smoothing
LOESS	Locally Weighted Smoothing
LOWESS	Locally Weighted Scatterplot Smoothing
LPA	Label Propagating Algorithms
m/z	Mass-to-Charge Ratio
MAIT	Metabolite Automatic Identification Toolkit
MAR	Missing at Random
MAVEN	Metabolomic Analysis and Visualization Engine
MCAR	Missing completely at random
MLPA	Multi-Label Propagating Algorithms
MNAR	Missing Not at Random
MoNA	Mass Bank of North America
MP	Mobile Phase
MS	Mass Spectrometry
MS-DIAL	Mass Spectrometry–Data Independent Analysis Software
MSEA	Metabolite Set Enrichment Analysis
MSI	Metabolomics Standard Initiative
MVI	Missing Value Imputation
ncGTW	Neighbor-wise Compound-Specific Graphical Time Warping
NMR	Nuclear Magnetic Resonance
ORA	Over-representation Analysis
OUKS	Omics Untargeted Key Script
PCA	Principal Component Analysis
PLS-DA	Partial Least Squares–Discriminant Analysis
QA	Quality Assurance
QC	Quality Control
QEA	Quantitative Enrichment Analysis
RF	Radio Frequency
RIE	Reduced Isotopic Envelope
RLE	Relative Log Expression

ROI	Region of Interest
RT	Retention Time
SNR	Signal to Noise Ratio
SOP	Standard Operating Protocol
SPC	Statistical Process Control
SSP	Single Sample Profiling
TIC	Total Ion Chromatography
UPLC	Ultrahigh-Performance Liquid Chromatography

1 Introduction

Omics sciences adopt a holistic view of the molecules that make up a cell, tissue, or organism [1] and are primarily aimed at the quantification of the genome, proteome, and metabolome. The genome, a term coined by Hans Winkler merging "gene" and "chromosome" [2, 3], is defined as the set of genes that comprise the heritable material [4]. The proteome was coined in a similar way in 1994 by Mark Wilkins [5] and stands for the entire complement of proteins expressed by the genome in a concrete spatial-temporal time. Finally, the metabolome comprises the whole set of chemical molecules produced by cells, tissue, or biological fluids [6].

Genomics, proteomics, and metabolomics are the three sciences that deal correspondingly with the genome, proteome, and metabolome.

Metabolomics represents an upstream readout of genetics and proteomics variations [7], being the closest to the real phenotype. On the one hand, this closeness opens a wide range of opportunities for disease diagnostics, drug discovery, or treatment follow-up, among others [1, 8], but on the other hand, it introduces a large variance coming from the environment. In addition, the whole metabolite population remains still unknown [9], complicating the whole statistical handling and interpretation. Metabolomics can therefore be seen as bridging the gap between genotype and phenotype [10], providing a more comprehensive view of how cells function, as well as identifying novel or striking changes in specific metabolites. Analysis and data mining of metabolomics data sets and their metadata can also lead to new hypotheses and new targets for biotechnology and biomedical research.

When analyzing a biological system, one can either be interested in a global profiling of the metabolome or in a defined set of metabolites. Those strategies are known as untargeted and targeted designs. Besides, a number of complementary approaches have to be established for extraction, detection, quantification, and identification of as many metabolites as possible [11, 12].

The number of metabolites to measure will mostly drive the election of the experimental design. Due to the huge diversity of

chemical structures and the large differences in abundance, there is no single technology available to analyze the entire metabolome. Mass spectrometry (MS) coupled to different chromatographic separation techniques, such as liquid (LC) or gas chromatography (GC) or nuclear magnetic resonance (NMR), are the main settings to analyze a large number of metabolites simultaneously.

LC–MS is applied most frequently in untargeted metabolomics studies because it offers a high number of advantages. The LC allows the separation of metabolites at different times before their analysis with the mass spectrometer. The combination of LC separation together with analysis increases the number of metabolites detected. The spatial separation of metabolites allows the detection of different isomers (metabolites that have the same mass) and helps detect metabolites at low concentrations. Besides, the year 2020 has seen an enormous rise in applications of ion mobility-mass spectrometry (IMS) methods of analysis in metabolomics, improving the separation and detection of different metabolites.

The raw data acquired from liquid chromatography–mass spectrometry instruments in an untargeted metabolomics study is composed of three-dimensional matrices, including mass-to-charge ratio [*m/z*], retention time [RT], and height or area, a measure of the analyte abundance. New instruments include IMS as a fourth dimension.

Resulting matrices are highly complex. First, the sensibility of the instrument induces the detection of hundreds of thousands of molecular features, including noise originating from the nonbiological background, worsened by the fact that each analyte can appear as multiple features. This leads to high dimensional matrices with plenty of false-positive detections and large collinearities. Second, the results of the experiment are very sensible to the instrument calibration, parameterization, and environmental settings. This impacts poor reproducibility and a general lack of standards to use as a reference for normalization and annotation. Third, hundreds or thousands of samples can be analyzed in a single study. Small alterations in the analytical settings might propagate, inducing nonlinearities as time shifts on the retention time, drifts, depletions of signal, and batch effects. This poses a statistical challenge in order to ensure comparability across samples and features. In addition, file sizes ranging from 1 to 6 gigabytes (GB) in function of the experimental settings might represent a computational challenge for most of the analysts. Finally, the incompleteness of the universe of known metabolites induces a challenge to match molecular features to known metabolites and to interpret them within a biological context. Despite the complexity of the analysis (i.e., nonlinear drifts, sample normalization, or feature identification), untargeted metabolomics provides an unbiased snapshot of the whole metabolome depicting a close insight into the biological system under study.

2 Experimental Devices for Metabolomics

To date, the two main analytical platforms to generate metabolomics data are NMR and MS. NMR allows one to analyze biological samples without undergoing detrimental sample processing at the cost of low sensitivity. On the contrary, MS is capable of isolating and detecting a huge amount of metabolites with great sensitivity but at the cost of a severe sample preprocessing [13]. In addition, MS detection can be divided into two: chromatographic separation and non-chromatographic separation [14]. Among others, the current trends for MS detection based on chromatographic separation are GC, LC, and capillary electrophoresis (CE). To increase selectivity and resolution, two-dimensional chromatography arises, commonly (GC × GC) and (LC × LC). On the other hand, direct infusion (DI) and flow injection (FI) platforms stand for non-chromatographic MS detection [15].

2.1 NMR

NMR uses magnetic fields and radio frequency (RF) to reveal information about the magnetic nuclei (16). On NMR, the sample is placed onto an electromagnetic field that aligns all biological magnetic elements in a similar way that Earth does with a compass.

During data acquisition, RF waves irradiate the sample and deflect this alignment on the selected nuclei, commonly 1H and 13C, in a process named saturation. After RF is turned off and nuclei relax back to the equilibrium or ground state. In the relaxation process, nuclei emit RF waves of the exact wavelength in which they were excited. Intensity, frequency or chemical shift, half-life, and phase are the four parameters that can be collected from RF irradiation [16].

The principal advantage of NMR is that the sample can be fully recovered after the analysis because it does not rely on sample preprocessing and analyte separation [17]. NMR has proved very useful for the structure characterization of unknown compounds [18]. Despite these advantages, NMR technique has scarce sensitivity when compared to MS methods [19]. Several approaches have been designed to overcome this low sensitivity problem [16].

2.2 CE–MS

Capillary electrophoresis (CE) combines capillary electrophoretic concepts with the early possibilities of ionizing molecules (electrospray ionization, ESI) [20]. The instrumental arrangement of CE–MS covers a capillary equipped with its injection port in one end and in the other the ESI interface [20]. Both ends of the capillary are connected to a high-voltage power supply. A sheath liquid system stabilizes the separation enhancing the analyte ionization. Ionized molecules arrive to the MS analyzer and go through a magnetic field and finally hit the detector.

2.3 GC–MS

GC–MS couples gas chromatography (GC) and a mass spectrometer (MS). In GC, as with most chromatography techniques, a mobile phase (MP) and a stationary phase are needed. The first is also called carrier gas and is usually an inert gas (e.g., helium, argon, nitrogen) that moves samples across the column. The stationary phase is a packed column. When a sample is forced to go through the column, its molecules interact differently with the stationary phase as a function of its affinity. As a consequence, molecules that were initially mixed appear separate at the elution of the column. Retention time is defined as the time spent for a molecule to go through the column [21]. MS configuration is coupled after the column. One of the main advantages is its high sensitivity and reproducibility [21], whereas a big drawback is that it can only profile apolar molecules that range the volatile to semi-volatile masses (i.e., <700 Da) [21].

2.4 LC–MS

In a similar way to GC–MS configuration, LC–MS couples liquid chromatography (LC) with a mass spectrometer (MS). HPLC and UHPLC stand for high-performance LC and ultrahigh-performance LC, respectively, and are the current high-resolution configurations of LC. Both are commonly used for metabolomics analysis.

LC uses a liquid as the carrier fluid or solvent. The solvent is forced to pass through a chromatographic column. This column is packed with pellicular or porous particles. Metabolites have different affinities to this adsorbent material generating different flow rates and, in consequence, different eluting times. High-resolution LC withstands pressures up to 400 atmospheres (for HPLC) or 1000 atmospheres (for UHPLC). This pressure by the system permits the packed column material to be composed of smaller particle sizes and thus have a great surface to area ratio leading to a better separation of the metabolites within the mixture.

Thus, LC allows the separation of the metabolites by differences in their retention times in the column. Coupled to the chromatographic column, there is a mass spectrometer for second separation in the mass to charge space.

2.5 Ion Mobility

Besides using GC, LC, or CE to separate metabolites and MS, MS/MS, and even MSn to distinguish them from one another, ion mobility is an additional separation dimension that is being introduced in several ways. FAIMS (high-field asymmetric waveform ion mobility spectrometry) and differential mobility operate in the ionization interface. Ions move through a small gap and are subjected to an oscillating, orthogonal and asymmetric, electric field. By applying a particular compensating voltage, individual ions can be coaxed to pass into the mass spectrometer. Further metabolite separation can be achieved by including low levels of metal ions in the mobile phase or by resolving the ions with isopropanol or other volatile solvents.

The other classes of ion mobility occur inside the mass spectrometer and allow for the calculation of the collision cross-section of a metabolite ion. This is an absolute value of a metabolite ion, independent of the instrument used for its measurement or lab where it is done.

3 Metabolomics Workflow

The final aim of metabolomics is to comprehend a biological system from the measure of its metabolites. A typical workflow covers from experimental design up to the functional interpretation of the results. As the LC–MS approach is the most widely used in the metabolomics field, this chapter is mainly focused on this analytical combination. The general workflow can be resumed in the following steps:

- Given a biological system to study, the first step is the *experimental design*. This step involves deciding whether the approach will be targeted or untargeted, sample collection, extraction protocols, and the analytical platform for obtaining data.
- *Data acquisition* implies the sample analysis according to the standard operating protocol (SOP) defined within the experimental design.
- *Data processing* may include peak picking, quality assurance, normalization, missing value imputation, transformation, and scaling. These steps are performed either through private or public tools.
- *Data analysis* consists of applying statistical approaches with the aim to obtain the best features in terms of separability of classes or explained variance from the working hypotheses. The challenge is then to identify those features that are related to the phenomenon studied and isolate them from the ones whose variation is not related to the phenomenon (i.e., drug response, phenotype, or diet, among others).
- In the case of untargeted, the next step is to undergo through *annotation* (i.e., match and label features with publicly available metabolites).
- Finally, *biological* insight is given through functional *interpretation* (i.e., pathway enrichment).

3.1 The Influence of Experimental Design and Data Acquisition on the Processing and Analysis of Metabolomics Data

A well-planned experimental design is crucial to ensure the successful outcome of any scientific study [22]. Experimental design includes all steps in the metabolomics pipeline from sample collection, preparation, and data acquisition to data processing and statistical analysis [23]. The role of the data analysis is not confined to the analysis of the data, but it should also take an active role in preventing the introduction of bias, selecting the appropriate

number of samples, developing an awareness of the influence of experimental design and data acquisition on the analysis of the data, and selecting the adequate data analysis techniques to get correct outcomes of the study.

3.1.1 Standard Operating Procedures (SOPs)

SOPs are routinely applied in scientific studies to promote good practices and transparency. In population-based studies, samples may be collected and processed at different sites by different people. To minimize the introduction of variation across sites, procedures should be fully defined in standard operating procedures (SOPs) and carried out by fully trained staff. SOPs should include constraints in the maximum time between sample collection and preparation, and sample preparation and data acquisition to ensure that sample instability is not a concern during large studies. Deviations from the outlined procedure should be recorded and considered in the final analysis and interpretation of the biological study.

3.1.2 Confounding Factors

In the collection of metadata, technical and clinical information about the samples is essential to identify confounding factors. Confounding factors are variables that are associated with both the features and the biological question under study. Ideally, confounding factors should be identified during the experimental design and included within the data acquisition procedure. It is key to determine if the biological findings may be misinterpreted as a result of a confounding factor. Typical confounders include demographic data, lifestyle, and physiological conditions such as diet, gender, ethnicity, age, and body mass index (BMI). It is preferential to modify the experimental design to eliminate confounding factors rather than adjust the data post-analysis.

3.1.3 Quality Control in the Analytical Stage

In untargeted metabolomics, the overall quantitative variation can be typically categorized into analytical and biological variations [24]. Analytical variation usually comes from the experimental procedures, such as sample handling and preparation, metabolite extraction, and LC–MS analysis. In comparison, biological variation is due to the intrinsic biological variances among samples.

Sources of variance that may be affecting the analytical stage include the following [25]:

- Equipment calibration. All the involved equipment, including pipettes, balances, freezers, etc., must be routinely calibrated.
- Batch effect. Experiments usually involve tens of samples that cannot be processed in a unique batch due to capacity restrictions (i.e., Rack Capacity). Changes from one batch to another introduce a huge source of unwanted variances.
- Selection of the column. Each column optimizes the separability in the time domain of a certain family of compounds. Poor selection of the column may lead to indistinguishable peaks.

- MP preparation. As well as column, MP also optimizes the separability in the time domain of a certain family of compounds. The total amount of MP must be pre-established to avoid refilling, one of the principal sources of unwanted variance.
- Carryover. This occurs when metabolites of a certain sample are still measurable after this sample is actually analyzed. There are basically two points where carryover can arise, on the autosampler (e.g., pipetting) or on the column (e.g., column aging).
- Matrix effects. Co-eluting molecules might alter the intensity of the compound of interest. This kind of effect might appear, among others, in the presence of contaminants, dirt on the source, or inadequate flow rate on the mobile phase.
- Temperature fluctuations. Temperature of the equipment and of the laboratory may remain constant during the whole experiment. Fluctuations may lead to equipment misbehavior, and it can alter the mobile phase and all its dependent features. Temperature should be monitored during all the experiment acquisition.

Several techniques have been designed by the analytical chemists to reduce as much as possible those sources of variance in their primary forms. Introducing diluted samples aids in the detection of the matrix effect (i.e., Ion suppression is dependent on the sample concentration) [26]. In addition, diluted samples establish an intensity trend that must be followed by all the reliable peaks [26]. With similar means, spiked samples can be introduced to the study. Spiked samples are those samples that contain compound standards with known behavior and concentration [27].

Another strategy is to use replicates. Among others, the use of technical replicates (i.e., repeated measurements of the same sample) is very useful in terms of improving reproducibility [22, 28].

A typical way to control the overall variance and behavior of the analysis while avoiding the inter-sample variance is to introduce quality controls (QCs). QCs are usually composed of a pool of all the samples present in the study. As QC samples do not suffer from inter-sample variance, they can be analyzed several times on a run allowing to properly characterize the overall equipment behavior [29]. This will be deeply discussed in Subheading 3.2.3.

3.1.4 Randomization

Randomization is an essential requirement of any scientific experiment to prevent the introduction of bias into a study. Inadequate randomization can negatively influence the outcomes of scientific research; study groups should be balanced with respect to known and unknown confounding factors [23]. In large-scale studies ($n > 100$), it is impossible to prepare and analyze all the samples in a single batch. Samples must be split into batches for the

preparation of samples and data acquisition. In those cases, randomization is performed to ensure that there is independence between any technical or biological covariate and the analytical routine plan (i.e., batch and injection order assignment). The lack of independence might introduce bias into the study data. Covariate distribution within the batches should be representative of the overall study.

3.2 Data Processing

Processing of raw LC–MS data typically is performed by the following steps: peak picking, peak alignment, peak matrix composition, data quality control, data preparation, data analysis, and, finally, functional analysis [30].

3.2.1 Peak Picking

The conversion of the raw data acquired by the experimental device into a list of peaks is a process commonly described as "peak picking" [31]. There is a wide range of software, both from the equipment vendors, e.g., MarkerView™ [32] by ABSciex Ltd., Progenesis QI by Waters Ltd. [33], MassHunter Profinder by Agilent Ltd. [34], Compound Discoverer™ by ThermoFisher Inc. [35], or MetaboScape by Bruker Inc. [36], and open source (e.g., mzMine2 [37], MS-DIAL [38], or XCMS [39]) software capable of addressing this task [31].

Peak picking is usually assessed following a two-step workflow. The first one applies several filters to smooth (e.g., Savitzky-Golay, Moving Average) and correct baseline nonlinearities (e.g., linear interpolation) [31]. The second step identifies metabolite peaks by detecting several parameter thresholds (e.g., signal to noise ratio, intensity, peak width).

GridMass [40] and centWave [41] are two of the most used algorithms for peak picking in metabolomics processing:

- GridMass [40] is a two-dimensional recursive peak detection algorithm. This method is implemented in Java [42] as a plugin in the mzMine2 package [37]. First, it ignores those regions of the matrix that are candidates to be artifacts (i.e., non-column-dependent metabolites that elute at the first seconds of the experiment) using a user-predefined threshold. Second, it tessellates the matrices with equally spaced probes. After, each probe explores its surrounding regions (delimited by the other probes) in search for its local maximum. This process is iterated until convergence is achieved. Peaks are integrated by taking into account the initial and final position of the probes. False positives are filtered by user-defined thresholds (e.g., minimum intensity, m/z tolerance, minimum and maximum retention time width).
- The CentWave algorithm is the current peak picking algorithm implemented in XCMS [39]. Previously Matched Filter [39] that divides the data into slices of user-defined m/z width and

determines the signal in each slice by taking the maximum intensity at each RT [43] was used. CentWave is based on the detection of regions of interest (ROIs) in the *m/z* domain combined with a Continuous Wavelet Transform (CWT) for its integration [44]. ROIs are determined as regions of delimited extend (i.e., peak width) with low *m/z* variance. Tolerance of the variance of the *m/z* and the extent of the region are user inputs. After, a CWT is fitted onto the intensity values of the ROI, the scale range of the CWT is automatically determined. After, false-positive peaks are reduced by making use of several user-defined thresholds such as minimum peak intensity and signal to noise ratio (SNR). Those parameters can be optimized using the Isotopologue Parameter Optimization (IPO) software.

More tools are available for the peak picking processing such as TidyMS [45], Skyline [46], apLCMS [47], deconMSn [48], or integrated mass spectrometry-based untargeted metabolomics data mining (IP4M) [49].

3.2.2 Peak Alignment

LC–MS experiments usually involve injecting tens of biological samples. The continuous injection inevitably ages the LC column (e.g., sample carryover, solid or stationary phase changes) [50]. Aging of the column and differences on temperature, pressure, or buffer composition, produce nonlinear shifts on the elution time of the metabolites [25]). Although several protocols to mitigate unwanted variances are implemented in the analytical stage (e.g., blanks and QC injections), they are not able to completely resolve the issue.

Shifts during analysis are critical when trying to compare peaks of several samples [50]. The computational procedure that attempts to correct the drift in the retention time across samples is called peak alignment. In short, peak alignment consists of constructing functions that align the peaks of any given sample to peaks of a given reference sample [50].

Peak alignment algorithms can be classified into two main groups:

- Warping algorithms: Warping approaches try to monotonically shift, stretch, and squeeze analytes eluting times until an objective function is minimized [50]. These algorithms try to minimize the RT distance among samples without using any landmark by equally weighting a distance function [51]. In order to boost the performance, algorithms define a reference sample [52] and prioritize property sharing ion features (e.g., correlation [53, 54], covariance [55], clustering [39]) to perform the correction. After this first approach, several functions can be fitted to correct the drift (i.e., piecewise interpolation [53, 56], polynomial functions [57], locally weighted

smoothing regression [58]). Combinations of the strategies for selecting the reference, sharing properties, and building the functions to fit samples are, in its majority, the actual benchmark on the warping alignment algorithms.

– Direct match algorithms: Direct match algorithms attempt to align chromatograms through samples based only on similarity metrics without any warping function [50]. These approaches are difficult to implement because of the data complexity and, in consequence, are less prevalent in the community. It starts by simplifying the problem by considering only a certain number of dimensions (i.e., total ion chromatogram (TIC), extracted ion chromatogram (EIC), reduced isotopic envelope (RIE) [50]). For example, Duran et al. [59] align RIEs that occur within a retention time window while Johnson et al. [60] match ion features to the reference sample by its similarity.

3.2.3 Quality Assurance and Quality Control in Metabolomics Analysis

According to ISO9000 [61], CITAC (the Cooperation on International Traceability in Analytical Chemistry), and EuraChem (A Focus for Analytical Chemistry in Europe) [62], quality assurance (QA) addresses the activities the laboratory undertakes to provide confidence that quality requirements will be fulfilled, whereas QC describes the individual measures that are used to actually fulfill the requirements.

QA can also be defined as all the planned and systematic activities implemented before samples are collected to provide confidence that a subsequent analytical process will fulfill predetermined requirements for quality [25]. Correspondingly, QC can be defined as the operational techniques and activities used to measure and report these quality requirements during and after data acquisition [29].

In untargeted metabolomics studies, blank and quality control (QC) samples are normally analyzed after data has been acquired to assess and ensure the analytical process is performed appropriately, providing a quantitative measure of quality for each metabolite detected or for the data set as a whole [29].

The blank sample is the solvent used for the metabolomics extraction after subjecting it to the complete extraction procedure. It is commonly used to reduce the dimensionality data set by filtering those metabolites with a blank contribution greater than the criteria fixed by the researcher (5% is commonly used) [29, 63].

The QC sample is a single sample that should be qualitatively and quantitatively representative of the entire collection of samples included in the study, providing an average of its metabolome. The QC sample that matches these criteria most accurately is a pooled sample since it represents both the sample matrix and metabolite composition of the samples. If it is not possible to create a pooled QC sample, an alternative QC sample should be used (e.g., a commercially available biological sample).

The QC sample should ideally provide a measure of the variance performed throughout the sample preparation, data acquisition, and data preprocessing steps. Sometimes, for example, when analyzing tissue samples, the pooled QC should be prepared after sample preparation protocol, pooling aliquots of each of the extracted samples. In these cases, the pooled QC will not provide a measure of the sample preparation variance but the variance after sample preparation.

As QC samples are assumed to be identical, preprocessing algorithms can be used to correct monitor drift, separate high- and low-quality data, equilibrate the analytical platform, and correct the signal drift allowing the integration of multiple analytical experiments/batches. These corrections allow the analytical variance reduction while maintaining the biological variance.

QC samples can also provide a quantitative measure of the technical variance or reproducibility. The relative standard deviation for each feature and the percentage of QC samples in which the feature was detected is applied as a filtering QC process. If the defined acceptance criteria are not met, then the metabolite feature is removed from the data set. The percentage of features that are usually removed during these procedures is 10–30% of all detected features.

Principal component analysis can be used to quickly assess the reproducibility of the QC samples in an analytical run. The principal component analysis (PCA) score plot illustrates the variability of the data within the study. The variance observed between the QC samples should be much lower than that observed between the biological samples. Introducing diluted QC samples can be used to reduce the matrix effect by filtering the features that do not follow the dilution factors [26].

The application of QCs in mass spectrometry-based untargeted studies has increased the capability to perform high-quality large-scale studies in terms of analytical variation.

In addition to the analytical and biological variation, in the case of poorly resolved peaks, computational variation is induced. This effect is worse in untargeted studies where noise and signal are not easily distinguishable. Huaxu Yu et al. have developed a QC-based workflow to automatically select either peak height or peak area based on their variations in repeated QC analyses to reduce computational variation without affecting analytical and biological variations [24].

3.2.4 Missing Value Imputation

Missing values can be problematic in the analysis of metabolomics data. There are three main reasons as to why a feature is missing from a mass spectrum:

- The metabolite is not present in the biological sample.
- The metabolite is present in the sample, but the concentration is below the limit of detection of the analytical technique used.

– The metabolite is present in the biological sample and at a level that would be detected by the instrument but was not integrated as a peak during the peak picking process.

Missing value imputation (MVI) has been studied for several decades to solve dataset incompleteness problems [64]. Missing values are a common occurrence in mass spectrometry-based metabolomics datasets and could have a significant effect on the statistical analysis if not treated appropriately.

Apart from their analytical origin, missing values can be classified into three groups [64]:

– Missing completely at random (MCAR). When the absence is unrelated to any observed variable or response.
– Missing at random (MAR). When the absence is related to one or more observed variables but not to the response.
– Missing not at random (MNAR). When the absence is related to the response itself.

Several statistical approaches are able to handle missing values (e.g., t-test and ANOVA on samples with missing data by disregarding the missing entries, or Bayesian PCA), but often it is more appropriate to apply methods that impute missing values prior to your statistical analysis [65]. The objective of such imputation methods is to produce a complete dataset by replacing missing values with plausible estimates.

The simplest method is to remove the metabolite feature containing a high number of missing values. In metabolomics, it is typical to use the 80/20 rule by removing features that are present at least in 80% of the samples. However, this method is only appropriate for large sample sets with a low percentage of missing values. If the sample size is limited, removing data may result in an inadequate sample size and limit the power of the statistical techniques applied in the data analysis. In addition, if the missing origin is MNAR, this strategy could potentially remove biomarkers (e.g., metabolite absence in cases and presence in controls).

One of the most routinely applied methods is mean value replacement. The missing metabolite feature is replaced by the mean value of the metabolite feature detected across the data set. In order to use the mean imputation method, the metabolite concentration should have a normal distribution—a Gaussian shape. However, this is not the case in most of the features detected in LC–MS studies.

Another method to impute missing values is the small value replacement. In metabolomics, it is typical that the missing value is replaced by a value half of the minimum peak intensity of the entire data set. This method would be expected to perform well when the missing values occur because the concentration of the metabolite is less than the limit of detection of the analytical method employed [64].

Other imputing methods rely on the distribution of related features in order to fill the missing value. K-nearest neighbor imputation (or KNN) uses Euclidean distance to identify the nearest neighbors of the missing feature and replace the missing value with an average of the nearest non-missing values [64]. Random Forest uses a proximity matrix to impute the missing values.

To improve the missing value imputation, a cross-validation criterion is applied. This helps reduce overfitting and produce more accurate imputations. Concretely, the data is split into training and test sets, and missing values are replaced in the training data. Hyperparameters (e.g., number of neighbors for KNN, number of trees for Random Forest) are fine-tuned to minimize the error of imputation in the training and testing dataset.

Some statistical approaches have been modified to handle a proportion of missing values; for example, missing values are imputed as part of the algorithm in Bayesian PCA. The missing values are determined using a principal component analysis regression with a Bayesian model. Although this approach has been applied for metabolomics data, it is mostly used in the processing of transcriptomics data.

In addition to these traditional methods for missing value imputation, new approaches are continuously being developed, such as rMisbeta, an R-package using robust estimators based on the minimum beta divergence method, which simultaneously imputes missing values and outliers [66].

3.2.5 Data Transformation, Scaling, and Normalization

In an LC–MS experiment, the signal drift is typically nonlinearly correlated with the injection order and batch number. The primary cause of this variation is the sample interaction with the analytical instrument, which leads to contamination producing variabilities in how metabolites are separated and detected. One of the most obvious symptoms of system aging is signal decay along the experiment. In addition, the limited capacity of both rack (i.e., vial container) and mobile phase inevitably introduces batch effects on the run. Finally, changes in the column pressure as well as temperature might contribute to the nonlinear drifts [25].

Data normalization, scaling, and standardization attempt to characterize and remove undesired variances while minimally affecting the underlying biological variance of the data. These preprocessing methods are often required before statistical analysis.

Scaling is transforming the data by applying a function to it. Normalization and standardization are both types of scaling. Standardization is changing the data to have a mean of zero and a standard deviation of one. Normalization for omics data usually refers to scaling each sample to a similar distribution of feature intensities. For example, one can express all peak intensities relative to their standard deviations. This is known as auto-scaling. Other well-known scaling methods in metabolomics include Pareto scaling, level scaling, and VAST scaling.

QCs can be used to remove unwanted variance. Since QCs don't contain intra-sample variability, they can be used as a standard to mathematically correct for signal drift over time and between batches and, in consequence, reduce the technical variation. This process increases the quality of the data while ensuring biological information and structure is not removed.

For those who seek more depth in the knowledge of strategies for data normalization and equalization, a good starting point is the review published by Wu et al. [67]. The simplest methods usually involve normalizing by the median or mean of the metabolites, assuming that all metabolites experiment with the same drift over the analyses. Logarithmic transformation is also a commonly used normalization technique since features tend to follow a log-normal distribution.

Another option makes use of internal standards (IS); known metabolites with known concentration are introduced to monitor its drift along the run [68]. Although this is the optimal solution, it is unfeasible in the majority of LC–MS experiments in terms of costs and invested time for IS production.

The major normalization methods used by LC/MS experimentalists are

- Quantile normalization methods. These methods consist of mapping time-varying changes in data density kernels forcing all samples to have identical peak-intensity distributions. These methods do not necessarily apply QC sample data [69].
- Regression normalization methods. These kinds of methods operate on each metabolite feature in isolation, correcting for sample-to-sample systematic variation using either linear or non-linear regression. The most prevalent smoothing algorithms are locally estimated scatterplot smoothing (LOESS) and its weighted version, locally weighted scatterplot smoothing (LOWESS) [70]. Advanced methods use QC samples to normalize samples allowing accurate correction of intra- and inter-batch variability.
- Multivariate statistics methods. Some approaches based on multivariate statistics assume that drift direction is retained on the first common principal component [71] and then its contribution is removed through component correction (CC). An evolution of CC is to consider that common variance (i.e., common principal component analysis (CPCA)) between samples and technical replicates where no biological variance is present (e.g., blanks) [72] retains drift information. Removing common contributions of first PCs followed by simple normalization technique is demonstrated as an effective intensity normalization approach.

3.3 Data Analysis

As described above, several data processing steps have to be carried out to remove systematic bias from the data, while preserving biological information. The next step is to use statistical techniques to extract the relevant and meaningful information from the processed data. Typically, the goal is to identify the metabolites that are significantly changing between classes of biological samples. This is not an easy step due to the wealth of information that needs to be inspected.

3.3.1 Data Visualization

Visualization of the data in easily interpretable plots and graphs is usually the first step in data analysis and is often referred to as exploratory data analysis (EDA). The main goals include assessing the data reproducibility, detecting possible outliers, and interesting metabolic differences between groups of samples.

In order to quickly identify quality issues in the data, statistical process control techniques are used. Statistical process control (SPC) techniques attempt to monitor the performance of a process over time to verify that it remains in a state of statistical control [73]. Basic monitoring charts such as Shewhart (i.e., mean intensity vs the injection order) [74], exponential weighted moving average (EWMA)[75], or cumulative sum (CUMSUM) [76] are commonly used to monitor the validity of the experiment using a unique variable.

Despite its wide usage, metabolomics data is high-dimensional and typically ill-conditioned (i.e., large number of detected features relative to the number of samples), leading to highly collinear datasets. Under those circumstances, the use of univariate quality control techniques is not suited. Then, data reduction or dimension reduction methods should be used.

PCA is often used for dimension reduction and visualization. Due to its unsupervised nature, PCA can also be used for exploration and quality control. The dimension reduction of the data in PCA is achieved by searching for a set of vectors (i.e., principal components) oriented so that it explains the largest percentage of the variance of the data. Principal components are constructed so that each one is orthogonal to the others and that each one retains more variance than the immediate previous. In short, PCA captures the largest variance in the data in a small set of vectors. This means that peaks with higher variance are likely to dominate the first principal components. This is undesirable in untargeted metabolomics where often the lower intense peaks are also of interest; hence, data scaling is desirable. It is important to note that the selection of the scaling method will imply slight differences in the output of PCA. For example, in Pareto scaling, the peak intensities are expressed relative to the square root of their standard deviations; hence, the contribution of high-variance peaks is reduced, but the "importance" of noisy signal is increased. A log-transformation can improve the PCA by reducing the scale of very large values and

reshaping the log-normal distribution of the peaks to a normal distribution and thus non-violating the PCA assumption of normality. The output of PCA are eigenvectors and eigenvalues. Eigenvalues and eigenvectors can be represented on score plots and loading plots. Score plots show a two-dimensional representation of the overall variance of the data, and the loading plots show which features contribute the most to each principal component.

In summarizing, PCA allows dimension reduction as well as selecting the best scaling/transformation/normalization approach for the dataset. Once the features are properly quantified in the intensity matrix, one can implement univariate or multivariate analysis methods to properly extract relevant information that properly characterizes the study [77].

3.3.2 Univariate Analysis

These methods analyze each feature independently to assess its statistical relevance. The main advantage of the univariate statistical tools is their easy interpretation [77] since the overall behavior of the data is not analyzed but rather each feature individually. The underlying assumption of univariate methods is that the metabolites are operating independently from each other. Clearly, this assumption is not correct since metabolites are related to each other via complex networks. Because of this, underlying metabolic relationships may not be underestimated. Such relationships can be detected by multivariate methods that take metabolite relationships into account. However, these methods can be difficult to apply to metabolomics data, especially when the number of variables in a data set is much larger than the number of samples. Therefore, univariate and multivariate approaches are complementary to each other and are both important approaches in the metabolomics toolbox.

The univariate feature selection is driven basically by the feature distribution [78]. Distribution normality is usually assessed through Kolmogorov–Smirnov tests and variance homogeneity through Barlett's tests.

If both tests are positive and features are known to follow normal distribution, ANOVA (analysis of variance, three or more groups) or t-test (two groups) can be used (so-called parametric methods) to determine differences between class and control. If the data distribution is not normal or unknown, nonparametric methods such as Mann–Whitney U test or Kruskal–Wallis can be used. Generally speaking, parametric models make more assumptions about the data and can therefore produce better results when these assumptions are correct. However, their results can be very misleading when the assumptions are incorrect.

For continuous outcome, linear regression, Pearson correlation is commonly used. A more extensive review on this topic can be found in [79].

It is important to note that multiple testing corrections are a major concern in omics data. The chance of getting a small *p*-value greatly increases when performing the same test hundreds or thousands of times. This is called a false-positive result. Multiple testing corrections are carried out to reduce the number of false positives. A common approach is the Bonferroni correction where the desired significance level is divided by the number of tests so that the overall significance level is close to the desired value. However, this test is often too restrictive, meaning that in many cases, groups that were actually different are not detected. An alternative approach is the Benjamini–Hochberg false discovery rate procedure, which aims to control the number of false-positive results among the variables where a significant difference was found.

These statistical methods explained above are available as both R and Python programming tools. An example of a user-friendly tool is MetaboAnalyst, which enables users the data analysis performance by interacting via a graphical web interface.

3.3.3 Multivariate Analysis Methods

In contrast to univariate, multivariate analysis takes simultaneously into account all the metabolic features [61]. Many multivariate approaches used in metabolomics require data reduction to reduce the data dimensionality. Multivariate analysis methods are mainly divided into two approaches:

- Unsupervised methods: These methods allow summarizing the metabolomics data, providing a feasible way to detect patterns related to the biological or experimental factors. PCA is the widest used unsupervised tool. This approach transforms the original data into a new orthogonal component subspace where its first components add most of the variance. The biological relevance of each metabolite is assessed by evaluating each metabolite contribution to the principal components, but as previously discussed, PCA is mostly used on QC procedures rather than inferring biological hypotheses. For example, a PCA showing a central distribution of QCs (i.e., pool of samples) indicates a healthy run [25]. Hierarchical clustering analysis (HCA) clusters metabolites by similarity (e.g., Euclidean Distance) in a hierarchical way. This distribution allows us to find similar metabolites that can be grouped as families, and in the same way also find similar families or groups of similar metabolites. These results are usually presented as a dendrogram.
- Supervised methods: These methods take advantage of the known response to identify the features that optimize the separability of classes in a classifier and maximize the explained variance. Partial least squares-discriminant analysis (PLS-DA) is one of the most popular supervised multivariate models in metabolomics. Similar to PCA, the output of a PLS-DA model

is scores and loadings; however, the scores now highlight the maximum differences between groups. The application of a supervised analysis method is always followed by validation steps to make sure that the observed differences are significant and generalize well to new data. Validation in its simplest form involves splitting the data table into two sets: a training set and a test set. The training set is used to construct the model. The test set is "blind" to the model training and is only used to assess how well the constructed model generalizes to new data. Often, this process is repeated several times using different splits of the data to obtain a more precise estimate of the model performance. This repeated splitting of the data to validate the model is known as cross-validation.

3.3.4 Peak Annotation

In any untargeted metabolomics study, raw data is collected without complete knowledge of which metabolites are present in the biological samples studied. The features captured during the study need to be annotated or identified as known metabolites. This is a critical step in the metabolomics workflow because no conclusion can be deduced from the data without knowing which metabolites are contributing the most to biological changes.

The terms identification and annotation are defined by the Chemical Analysis Working Group (CAWG) for Metabolomics Standards Initiative (MSI) [80]. Identification is the highest level of confidence and requires comparison of experimental data collected for the biological sample to an authentic chemical standard whose chemical structure is known. However, chemical standards for all known metabolites are yet to be commercialized and in consequence spectral standards databases and libraries are far from complete. Due to the latter issue, most of the features must be assigned to metabolites through a step called peak annotation. Peak annotation has lower confidence in the assignment of a chemical structure.

Peak annotation generally uses the three-dimensional space of LC–MS: first, the chromatographic retention time that is related to the chemical and physical properties of the metabolite (metabolites interact with the chromatographic stationary phase depending on their properties); second, the mass of the metabolite (with four decimal places) to calculate the empirical formula. In the cases of MSn, the fragmentation mass spectrum of each metabolite is used. MSn provides a greater resolution since it gives information related to the chemical structure of the metabolite, allowing the annotation of a unique metabolite structure.

Determination of the accurate mass will ideally provide a single empirical formula, though multiple empirical formulas can be reported in most cases. A single empirical formula can be linked to multiple metabolites; these metabolites are called isomers and

isomers have the same empirical formula but different three-dimensional spatial arrangements. A fragmentation mass spectrum can sometimes, but not always, differentiate between two or more isomers. The collection of a fragmentation mass spectrum is also referred to an MSn mass spectrum. Finally, the intensity profile is used to aggregate features that potentially originated from the same metabolite.

For LC–MS, there are also metabolite specific libraries that help researchers with the annotation step, including METLIN [81], MassBank of North America—MoNA [82], mzCloud [83], MassBank [84], and Global Natural Products Social Molecular Networking database—GNPS [85].

It is important as a first step to match biologically derived mass spectra to mass spectra in libraries where the data for both have been acquired identically with the same analytical platform, collision energy, and fragmentation mechanism. If this approach is not successful, then searching for other mass spectral library data, not necessarily acquired in the same approach, can be applied with caution. Several software tools, including xMSannotator [86], RAMclust [87], CAMERA [88], MetAssign [89], and mWISE [90], are available for metabolite annotation. Numerous compound repositories for metabolite annotation such as PubChem [91], Chemical Entities of Biological Interest (ChEBI) [92], KNApSAcK [93], LipidBank and LIPID MAPS [94], and Kyoto Encyclopedia of Genes and Genomes (KEGG) [95] are also accessible for metabolite annotation.

3.4 Functional Analysis

Once an interesting set of peaks (i.e., in terms of separability) are properly matched to some chemical compounds, functional analysis can be performed. Functional analysis attempts to give a deeper insight into complex biochemical relationships and introduce orthogonal information to get a better comprehension about the issue under study. Usually, two strategies are taken, pathway and network enrichment, although some tools combine both strategies [96].

The former uses biological knowledge to analyze patterns from an integrative point of view; the latter builds networks based on the high correlation degree and applies modularity analysis, among other techniques [77].

3.4.1 Pathway Enrichment

Metabolic pathways are groups of metabolites that are linked by one or several enzymatic reactions promoting the same biological process [77]. Methods to analyze pathway enrichment are commonly known as metabolite set enrichment analysis (MSEA) introduced in a similar way as gene set enrichment analysis (GSEA) [97]. The following revises three approaches implemented on MSEA [98]:

- Over-representation analysis (ORA): Fisher's exact test is implemented on a set of metabolites in order to quantify if they are over-represented on a given metabolic pathway.
- Quantitative enrichment analysis (QEA): Pathways are considered enriched when the concentration of their metabolites differentiates significantly over the rest.
- Single sample profiling (SSP): This method compares the intensity of the peaks from a set of metabolites to the normal intensity ranges of a sample.

Tools such as MetaboAnalyst [99] and Cytoscape [100] allow researchers to properly perform analysis of metabolites from a pathway level.

3.4.2 Network Analysis

Networks that are related to metabolite expression are known as metabolic networks. In those networks, nodes are metabolites and their relations are represented on the edges. This representation is also called a graph. Graphs can either be directed (i.e., information flows in a specific direction between nodes) or undirected (i.e., information flows forward and backward between nodes).

In correlation-based analysis, edges between nodes correspond to the mathematical pairwise correlation between metabolites. Correlation-based graphs are an example of undirected graphs. Other strategies run enzymatic reactions in-between metabolites on the edges [96, 101]. Those cases are an example of directed graphs. Measures of centrality (i.e., closeness of neighbors in function of the number of connections) are used to know how relevant a metabolite is within a specific network. If the network is biology based (i.e., edges encode enzymatic reactions), one can depict the importance of a metabolite in a specific pathway where this reaction is involved.

Other strategies compute label propagating algorithms (LPAs) and multi-LPA (MLPA) using labels of some nodes to propagate and mark unlabeled nodes in the network [102]. This strategy allows finding sub-modules within the network that can be suited under a certain label. For example, by using relations of metabolites with diseases (i.e., KEGG Disease Database), one can depict disease-related nodes. DiffuStats [103] is an R-package [104] that allows one to easily implement this kind of strategy on a given network.

4 Computational Tools

4.1 R-Based Tools

R programming language [104] is the widest election for metabolomics application. This contributed to a great expansion of the statistical tools implemented in this field. This knowledge is, in its majority, stored on the Bioconductor [105] repository. Some of the widest used packages are described below.

XCMS is the most used R-package for metabolomics applications. It is capable of analyzing both GC–MS and LC–MS data. Its functionalities include peak picking (centWave and matched filter algorithms), peak alignment (LOESS and Obiwarp [56]), and gap filling (i.e., imputation of non-detected peaks). It also allows statistical analysis using ANOVA and Student's t-Test in function of the number of classes present in the study. In addition, XCMS has an online implementation [106]. CAMERA is the widest election to perform peak annotation. This package requires as an input a peak table that outputs XCMS. It includes a table of adduct transformations that is fully user-customizable. CAMERA uses peak correlation and retention time windowing to group peaks previously to make the annotation. Similar to XCMS, CAMERA outputs a table with the peak groups and annotations found during the pipeline.

New R-packages are being continuously developed in order to improve any step of the data processing and analysis. Some examples are MaRR (a nonparametric approach that detects the change from reproducible to irreproducible signals using a maximal rank statistic [107]); and OUKS (an R-based open-source collection of scripts—Omics Untargeted Key Script, which provides comprehensive data processing by integrating various R-packages and metabolomics software tools, allowing the creation of a custom pipeline [108]). Metabolomics Analysis—An R Tool, SMART [109], and Specmine [110] are also examples of software packages written in R.

In addition, there are tools or packages to automatize or improve postprocessing steps, such as correlation-based removal of multiplicities (CROP) or neighbor-wise compound-specific graphical time warping (ncGTW). The first one, CROP, implemented as an R-package, is a visual postprocessing tool that removes redundant features from LC–MS/MS-based untargeted metabolomic data sets [111], where it groups highly correlated features within a defined RT window avoiding the condition of specific m/z difference making it a second-tier strategy for multiplicities reduction. The output is a graphical representation of a correlation network, allowing a good understanding of the cluster composition that can aid in further parameter tuning. The second one, ncGTW, is an integrated reference-free profile alignment method, implemented as an R-package and is available as a plugin for XCMS that aids in detecting and fixing the bad alignments (misaligned feature groups) in the LC–MS data to render accurate grouping and peak-filling [112].

Remove unwanted variation in a hierarchical structure (hRUV) is an R-package (also available as Shiny app) that aids in the removal of unwanted variation from large scale LC–MS metabolomics studies which it accomplishes by progressively merging the adjustments in neighboring batches [113]. The package uses sample replicates to integrate data from several batches for the removal of intra-batch

signal drift and inter-batch unwanted variation and outperforms existing tools while retaining biological variation. For assessment of the results, the user can visualize results as three kinds of diagnostic plots, i.e., PCA plots, relative log expression (RLE) plots, and metabolite run plots.

MetumpX is an R-package that facilitates easy download and installation of 103 tools spread across the standard untargeted MS-based metabolomics pipeline [114]. The package can aid in the automatic installation of software pipelines, truly speeding up the learning curve to build software workstations.

Other R-packages like MeTaQuaC aids in the implementation of concepts and methods for Biocrates kits and its application in targeted LC–MS metabolomics workflows and creates a QC report containing visualization and informative scores and provides summary statistics and unsupervised multivariate analysis methods, among others [115].

Dbnorm R-package allows visualization and removal of technical heterogeneity from large-scale metabolomics dataset, after allowing inspection at both macroscopic and microscopic scales at both sample batch and metabolic feature levels, respectively [116].

MetaClean, available as an R-package, combines 11 peak quality metrics and 8 machine learning algorithms to automatically detect poorly integrated peaks [117].

MAIT [118], which stands for Metabolite Automatic Identification Toolkit, is a wrapper tool. It combines in a unique end-to-end tool all the XCMS and CAMERA steps. It also adds statistical tools to validate the analysis and gives a novel alternative for the annotation based on biotransformation using the Human Metabolome Database (HMDB) [9].

Several software tools including xMSannotator [86], RAMclust [87], mWISE [90], and MetAssign [89] are available for metabolite annotation. Numerous compound repositories for metabolite annotation such as PubChem [91], ChEBI [92], KNApSAcK [93], LipidBank and LIPID MAPS [94], and KEGG [95]) are also accessible, which are obliging for metabolite annotation.

Metabox [119] is an R-based software framework for metabolomics data processing, statistical analysis, and functional analysis.

There are more software tools or packages developed to aid compound identification, such as Metabolomic Analysis and Visualization Engine (MAVEN) [120], Mass Spectrometry–Data Independent Analysis software (MS-DIAL) [38], OpenMS [121], and adaptive processing of liquid chromatography–mass spectrometry (apLCMS) [47]. All the major software programs are compatible with the open format such as mzML for feature extraction and quantification [122, 123].

Finally, there are also multifunctional tools that facilitate the data processing, annotation, statistical analysis, and interpretation of metabolite data.

4.2 XCMS Online

XCMS Online [106], Scripps Center for Metabolomics and Mass Spectrometry, is an integrated freely accessible cloud-based platform to facilitate raw MS spectra processing and visualization of untargeted LC–MS data. XCMS Online retains the same features as the original XCMS software [39] with more flexibility in web browsing and also provides a direct link to the METLIN [81] database for metabolite identification. It also provides interactive metabolomics cloud plots, retention time correction plots, univariate statistical analysis, and pathway information.

4.3 MetaboAnalyst

MetaboAnalyst 5.0 [99] is the most complete computational tool based on a web server to perform metabolomics data analysis, visualization, and interpretation. It includes tools to perform annotation, enrichment and pathway analysis, power and statistical analysis, batch correction, or database ID. For batch correction, a novel algorithm called WaveICA has been incorporated into the pipeline [124]. In addition, this algorithm has been recently improved (WaveICA 2.0) to remove inter-batch and run order effects for metabolomics data without using batch information [125].

Although it gives huge advantages in data processing, it does not include functionalities from the raw files and its structure is very rigid, which complicates an experienced user to undergo more complex and deeper analysis.

4.4 Workflow4metabolomics

Workflow4metabolomics [126] is developed by the French Bioinformatics Institute for data processing, statistical analysis, and interpretation of LC–MS data. It is built on a Galaxy environment with unique computational modules like data normalization, multivariate analysis, and annotation. Galaxy-M [127] is developed for LC–MS and DIMS metabolomic data analysis. It provides processing of raw data to preparation of statistical analysis and annotation in the galaxy-based platform. Both Workflow4metabolomics and Galaxy-M require implementations of XCMS [39] and CAMERA [88] packages for the analysis.

4.5 MZmine 2

MZmine 2 [37] is developed for both targeted and untargeted LC–MS data. It is an open-source data processing toolbox obtained under the license of GNU GPL (General Public License).

4.6 MS-Dial

MS-DIAL 4.0 [38] is developed for tandem mass fragmentation data in data-independent acquisition mode along with comprehensive identification and quantification of small molecules including lipids by mass spectral deconvolution.

4.7 IP4M

Integrated mass spectrometry-based untargeted data mining (IP4M) [49] covers all eight modules include data preprocessing, peak annotation, statistical analysis, pathway and enrichment

analysis, classification and biomarker detection, correlation analysis, regression analysis, and sample size and power analysis for untargeted LC–MS and GC–MS metabolomics data.

4.8 iMAP

iMAP has been developed to improve the IP4M modules and workflows according to the IP4M user's needs. Compared with IP4M, iMAP provides more optional methods and workflows with more adjustable parameters and improved figures. It also includes new modules for predictive model building and validation and correlation-based network construction and analysis resulting in a more mature, comprehensive, and modern platform to empower the metabolomics community [128].

5 Metabolomic Databases

In terms of enrichment, it's basic to interpret the results of a study in the frame of the current knowledge. This knowledge is stored in several databases. Those databases can be summarized in two groups: chemical and spectral databases and metabolomics databases. The first one contains factual information related to chemistry, chemical compounds, and their spectra. Factual is referred to images, numbers, texts, and graphics that describe their properties [129]. ChEBI [92] is a physiochemical metabolomics database. The latter includes databases that also take into account biological information that is complementary to the chemical information (i.e., disease and pathway associations, reactions, relationships, etc.). KEGG [95], HMDB [9], and Reactome [130] are examples of comprehensive metabolomics databases.

5.1 HMDB

The HMDB [9] evolved from physiochemical to metabolism-oriented database over its versions. Nowadays, it is a public web-enabled database that contains comprehensive information about the human metabolites together with their biological roles, physiological concentrations, disease associations, and chemical reactions among others.

In version HMDB 4.0 launched in 2018, HMDB contains information of 114.100 metabolites, 18.192 reactions, and 764 drugs linked to 494 metabolites through 2.497 associations, with up to 25.570 illustrated pathways divided into Metabolism or Catabolism Pathways (25.086), Metabolite Disease Pathways (213), Metabolite-Physiology Pathways (6), Drug Action Pathways (383), Drug Metabolism Pathways (64), and Metabolite Signaling Pathways (18).

In addition, there are tools such as hmdbQuery [131] in R language [104] that have been launched through Bioconductor [105] repository that eases the computational interaction with the HMDB database.

5.2 KEGG

KEGG is a broadly used database in its majority because of its curation [95]. KEGG was originally developed in 1995 as an integrated database resource for the biological interpretation of completely sequenced genomes by KEGG pathway mapping. Nowadays, KEGG content is mostly hand curated from research papers, and recently, drug labels and other regulatory documents have started to be analyzed. KEGG can be classified into four categories.

Same as HMDB, there are packages such as KEGGREST [132] for the R [104] in the Bioconductor that facilitate the interaction with the KEGG database.

Acknowledgments

This work would never have been possible without funding from the Call for "Ayudas para la Vinculación de Ténicos de Apoyo a ECAIs"; Funding Agency: Andalusian Health Service (References: RF2-0003-2020 and RF2-0003-2018). Ángela Peralbo-Molina would like to personally thank Pol Solà-Santos and Alexandre Perera-Luna for the opportunity to work together on this chapter.

References

1. Horgan RP, Kenny LC (2011) 'Omic' technologies: genomics, transcriptomics, proteomics and metabolomics. Obstet Gynaecol 13: 189–195. https://doi.org/10.1576/TOAG.13.3.189.27672
2. Stencel A, Crespi B (2013) What is a genome? Mol Ecol 22:3437–3443. https://doi.org/10.1111/MEC.12355
3. Lederberg J, Mccray AT (2001) 'Ome sweet' omics-a genealogical treasury of words. The Scientist 15:8–9
4. Ridley M (1999) Genome: the autobiography of a species in 23 chapters. 344
5. Wilkins M (2014) Proteomics data mining. Expert Rev Proteomics 6:599–603. https://doi.org/10.1586/EPR.09.81
6. Patti GJ, Yanes O, Siuzdak G Metabolomics: the apogee of the omic triology NIH public access. Nat Rev Mol Cell Biol 13:263–269. https://doi.org/10.1038/nrm3314
7. Shah SH, Newgard CB (2015) Integrated metabolomics and genomics: systems approaches to biomarkers and mechanisms of cardiovascular disease. Circ Cardiovasc Genet 8:410. https://doi.org/10.1161/CIRCGENETICS.114.000223
8. Cuperlovic-Culf M (2018) Machine learning methods for analysis of metabolic data and metabolic pathway modeling. Meta 8. https://doi.org/10.3390/METABO8010004
9. Wishart DS, Feunang YD, Marcu A et al (2018) HMDB 4.0: the human metabolome database for 2018. Nucleic Acids Res 46: D608–D617. https://doi.org/10.1093/NAR/GKX1089
10. Fiehn O (2002) Metabolomics – the link between genotypes and phenotypes. Plant Mol Biol 48(1):155–171. https://doi.org/10.1023/A:1013713905833
11. Roessner U, Beckles D (2009) Metabolite measurements. Plant Metab Netw:39–69. https://doi.org/10.1007/978-0-387-78745-9_3
12. Villas-Bas SG, Roessner U, Hansen MAE et al (2007) Metabolome analysis. Wiley, Hoboken. https://doi.org/10.1002/0470105518
13. Metz TO, Zhang Q, Page JS et al (2007) The future of liquid chromatography-mass spectrometry (LC-MS) in metabolic profiling and metabolomic studies for biomarker discovery. Biomark Med 1:159. https://doi.org/10.2217/17520363.1.1.159
14. Fuhrer T, Zamboni N (2015) High-throughput discovery metabolomics. Curr Opin Biotechnol 31:73–78. https://doi.org/10.1016/J.COPBIO.2014.08.006
15. Yan M, Xu G (2018) Current and future perspectives of functional metabolomics in disease studies – a review. Anal Chim Acta 1037: 41–54. https://doi.org/10.1016/J.ACA.2018.04.006

16. Bothwell JHF, Griffin JL (2011) An introduction to biological nuclear magnetic resonance spectroscopy. Biol Rev 86:493–510. https://doi.org/10.1111/J.1469-185X.2010.00157.X
17. Wang Y, Liu S, Hu Y et al (2015) Current state of the art of mass spectrometry-based metabolomics studies – a review focusing on wide coverage, high throughput and easy identification. RSC Adv 5:78728–78737. https://doi.org/10.1039/C5RA14058G
18. Pablos J, Villas-Bôas SG, Mas S et al (2005) Mass spectrometry in metabolome analysis. Mass Spectrom Rev 24(5):613–646. https://doi.org/10.1002/mas.20032
19. Zhang A, Sun H, Wang P et al (2011) Modern analytical techniques in metabolomics analysis. Analyst 137:293–300. https://doi.org/10.1039/C1AN15605E
20. Schmitt-Kopplin P, Frommberger M (2003) Capillary electrophoresis – mass spectrometry: 15 years of developments and applications. Electrophoresis 24:3837–3867. https://doi.org/10.1002/ELPS.200305659
21. Theodoridis G, Gika HG, Wilson ID (2011) Mass spectrometry-based holistic analytical approaches for metabolite profiling in systems biology studies. Mass Spectrom Rev 30: 884–906. https://doi.org/10.1002/MAS.20306
22. Rodrigues AM, Ribeiro-Barros AI, António C (2019) Experimental design and sample preparation in forest tree metabolomics. Meta 9. https://doi.org/10.3390/METABO9120285
23. Berg M, Vanaerschot M, Jankevics A et al (2013) LC-MS metabolomics from study design to data-analysis – using a versatile pathogen as a test case. Comput Struct Biotechnol J 4:e201301002. https://doi.org/10.5936/CSBJ.201301002
24. Yu H, Chen Y, Huan T (2021) Computational variation: an underinvestigated quantitative variability caused by automated data processing in untargeted metabolomics. Anal Chem. https://doi.org/10.1021/ACS.ANALCHEM.0C03381
25. Dudzik D, Barbas-Bernardos C, García A, Barbas C (2018) Quality assurance procedures for mass spectrometry untargeted metabolomics. A review. J Pharm Biomed Anal 147:149–173. https://doi.org/10.1016/J.JPBA.2017.07.044
26. Sands CJ, Gómez-Romero M, Correia G et al (2021) Representing the metabolome with high fidelity: range and response as quality control factors in LC-MS-based global profiling. Anal Chem 93:1924–1933. https://doi.org/10.1021/ACS.ANALCHEM.0C03848
27. Guan F, You Y, Fay S et al (2021) Novel algorithms for comprehensive untargeted detection of doping agents in biological samples. Anal Chem 93:7746–7753. https://doi.org/10.1021/ACS.ANALCHEM.1C01273
28. Caesar LK, Kvalheim OM, Cech NB (2018) Hierarchical cluster analysis of technical replicates to identify interferents in untargeted mass spectrometry metabolomics. Anal Chim Acta 1021:69–77. https://doi.org/10.1016/J.ACA.2018.03.013
29. Broadhurst D, Goodacre R, Reinke SN et al (2018) Guidelines and considerations for the use of system suitability and quality control samples in mass spectrometry assays applied in untargeted clinical metabolomic studies. Metabolomics 14(6):1–17. https://doi.org/10.1007/S11306-018-1367-3
30. Hendriks MMWB, van Eeuwijk FA, Jellema RH et al (2011) Data-processing strategies for metabolomics studies. TrAC Trends Anal Chem 30:1685–1698. https://doi.org/10.1016/J.TRAC.2011.04.019
31. Rafiei A, Sleno L (2014) Comparison of peak-picking workflows for untargeted liquid chromatography/high-resolution mass spectrometry metabolomics data analysis. Rapid Commun Mass Spectrom 29:119–127. https://doi.org/10.1002/RCM.7094
32. MarkerView TM Software 1.3.1 for Statistical analysis of MS datasets (2017)
33. Progenesis QI for proteomics-What's new in the latest release? (2017)
34. MassHunter Profinder Software Quick Start Guide
35. Compound Discoverer Software | Thermo Fisher Scientific - ES. https://www.thermofisher.com/es/es/home/industrial/mass-spectrometry/liquid-chromatography-mass-spectrometry-lc-ms/lc-ms-software/multi-omics-data-analysis/compound-discoverer-software.html. Accessed 22 Oct 2021
36. MetaboScape | Bruker. https://www.bruker.com/en/products-and-solutions/mass-spectrometry/ms-software/metaboscape.html. Accessed 22 Oct 2021
37. Pluskal T, Castillo S, Villar-Briones A, Orešič M (2010) MZmine 2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. BMC Bioinf 11:1–11. https://doi.org/10.1186/1471-2105-11-395

38. Tsugawa H, Cajka T, Kind T et al (2015) MS-DIAL: data independent MS/MS deconvolution for comprehensive metabolome analysis. Nat Methods 12:523. https://doi.org/10.1038/NMETH.3393
39. Smith CA, Want EJ, O'Maille G et al (2006) XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. Anal Chem 78:779–787. https://doi.org/10.1021/AC051437Y
40. Treviño V, Yañez-Garza IL, Rodriguez-López CE et al (2015) GridMass: a fast two-dimensional feature detection method for LC/MS. J Mass Spectrom 50:165–174. https://doi.org/10.1002/JMS.3512
41. Tautenhahn R, Böttcher C, Neumann S (2008) Highly sensitive feature detection for high resolution LC/MS. BMC Bioinf 9:1–16. https://doi.org/10.1186/1471-2105-9-504
42. Arnold K (2000) The Java programming language, 3rd edn. Addison-Wesley, Boston
43. Castillo S, Gopalacharyulu P, Yetukuri L, Orešič M (2011) Algorithms and tools for the preprocessing of LC–MS metabolomics data. Chemom Intell Lab Syst 108:23–32. https://doi.org/10.1016/J.CHEMOLAB.2011.03.010
44. Du P, Kibbe WA, Lin SM (2006) Improved peak detection in mass spectrum by incorporating continuous wavelet transform-based pattern matching. Bioinformatics 22:2059–2065. https://doi.org/10.1093/BIOINFORMATICS/BTL355
45. Riquelme G, Zabalegui N, Marchi P et al (2020) A python-based pipeline for preprocessing LC–MS data for untargeted metabolomics workflows. Metabolites 10:416. https://doi.org/10.3390/METABO10100416
46. Adams KJ, Pratt B, Bose N et al (2020) Skyline for small molecules: a unifying software package for quantitative metabolomics. J Proteome Res 19:1447–1458. https://doi.org/10.1021/ACS.JPROTEOME.9B00640
47. Yu T, Park Y, Johnson JM, Jones DP (2009) apLCMS-adaptive processing of high-resolution LC/MS data. Bioinformatics 25:1930–1936. https://doi.org/10.1093/BIOINFORMATICS/BTP291
48. Mayampurath AM, Jaitly N, Purvine SO et al (2008) DeconMSn: a software tool for accurate parent ion monoisotopic mass determination for tandem mass spectra. Bioinformatics (Oxford, England) 24:1021. https://doi.org/10.1093/BIOINFORMATICS/BTN063
49. Liang D, Liu Q, Zhou K et al (2020) IP4M: an integrated platform for mass spectrometry-based metabolomics data mining. BMC Bioinf 21. https://doi.org/10.1186/S12859-020-03786-X
50. Smith R, Ventura D, Prince JT (2013) LC-MS alignment in theory and practice: a comprehensive algorithmic review. Brief Bioinform 16:104–117. https://doi.org/10.1093/BIB/BBT080
51. Kassidas A, MacGregor JF, Taylor PA (1998) Synchronization of batch trajectories using dynamic time warping. AICHE J 44:864–875. https://doi.org/10.1002/AIC.690440412
52. Skov T, van den Berg F, Tomasi G, Bro R (2006) Automated alignment of chromatographic data. J Chemom 20:484–497. https://doi.org/10.1002/CEM.1031
53. Nielsen NPV, Carstensen JM, Smedsgaard J (1998) Aligning of single and multiple wavelength chromatographic profiles for chemometric data analysis using correlation optimised warping. J Chromatogr A 805:17–35. https://doi.org/10.1016/S0021-9673(98)00021-1
54. Sadygov RG, Maroto FM, Hühmer AFR (2006) ChromAlign: a two-step algorithmic procedure for time alignment of three-dimensional LC-MS chromatographic surfaces. Anal Chem 78:8207–8217. https://doi.org/10.1021/AC060923Y
55. Bylund D, Danielsson R, Malmquist G, Markides KE (2002) Chromatographic alignment by warping and dynamic programming as a pre-processing tool for PARAFAC modelling of liquid chromatography-mass spectrometry data. J Chromatogr A 961:237–244. https://doi.org/10.1016/S0021-9673(02)00588-5
56. Prince JT, Marcotte EM (2006) Chromatographic alignment of ESI-LC-MS proteomics data sets by ordered bijective interpolated warping. Anal Chem 78:6140–6152. https://doi.org/10.1021/AC0605344
57. Eilers PHC (2004) Parametric time warping. Anal Chem 76:404–411. https://doi.org/10.1021/AC034800E
58. Higgs RE, Knierman MD, Gelfanova V et al (2005) Comprehensive label-free method for the relative quantification of proteins from biological samples. J Proteome Res 4:1442–1450. https://doi.org/10.1021/PR050109B

59. Duran AL, Yang J, Wang L, Sumner LW (2003) Metabolomics spectral formatting, alignment and conversion tools (MSFACTs). Bioinformatics 19:2283–2293. https://doi.org/10.1093/BIOINFORMATICS/BTG315
60. Johnson KJ, Wright BW, Jarman KH, Synovec RE (2003) High-speed peak matching algorithm for retention time alignment of gas chromatographic data for chemometric analysis. J Chromatogr A 996:141–155. https://doi.org/10.1016/S0021-9673(03)00616-2
61. ISO 9000:2015(en), Quality management systems – fundamentals and vocabulary. https://www.iso.org/obp/ui#iso:std:iso:9000:ed-4:v1:en. Accessed 14 Oct 2021
62. Barwick V (2016) Eurachem/CITAC guide: guide to quality in analytical chemistry: an aid to accreditation. Retrieved Feb 19, 2018. https://www.eurachem.org/images/stories/Guides/pdf/Eurachem_CITAC_QAC_2016_EN.pdf
63. Martínez-Sena T, Luongo G, Sanjuan-Herráez D et al (2019) Monitoring of system conditioning after blank injections in untargeted UPLC-MS metabolomic analysis. Sci Rep 9:9822. https://doi.org/10.1038/S41598-019-46371-W
64. Lin W-C, Tsai C-F (2019) Missing value imputation: a review and analysis of the literature (2006–2017). Artif Intell Rev 53(2): 1487–1509. https://doi.org/10.1007/S10462-019-09709-4
65. Rashid W, Gupta MK (2021) A perspective of missing value imputation approaches. Adv Intell Syst Comput 1086:307–315. https://doi.org/10.1007/978-981-15-1275-9_25
66. Shahjaman M, Rahman MR, Islam T et al (2021) rMisbeta: a robust missing value imputation approach in transcriptomics and metabolomics data. Comput Biol Med 138: 104911. https://doi.org/10.1016/J.COMPBIOMED.2021.104911
67. Wu Y, Li L (2016) Sample normalization methods in quantitative metabolomics. J Chromatogr A 1430:80–95. https://doi.org/10.1016/J.CHROMA.2015.12.007
68. Sysi-Aho M, Katajamaa M, Yetukuri L, Orešič M (2007) Normalization method for metabolomics data using optimal selection of multiple internal standards. BMC Bioinf 8:1–17. https://doi.org/10.1186/1471-2105-8-93
69. Draisma HHM, Reijmers TH, van der Kloet F et al (2010) Equating, or correction for between-block effects with application to body fluid LC–MS and NMR metabolomics data sets. Anal Chem 82:1039–1046. https://doi.org/10.1021/AC902346A
70. Dunn WB, Broadhurst D, Begley P et al (2011) Procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid chromatography coupled to mass spectrometry. Nat Protoc 6(7): 1060–1083. https://doi.org/10.1038/nprot.2011.335
71. Artursson T, Eklo T, Lundstro I et al (2000) Drift correction for gas sensors using multivariate methods. J Chemom 14:711–723
72. Fernández-Albert F, Llorach R, Garcia-Aloy M et al (2014) Intensity drift removal in LC/MS metabolomics by common variance compensation. Bioinformatics 30: 2899–2905. https://doi.org/10.1093/BIOINFORMATICS/BTU423
73. Bayart D (2005) Economic control of quality of manufactured product (1931). In: Landmark writings in western mathematics 1640–1940. Elsevier, Amsterdam, pp 926–935. https://doi.org/10.1016/B978-044450871-3/50153-4
74. Roberts SW (1959) Control chart tests based on geometric moving averages. Technometrics 1:239–250. https://doi.org/10.1080/00401706.1959.10489860
75. Hunter JS (2018) The exponentially weighted moving average. J Qual Technol 18:203–210. https://doi.org/10.1080/00224065.1986.11979014
76. Woodward RH, Goldsmith PL. Cumulative sum techniques. Trove. https://trove.nla.gov.au/work/22348696. Accessed 10 Jan 2022
77. Alonso A, Marsal S, Julià A (2015) Analytical methods in untargeted metabolomics: state of the art in 2015. Front Bioeng Biotechnol 3:. https://doi.org/10.3389/FBIOE.2015.00023
78. Broadhurst DI, Kell DB Statistical strategies for avoiding false discoveries in metabolomics and related experiments. Metabolomics 2(4). https://doi.org/10.1007/s11306-006-0037-z
79. Gardinassi LG, Xia J, Safo SE, Li S (2017) Bioinformatics tools for the interpretation of metabolomics data. Curr Pharmacol Rep 6(3):374–383. https://doi.org/10.1007/S40495-017-0107-0
80. Sumner LW, Amberg A, Barrett D et al (2007) Proposed minimum reporting standards for chemical analysis chemical analysis working group (CAWG) metabolomics standards initiative (MSI). Metab Off J Metab Soc

3:211. https://doi.org/10.1007/S11306-007-0082-2

81. Smith CA, O'Maille G, Want EJ et al (2005) METLIN: a metabolite mass spectral database. Ther Drug Monit 27:747–751. https://doi.org/10.1097/01.FTD.0000179845.53213.39
82. MassBank of North America. https://mona.fiehnlab.ucdavis.edu/. Accessed 19 Oct 2021
83. mzCloud – Advanced Mass Spectral Database. https://www.mzcloud.org/. Accessed 19 Oct 2021
84. Horai H, Arita M, Kanaya S et al (2010) MassBank: a public repository for sharing mass spectral data for life sciences. J Mass Spectrom 45:703–714. https://doi.org/10.1002/JMS.1777
85. Wang M, Carver JJ, Phelan V et al (2016) Sharing and community curation of mass spectrometry data with global natural products social molecular networking. Nat Biotechnol 34(8):828–837. https://doi.org/10.1038/nbt.3597
86. Uppal K, Walker DI, Jones DP (2017) xMSannotator: an R package for network-based annotation of high-resolution metabolomics data. Anal Chem 89:1063–1067. https://doi.org/10.1021/ACS.ANALCHEM.6B01214
87. Broeckling CD, Afsar FA, Neumann S et al (2014) RAMClust: a novel feature clustering method enables spectral-matching-based annotation for metabolomics data. Anal Chem 86:6812–6817. https://doi.org/10.1021/AC501530D
88. Kuhl C, Tautenhahn R, Böttcher C et al (2012) CAMERA: an integrated strategy for compound spectra extraction and annotation of LC/MS data sets. Anal Chem 84:283. https://doi.org/10.1021/AC202450G
89. Daly R, Rogers S, Wandy J et al (2014) MetAssign: probabilistic annotation of metabolites from LC–MS data using a Bayesian clustering approach. Bioinformatics 30:2764. https://doi.org/10.1093/BIOINFORMATICS/BTU370
90. Barranco-Altirriba M, Solà-Santos P, Picart-Armada S et al (2021) MWISE: an algorithm for context-based annotation of liquid chromatography-mass spectrometry features through diffusion in graphs. Anal Chem 93:10772–10778. https://doi.org/10.1021/ACS.ANALCHEM.1C00238/SUPPL_FILE/AC1C00238_SI_001.PDF
91. Wheeler DL, Barrett T, Benson DA et al (2008) Database resources of the National Center for Biotechnology Information. Nucleic Acids Res 36:D13–D21. https://doi.org/10.1093/NAR/GKM1000
92. Degtyarenko K, de Matos P, Ennis M et al (2008) ChEBI: a database and ontology for chemical entities of biological interest. Nucleic Acids Res 36:D344–D350. https://doi.org/10.1093/NAR/GKM791
93. Shinbo Y, Nakamura Y, Altaf-Ul-Amin M et al (2006) KNApSAcK: a comprehensive species-metabolite relationship database. Biotechnol Agric For 57:165–181. https://doi.org/10.1007/3-540-29782-0_13
94. Sud M, Fahy E, Cotter D et al (2007) LMSD: LIPID MAPS structure database. Nucleic Acids Res 35:D527–D532. https://doi.org/10.1093/NAR/GKL838
95. Kanehisa M, Goto S (2000) KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Res 28:27–30. https://doi.org/10.1093/NAR/28.1.27
96. Picart-Armada S, Fernández-Albert F, Vinaixa M et al (2017) Null diffusion-based enrichment for metabolomics data. PLoS One 12:e0189012. https://doi.org/10.1371/JOURNAL.PONE.0189012
97. Khatri P, Sirota M, Butte AJ (2012) Ten years of pathway analysis: current approaches and outstanding challenges. PLoS Comput Biol 8:e1002375. https://doi.org/10.1371/JOURNAL.PCBI.1002375
98. Xia J, Wishart DS (2010) MSEA: a web-based tool to identify biologically meaningful patterns in quantitative metabolomic data. Nucleic Acids Res 38:W71. https://doi.org/10.1093/NAR/GKQ329
99. Pang Z, Chong J, Zhou G et al (2021) MetaboAnalyst 5.0: narrowing the gap between raw spectra and functional insights. Nucleic Acids Res 49:W388–W396. https://doi.org/10.1093/NAR/GKAB382
100. Shannon P, Markiel A, Ozier O et al (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 13:2498–2504. https://doi.org/10.1101/GR.1239303
101. Orth JD, Thiele I, Palsson BØ (2010) What is flux balance analysis? Nat Biotechnol (3):28, 245–248. https://doi.org/10.1038/nbt.1614
102. Jiang X, Zhang H, Quan X et al (2017) Disease-related gene module detection based on a multi-label propagation clustering algorithm. PLoS One 12:e0178006. https://doi.org/10.1371/JOURNAL.PONE.0178006
103. Picart-Armada S, Thompson WK, Buil A, Perera-Lluna A (2018) diffuStats: an R package to compute diffusion-based scores on

biological networks. Bioinformatics 34: 533–534. https://doi.org/10.1093/BIOINFORMATICS/BTX632

104. The R Development Core Team (2008) R: a language and environment for statistical computing
105. Huber W, Carey VJ, Gentleman R et al (2015) Orchestrating high-throughput genomic analysis with Bioconductor. Nat Methods 12:115–121. https://doi.org/10.1038/NMETH.3252
106. Tautenhahn R, Patti GJ, Rinehart D, Siuzdak G (2012) XCMS online: a web-based platform to process untargeted metabolomic data. Anal Chem 84:5035–5039. https://doi.org/10.1021/AC300698C
107. Ghosh T, Philtron D, Zhang W et al (2021) Reproducibility of mass spectrometry based metabolomics data. BMC Bioinf 22:1–25. https://doi.org/10.1186/S12859-021-04336-9
108. Plyushchenko IV, Fedorova ES, Potoldykova NV et al (2021) Omics untargeted key script: R-based software toolbox for untargeted metabolomics with bladder cancer biomarkers discovery case study. J Proteome Res 21(3): 833–847. https://doi.org/10.1021/ACS.JPROTEOME.1C00392
109. Liang Y-J, Lin Y-T, Chen C-W et al (2016) SMART: statistical metabolomics analysis – an R tool. Anal Chem 88:6334–6341. https://doi.org/10.1021/ACS.ANALCHEM.6B00603
110. Costa C, Maraschin M, Rocha M (2016) An R package for the integrated analysis of metabolomics and spectral data. Comput Methods Prog Biomed 129:117–124. https://doi.org/10.1016/J.CMPB.2016.01.008
111. Kouřil Š, de Sousa J, Václavík J et al (2020) CROP: correlation-based reduction of feature multiplicities in untargeted metabolomic data. Bioinformatics 36:2941–2942. https://doi.org/10.1093/BIOINFORMATICS/BTAA012
112. Wu CT, Wang Y, Wang Y et al (2020) Targeted realignment of LC-MS profiles by neighbor-wise compound-specific graphical time warping with misalignment detection. Bioinformatics 36:2862–2871. https://doi.org/10.1093/BIOINFORMATICS/BTAA037
113. Kim T, Tang O, Vernon ST et al (2020) hRUV: hierarchical approach to removal of unwanted variation for large-scale metabolomics data. bioRxiv:2020.12.21.423723. https://doi.org/10.1101/2020.12.21.423723
114. Wajid B, Iqbal H, Jamil M et al (2020) MetumpX – a metabolomics support package for untargeted mass spectrometry. Bioinformatics 36:1647–1648. https://doi.org/10.1093/BIOINFORMATICS/BTZ765
115. Kuhring M, Eisenberger A, Schmidt V et al (2020) Concepts and software package for efficient quality control in targeted metabolomics studies – MeTaQuaC. bioRxiv:2020.01.10.901710. https://doi.org/10.1101/2020.01.10.901710
116. Bararpour N, Gilardi F, Carmeli C et al (2021) DBnorm as an R package for the comparison and selection of appropriate statistical methods for batch effect correction in metabolomic studies. Sci Rep 11:1–13. https://doi.org/10.1038/s41598-021-84824-3
117. Chetnik K, Petrick L, Pandey G (2020) MetaClean: a machine learning-based classifier for reduced false positive peak detection in untargeted LC–MS metabolomics data. Metab Off J Metab Soc 16:117. https://doi.org/10.1007/S11306-020-01738-3
118. Fernández-Albert F, Llorach R, Andrés-Lacueva C, Perera A (2014) An R package to analyse LC/MS metabolomic data: MAIT (metabolite automatic identification toolkit). Bioinformatics 30:1937–1939. https://doi.org/10.1093/BIOINFORMATICS/BTU136
119. Wanichthanarak K, Fan S, Grapov D et al (2017) Metabox: a toolbox for metabolomic data analysis, interpretation and integrative exploration. PLoS One 12:e0171046. https://doi.org/10.1371/JOURNAL.PONE.0171046
120. Clasquin MF, Melamud E, Rabinowitz JD (2012) LC-MS data processing with MAVEN: a metabolomic analysis and visualization engine. Curr Protoc Bioinformatic 14: Unit14.11. https://doi.org/10.1002/0471250953.BI1411S37
121. Sturm M, Bertsch A, Gröpl C et al (2008) OpenMS – an open-source software framework for mass spectrometry. BMC Bioinf 9: 1–11. https://doi.org/10.1186/1471-2105-9-163
122. Kessner D, Chambers M, Burke R et al (2008) ProteoWizard: open source software for rapid proteomics tools development. Bioinformatics 24:2534–2536. https://doi.org/10.1093/BIOINFORMATICS/BTN323
123. Martens L, Chambers M, Sturm M et al (2011) mzML—a community standard for mass spectrometry data. Mol Cell Proteomics 10:R110.000133. https://doi.org/10.1074/MCP.R110.000133

124. Deng K, Zhang F, Tan Q, Huang Y, Song W, Rong Z, Zhu Z-J, Li K, Li Z (2019) WaveICA: a novel algorithm to remove batch effects for large-scale untargeted metabolomics data based on wavelet analysis. Anal Chim Acta 1061:60–69. https://doi.org/10.1016/j.aca.2019.02.010
125. Deng K, Zhao F, Rong Z et al (2021) WaveICA 2.0: a novel batch effect removal method for untargeted metabolomics data without using batch information. Metabolomics 17:1–8. https://doi.org/10.1007/S11306-021-01839-7
126. Giacomoni F, le Corguillé G, Monsoor M et al (2015) Workflow4Metabolomics: a collaborative research infrastructure for computational metabolomics. Bioinformatics 31: 1493–1495. https://doi.org/10.1093/BIOINFORMATICS/BTU813
127. Davidson RL, Weber RJM, Liu H et al (2016) Galaxy-M: a Galaxy workflow for processing and analyzing direct infusion and liquid chromatography mass spectrometry-based metabolomics data. GigaScience:5. https://doi.org/10.1186/S13742-016-0115-8
128. Zhou D, Zhu W, Sun T et al (2021) iMAP: a web server for metabolomics data integrative analysis. Front Chem 0:262. https://doi.org/10.3389/FCHEM.2021.659656
129. Rumble JR Jr, Lide DR Jr (2002) Chemical and spectral databases: a look into the future. J Chem Inf Comput Sci 25:231–235. https://doi.org/10.1021/CI00047A018
130. Fabregat A, Sidiropoulos K, Viteri G et al (2017) Reactome pathway analysis: a high-performance in-memory approach. BMC Bioinf 18:1–9. https://doi.org/10.1186/S12859-017-1559-2
131. Bioconductor – hmdbQuery. https://www.bioconductor.org/packages/release/bioc/html/hmdbQuery.html. Accessed 20 Oct 2021
132. Bioconductor – KEGGREST. https://bioconductor.org/packages/release/bioc/html/KEGGREST.html. Accessed 20 Oct 2021

Chapter 21

Data Processing and Analysis in Liquid Chromatography–Mass Spectrometry-Based Targeted Metabolomics

Masahiro Sugimoto, Yumi Aizawa, and Atsumi Tomita

Abstract

Mass spectrometry (MS)-based metabolomics provides high-dimensional datasets; that is, the data include various metabolite features. Data analysis begins by converting the raw data obtained from the MS to produce a data matrix (metabolite × concentrations). This is followed by several steps, such as peak integration, alignment of multiple data, metabolite identification, and calculation of metabolite concentrations. Each step yields the analytical results and the accompanying information used for the quality assessment of the anterior steps. Thus, the measurement quality can be analyzed through data processing. Here, we introduce a typical data processing procedure and describe a method to utilize the intermediate data as quality control. Subsequently, commonly used data analysis methods for metabolomics data, such as statistical analyses, are also introduced.

Key words Mass spectrometry, Data processing, Statistical analysis, Multivariate analysis

1 Introduction

Metabolomics enables us a holistic view of metabolic pathways to identify and quantify various metabolites. However, analytical techniques employed still have limitations, and a single method cannot simultaneously analyze the whole metabolites that participate in a pathway; therefore, several methods need to be integrated to cover a wider range [1]. Nuclear magnetic resonance (NMR), which consists of nondestructive sample processing [2], and mass spectrometry (MS) are routinely used to profile metabolites. Besides, mass spectrometry (MS) with some separation systems are also common methods showing more sensitivity and thus usually result in the detection of a larger number of metabolites. Gas chromatography–mass spectrometry (GC–MS) is a separation system that has been used to analyze volatile and nonvolatile compounds with chemical derivatization [3]. Usually, electron ionization (EI) is

Raúl González-Domínguez (ed.), *Mass Spectrometry for Metabolomics*,
Methods in Molecular Biology, vol. 2571, https://doi.org/10.1007/978-1-0716-2699-3_21,
© The Author(s), under exclusive license to Springer Science+Business Media, LLC, part of Springer Nature 2023

used as an interface for MS, where molecular ions are fragmented and the peaks are identified by a homology search algorithm. Capillary electrophoresis–MS (CE–MS) [4] and liquid chromatography (LC)–MS [5] analyze charged and solvent-soluble metabolites, respectively. Both LC and CE are usually integrated with an electrospray ionization (ESI) interface that ionizes the molecules without compromising the signal. Although it is expected that a single peak corresponding to a single metabolite is observed, various other peaks, including isotopic, fragment, and noise peaks, are also observed. Furthermore, these peaks are sometimes skewed by background noise. In addition, the retention or migration times of each peak may shift, and the alignment of multiple samples is necessary; for example, the corresponding peak must be assigned to the correct metabolite among various samples by eliminating such shift data. Considering the features of the data that is instrument- and sample processing method-specific, appropriate data processing is necessary.

Given the various types of separation systems and MS types, we have only focused here on the data processing of LC–MS. In particular, LC–ESI–time-of-flight-MS (LC–ESI–TOFMS) is of interest, while the processes, which we have introduced, are similar to those of a single MS, such as CE–TOFMS. This combination is widely used for various applications, including targeted and non-targeted analyses. Metabolite identification using GC–MS and tandem MS (e.g., MS/MS) differs from that using LC–MS (single MS) because the former uses fragmented data, whereas the latter uses less informative data for each metabolite. However, the concepts of noise reduction, alignment, quantification, normalization, and quality control (QC) are similar among these methods.

This chapter is composed of two parts: (1) raw data processing and (2) data analysis. The first part includes straightforward analytical processes. However, QC or quality assessment of each step is also an important aspect [6, 7]. Hence, the effects of the quality assessment of sample processing, measurement, data processing itself, and the intra- and inter-batch are within the scope of this study. The overall analytical flow is illustrated in Fig. 1. The pre-conditions of sampling (e.g., the fasting duration and the last meal of the patient before the sampling), sampling method, sample storage, and sample transfer also have important effects on the metabolomics profiles in clinical research [8–10]. This warrants establishing the standard of operations (SOPs) to yield reproducible profiles [11]; however, this study does not address this aspect.

Recently, artificial intelligence (AI), especially machine learning methods, has been introduced for QC of data processing in metabolomics [12]. To use these machine learning methods, the answer, i.e., the good and bad quality datasets, needs to be prepared and the machine learning models need to be trained to predict the quality of the newly observed data. There are two types of data processing:

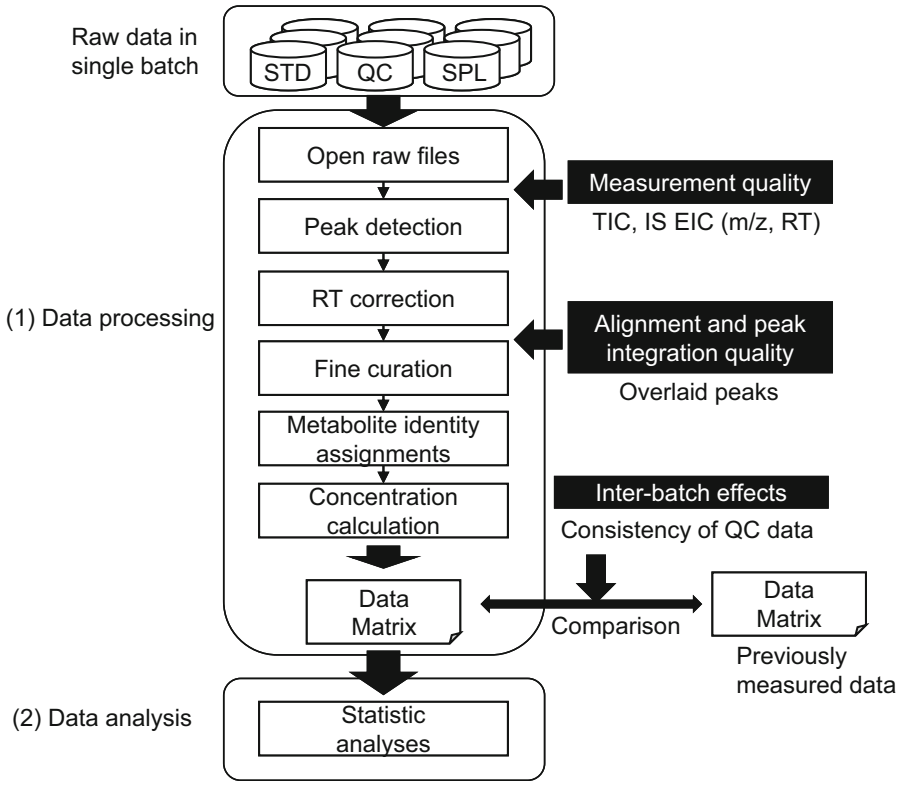

Fig. 1 The overall flow of data processing and data analysis. The standard mixture, QC samples, and sample files are simultaneously analyzed. In data processing (1), these data are opened and processed to yield a data matrix, which includes metabolite concentrations. Through the process, measurement, alignment, and peak integration qualities are evaluated based on total ion chromatogram, IS, EIC, and overlaid peaks. The data matrix, especially the data of QC samples in the matrix, is compared to that of previous batch measurements. Finally, the data matrix is used for downstream data analysis (2), such as statistical analyses

targeted and nontargeted analyses [13]. The latter includes uncharacterized peaks without corresponding metabolite compounds and the application of sophisticated QC using AI-based techniques [14]. Here, we introduce the fundamental process of the targeted metabolomics data analysis without the use of AI. We have used Agilent Mass Hunter software, which has been used for various versions of GC–MS, CE–MS, and LC–MS data. Although the application of this software is limited to Agilent products, the protocols presented here are not software-specific. Furthermore, we have used human serum as an example [15], but the protocol is applicable to any type of sample. However, depending on the sample type, the metabolite concentration should be normalized; for example, if urine samples are used, the metabolite division by creatinine or total peak area must be created to eliminate biases of whole concentrations of each urine sample [16].

The second part includes downstream analysis after obtaining the quantified concentration of each metabolite, such as statistical analysis. Statistical analysis is more widely applicable as it can be used for any type of metabolomic data, including data from NMR and other separation systems integrated with MS. The statistical analyses require updated software and databases because each tool is optimized for the specific dataset [17]. Here, we have described a protocol using free open web tools to conduct common statistical analysis of metabolomics data [18].

2 Methods

2.1 Designing Sample Sequence

Prepare the sequence to be measured to identify the metabolite names and quantify their concentrations as follows (Fig. 2):

1. Prepare standard mixtures, including standard metabolites and internal standards (IS). The vials, including this mixture, were placed before and after all the other vials.
2. Prepare the processed samples. The IS should be mixed during preprocessing.
3. Insert the QC samples, which also contain the IS, at every 10 times samples (*see* **Note 1**).

2.2 Raw Data Processing

1. Open all the raw data (. d files) in a single batch in MassHunter (Workstation Software, Qualitative Analysis Navigator).
2. Use the overlaid total ion chromatogram (TIC) (Fig. 3) to evaluate the instrument conditions (*see* **Notes 2** and **3**).
3. Display the IS peaks (Fig. 4) to evaluate the instrument and measurement conditions (*see* **Note 4**). An error of 0.1 min in retention time (RT) and 10 ppm of the *m/z* value between batches is acceptable.
4. Close the software.

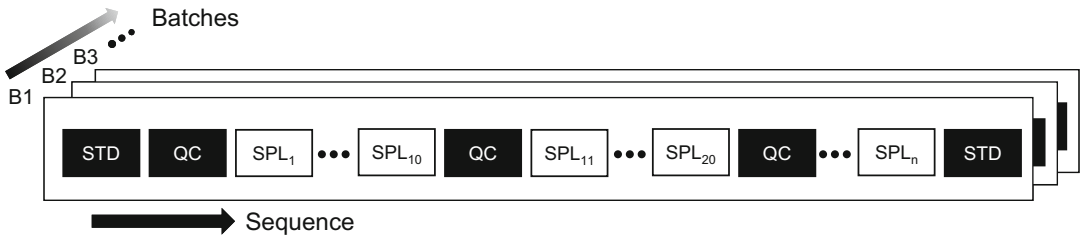

Fig. 2 A typical sequence design of a single batch. A batch includes standard mixtures (STD), which are measured first and last in the batch. QC samples are inserted every 10 samples ($SPL_{1...n}$). The QC samples are also measured in all batches (B_1, B_2, B_3)

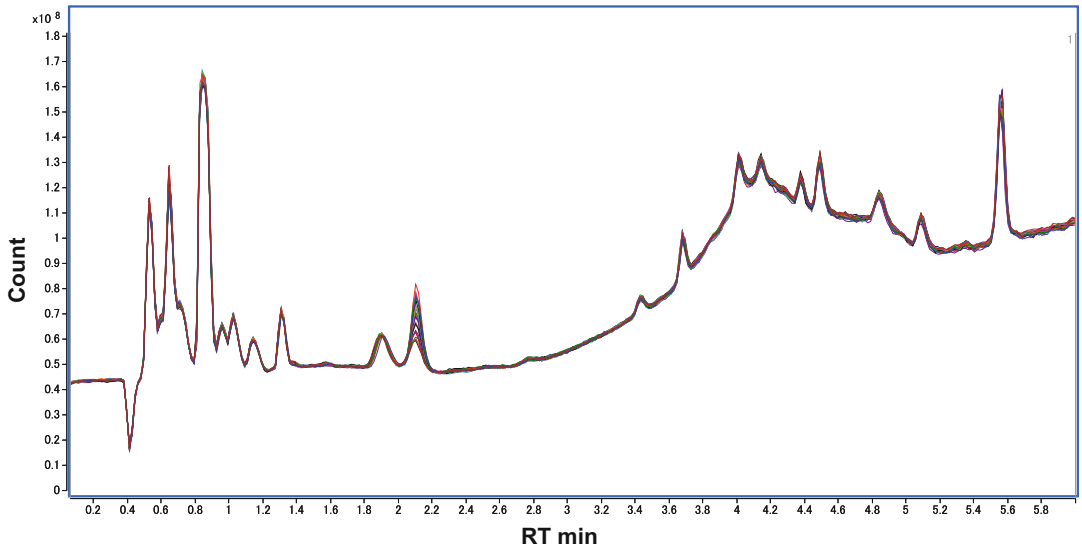

Fig. 3 The TIC data of 20 samples are overlaid

5. Create a batch file, a type of project management file, and register all raw data files after restarting MassHunter.
6. Load a method file that describes the predefined *m/z* and RT values to detect peaks and integrate their peak area. A tolerance of ±1.0 min RT and m/z ±50 ppm is used.
7. Apply RT to all sample data to ensure that the RT in the samples is corrected to the RT in the standard mixture; i.e., the alignment of multiple samples is ensured.
8. Display the overlaid peaks in the extracted ion chromatography (EIC) to evaluate the alignment quality.
9. Conduct peak detection and integration with different options if incorrectly detected or aligned peaks are observed (Fig. 5, see SPL5).
10. Assign metabolite identities to each peak based on *m/z* and RT. Eliminate the detected peaks that exhibit differences larger than 0.1 min for RT and 20 ppm for the *m/z* value. Peaks showing signal/noise (S/N) ratios less than 3 are treated as undetectable peaks.
11. Export the relative area file as a CSV file.
12. Calculate the metabolite concentration based on the area of the metabolite peak divided by the area of IS peak (relative peak area). Then, based on the relative peak areas in the sample and standard mixture, calculate the metabolite concentration (*see* **Note 5**).

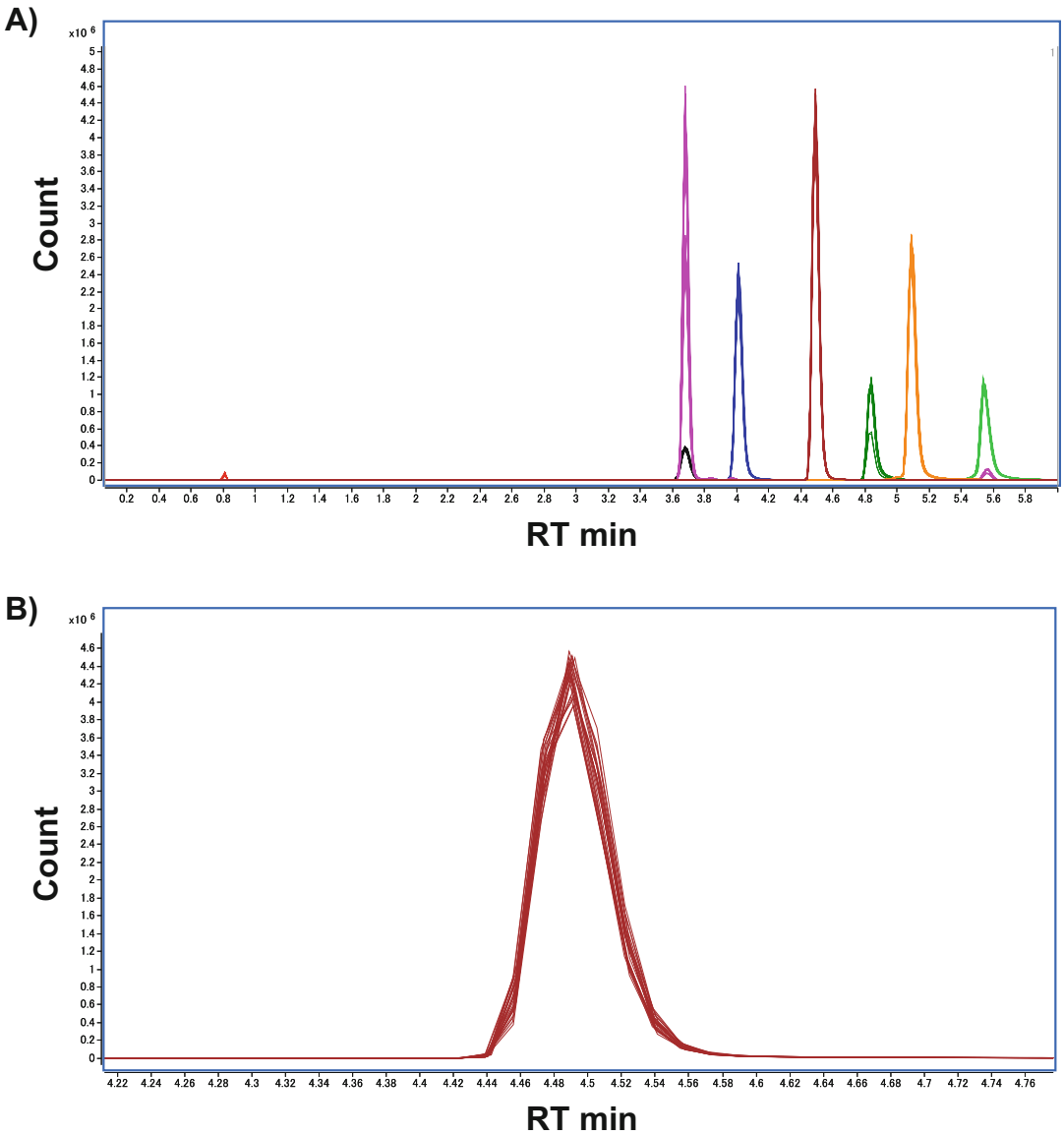

Fig. 4 Overlaid chromatograms of IS. (**a**) EIC and (**b**) peaks of an IS

2.3 Statistical Analysis of Quantified Data

1. Prepare the data matrix as a CSV file or a tab-separated TXT file. The first column includes the metabolite names or metabolite identities, such as KEGG and HMDB IDs. The first and second lines include sample and group names, respectively. For example, to compare two groups, such as the patient and healthy control groups, the characters P and C are used as group names. The other cells included metabolite concentrations (numerical values). Matrix-transported data are also acceptable.

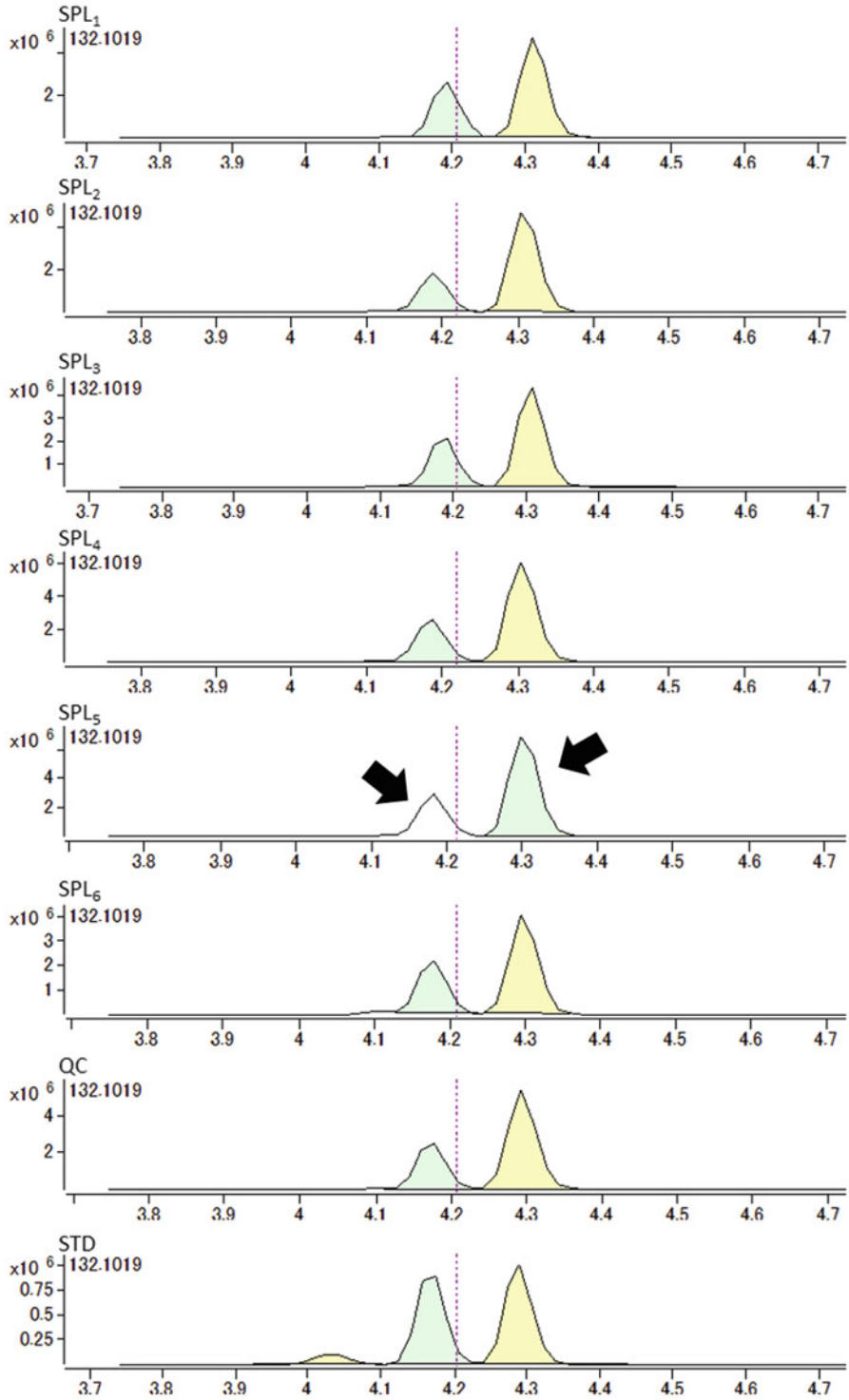

Fig. 5 An example of aligned extracted ion chromatograms. In this example, the left peak of SPL_5 was not detected and the right peak is misaligned

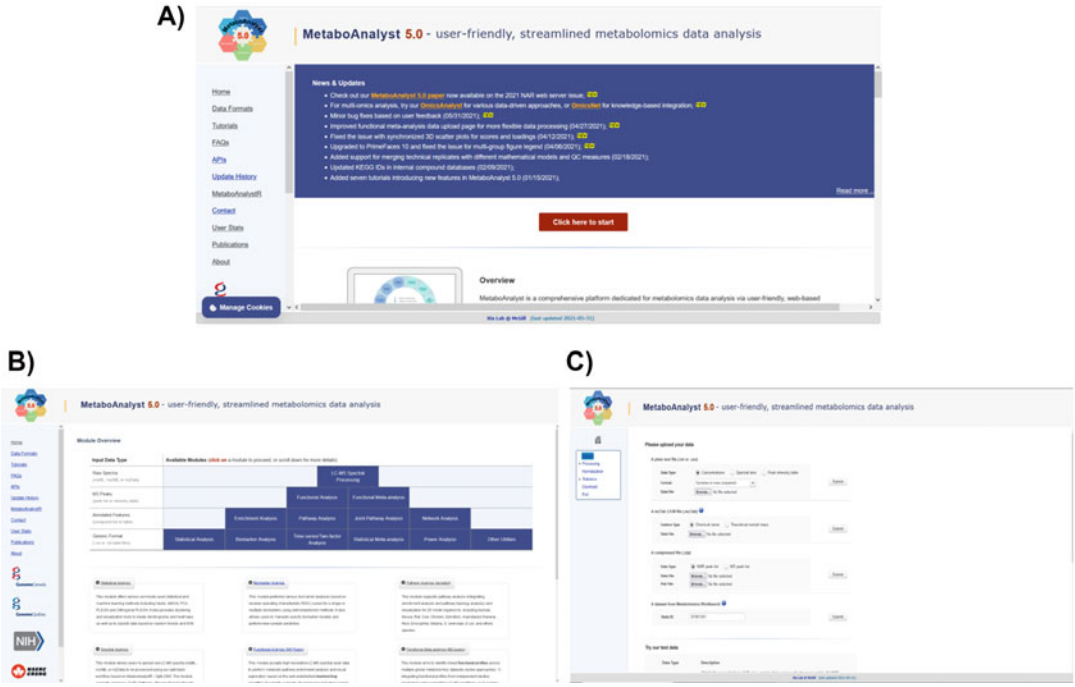

Fig. 6 Website of MetaboAnalyst. (**a**) Top page, (**b**) the menu of the data analysis, and (**c**) the data upload page

2. Access the website of MetaboAnalyst (https://www.metaboanalyst.ca/) [19] (Fig. 6) and upload the prepared data.
3. Select the principal component (PC) analysis with the data normalization option. There are three options: sample normalization, data transformation, and data scaling (*see* **Note 6**).
4. Visualize the score and loading plots (Fig. 7). If outlier samples are observed in the score plots, they are eliminated and reanalyzed.
5. Conduct discrimination analysis (Fig. 8). There are several methods, such as Volcano plots and partial least squares discriminant analysis. The former is used for discriminating two group data and the latter is used for two or more group data.
6. Conduct enrichment analysis (Fig. 9) to detect the presence of biologically meaningful metabolite groups (with or without concentration data) that are different between two metabolism states.
7. Conduct pathway analysis (Fig. 10) to rank the pathways based on the differences in the metabolic profiles, implying a pathway impact. The metabolites exhibiting a difference are located in the topologically central node; that is, the metabolites with a broad impact on the pathway are evaluated as a pathway impact (*see* **Note 7**).

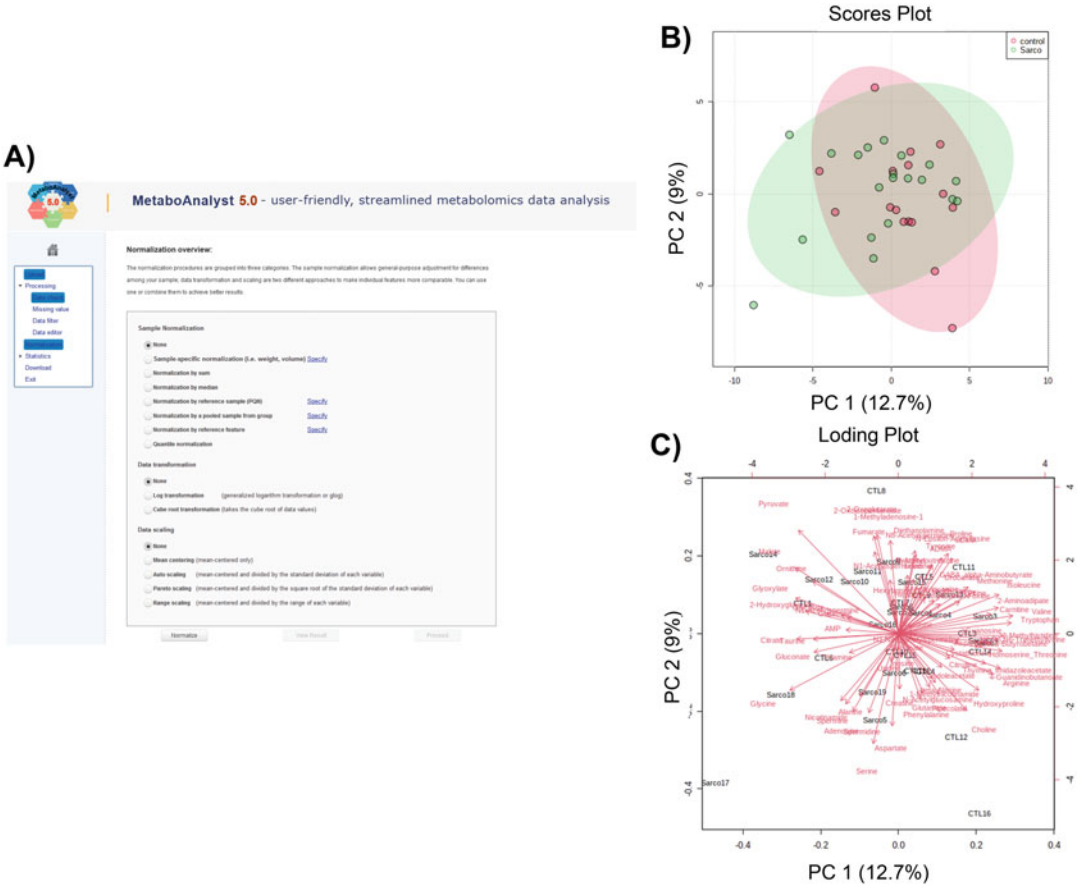

Fig. 7 PC analysis. (**a**) Data normalization page, (**b**) score plots, and (**c**) loading plots. The overall concentration of each sample is normalized using a sample normalization option. For example, select a normalization by the reference feature option to divide the metabolite concentration in urine samples by the creatinine concentration. Change the distribution of the metabolite concentration distribution by a data transformation option. Select a data scaling option, if necessary. For example, an auto-scaling option transfers the metabolite concentration into Z-scores. In score plots, a dot indicates a sample, and a shorter distance between two dots indicates a high similarity between the metabolic profiles of the two samples. Loading plots indicate the contribution of the metabolite concentrations

3 Notes

1. QC samples are used for correcting quantified values of each metabolite if unexpected bias among multiple biases is observed. Therefore the samples of which all peaks are detected in enough S/N ratio is preferable; i.e., all peaks should be detected at the center of dynamic ranges of each metabolite. Since QC samples are measured multiple times, multiple samples were mixed to increase the volume.

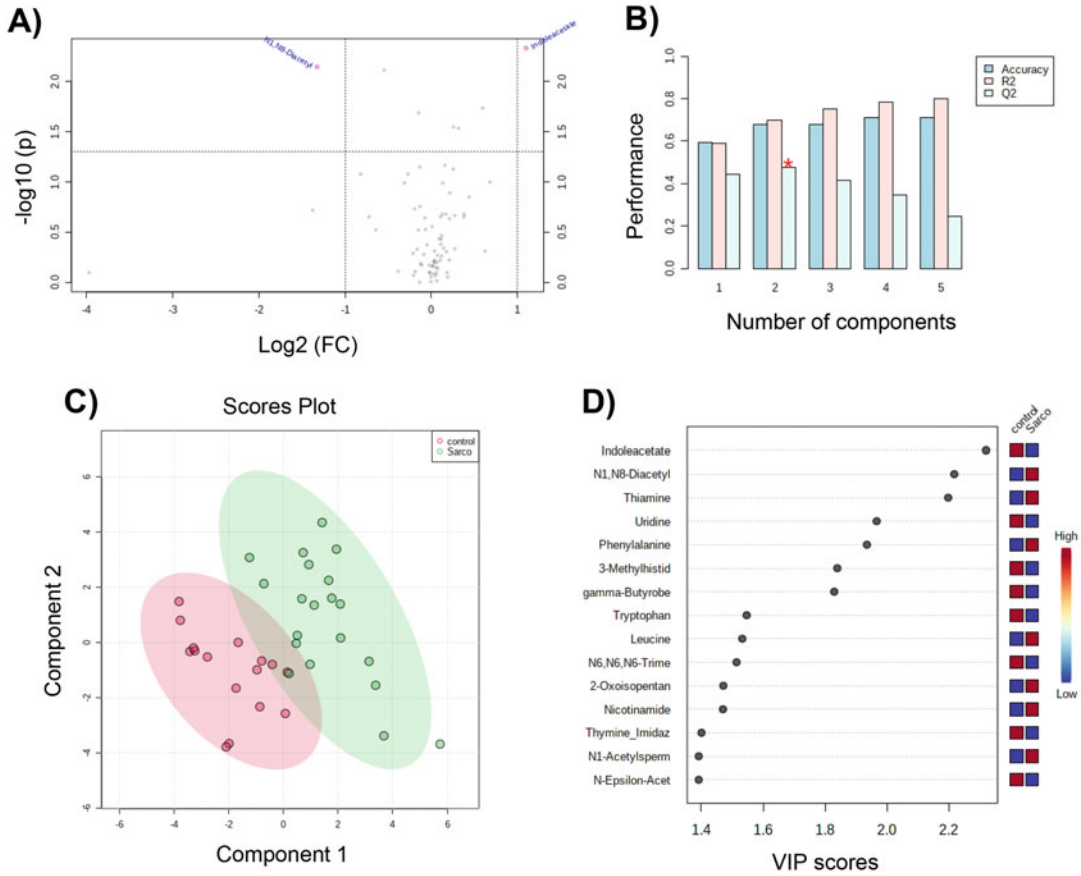

Fig. 8 Discrimination analyses. (**a**) Volcano plots and (**b**–**d**) partial least squares-discriminant analysis (PLS-DA). (**a**) In volcano plots, the data showing a higher absolute value in the x- and y-axes indicate higher fold change and more significant difference, respectively, between the two groups. (**b**) The accuracy of the prediction of the PLS-DA model. R^2 and Q^2 are calculated using all data and cross-validation. The number of components should be selected by the smallest difference between R^2 and Q^2 or the highest Q^2 value. At least, $Q^2 > 0.45$ is empirically required when PLS-DA is used for subsequent analyses. (**c**) Score plots of the PLS-DA, which indicate similarities in the samples based on the PLS-DA model. (**d**)Variable importance for prediction (VIP) score of the PLS-DA. The metabolites with high VIP scores significantly contribute to the separation

2. The TIC should be evaluated for the quality of the MS data. Figure 11 depicts examples of standard mixture TICs conducted at various intervals from 1 to 11 months after the last MS maintenance. The TICs performed later exhibit a higher count (more background noise) than those performed earlier. In the case of high background noise, the S/N ratio of the various peaks should be calculated, maintenance on the instrument should be performed, and the samples should be reanalyzed.

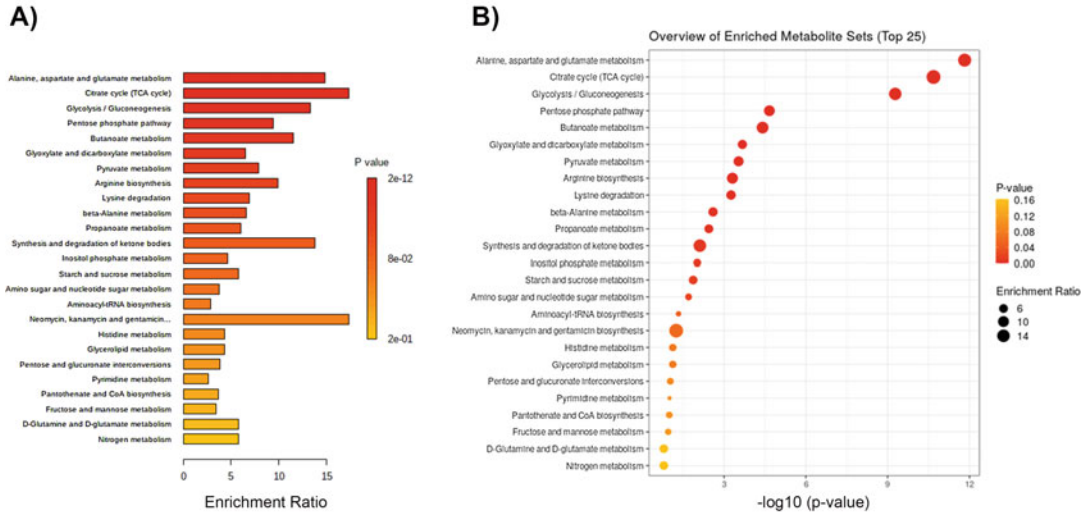

Fig. 9 Pathway enrichment analysis. (**a**) The rank of the enrichment pathways and (**b**) the rank of the enrichment metabolites. The pathways with a larger number of enriched metabolites are ranked higher

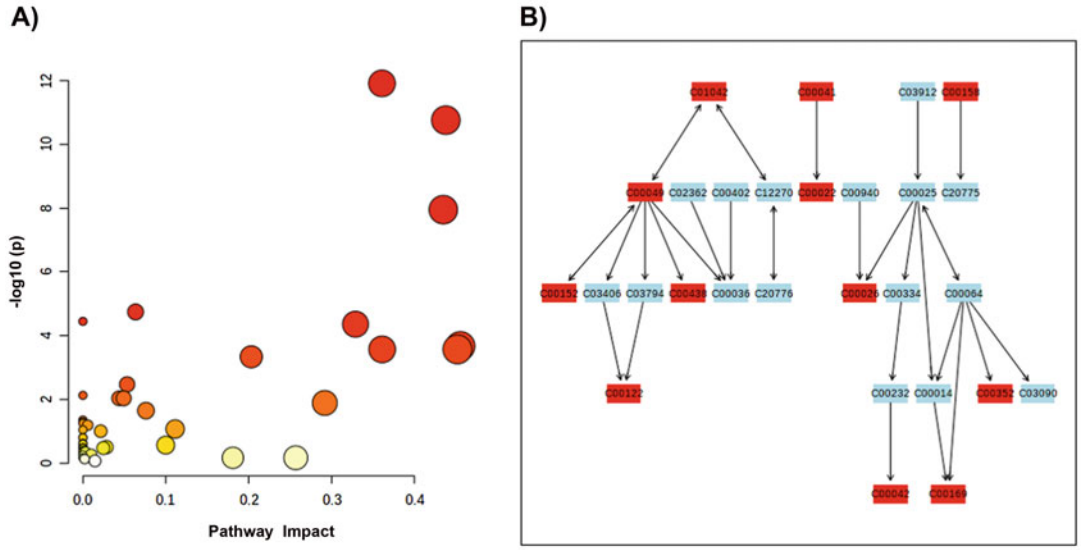

Fig. 10 Pathway analysis. (**a**) Pathway impact (x-axis) and significant differences (y-axis). (**b**) A pathway. The pathways including metabolites with a significant difference are located at higher y-values. When these metabolites are located at the topologically center position of the pathway, the pathway is located at higher x-values

3. The overall TIC is not horizontally flat compared to the ones in the previous batches (Fig. 12). This may be due to the imprecise conditions of nitrogen gas; for example, the temperature of the gas provider may be inaccurate. In such a case, conduct maintenance on the gas provider and restart the gas provider to set its temperature at the target value.

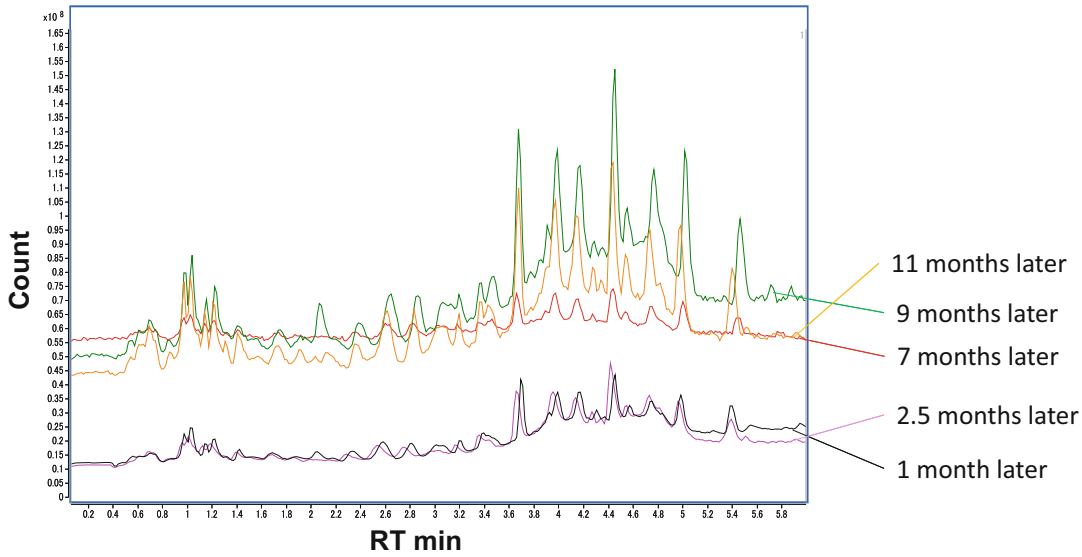

Fig. 11 The difference of the overlaid TIC after various periods (1 month to 11 months) after the maintenance of the LC/MS system

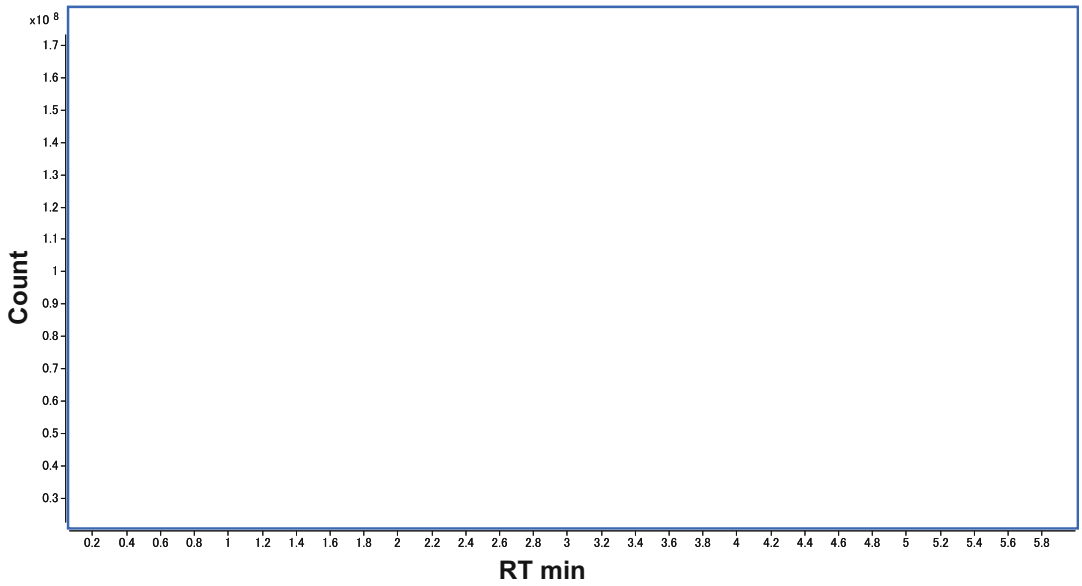

Fig. 12 Examples of unstable TIC. The green TIC is observed in usual cases. The purpose TIC is observed at the unexpectedly high temperature of nitrogen gas

4. The RT values of the peaks of the IS were compared with those of the previous batches (Fig. 13). A delay in the RT (shift right in the EIC) implies that the filter attached before the LC columns is contaminated. Hence, it is replaced with a new filter and the sample is reanalyzed.

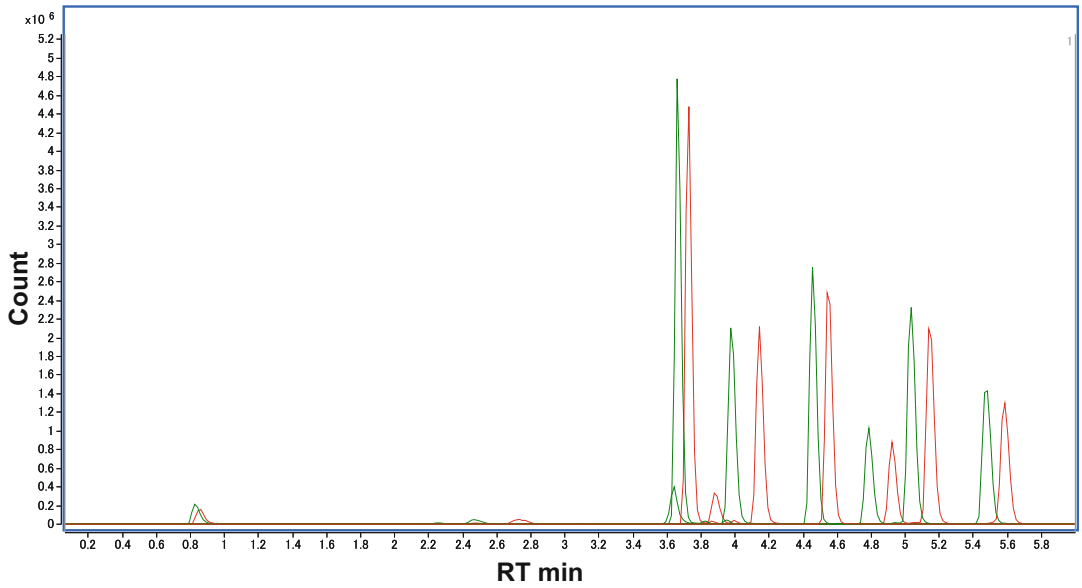

Fig. 13 Overlaid EIC of internal standard peaks. The green EIC is observed in usual cases. The red EIC is shifted to the right because of the contaminated filter before the LC columns

5. When the data matrix (samples × metabolite concentrations) is obtained, a quality assessment of the quantified values should be conducted before downstream statistical analyses. A simple method is to compare the quantification levels of the standard mixture and QC samples with those of the previously measured. Two types of unexpected biases can be considered: (1) internal batch bias and (2) inter-batch bias. To eliminate bias in each batch, QC samples were measured among the sequences of several samples, and the sensitivity of MS was estimated by the peak intensities of the QC samples before and after the sequence using linear regressions [20]. Correlation analyses or more sophisticated normalization using QC samples have also been developed for large-scale, including multi-batch analyses [21, 22].
6. Unexpected bias in the overall metabolite concentration among the samples should be eliminated by sample normalization. The data are distributed in a Poisson distribution, and a Gaussian distribution is preferable for subsequent analysis; thus, log transformation should be selected for data transformation. The metabolite concentration is transferred to the Z-score by auto-scaling of the data scaling.
7. Pathway analysis ranks the pathways based on pathway impact and significant differences. Pathways showing higher values in both the x- and y-axes should be the focus of subsequent analyses.

Acknowledgments

This research was funded by grants from JSPS KAKENHI (grant number 20B205) and JST OPERA (grant number JPMJOP1842).

References

1. Amara CS, Vantaku V, Lotan Y, Putluri N (2019) Recent advances in the metabolomic study of bladder cancer. Expert Rev Proteomics 16(4):315–324. https://doi.org/10.1080/14789450.2019.1583105
2. Gardner A, Parkes HG, Carpenter GH, So PW (2018) Developing and standardizing a protocol for quantitative proton nuclear magnetic resonance ((1)H NMR) spectroscopy of saliva. J Proteome Res 17(4):1521–1531. https://doi.org/10.1021/acs.jproteome.7b00847
3. Lubes G, Goodarzi M (2018) GC–MS based metabolomics used for the identification of cancer volatile organic compounds as biomarkers. J Pharm Biomed Anal 147:313–322
4. Zhang W, Ramautar R (2021) CE-MS for metabolomics: developments and applications in the period 2018–2020. Electrophoresis 42(4):381–401. https://doi.org/10.1002/elps.202000203
5. Roca M, Alcoriza MI, Garcia-Cañaveras JC, Lahoz A (2021) Reviewing the metabolome coverage provided by LC-MS: focus on sample preparation and chromatography-a tutorial. Anal Chim Acta 1147:38–55. https://doi.org/10.1016/j.aca.2020.12.025
6. Sugimoto M, Kawakami M, Robert M, Soga T, Tomita M (2012) Bioinformatics tools for mass spectroscopy-based metabolomic data processing and analysis. Curr Bioinforma 7(1):96–108. https://doi.org/10.2174/157489312799304431
7. Baima G, Iaderosa G, Citterio F, Grossi S, Romano F, Berta GN, Buduneli N, Aimetti M (2021) Salivary metabolomics for the diagnosis of periodontal diseases: a systematic review with methodological quality assessment. Metab Off J Metab Soc 17(1):1. https://doi.org/10.1007/s11306-020-01754-3
8. Ishikawa S, Sugimoto M, Kitabatake K, Tu M, Sugano A, Yamamori I, Iba A, Yusa K, Kaneko M, Ota S, Hiwatari K, Enomoto A, Masaru T, Iino M (2017) Effect of timing of collection of salivary metabolomic biomarkers on oral cancer detection. Amino Acids 49(4): 761–770. https://doi.org/10.1007/s00726-017-2378-5
9. Nakajima T, Katsumata K, Kuwabara H, Soya R, Enomoto M, Ishizaki T, Tsuchida A, Mori M, Hiwatari K, Soga T, Tomita M, Sugimoto M (2018) Urinary polyamine biomarker panels with machine-learning differentiated colorectal cancers, benign disease, and healthy controls. Int J Mol Sci 19(3). https://doi.org/10.3390/ijms19030756
10. Hirayama A, Sugimoto M, Suzuki A, Hatakeyama Y, Enomoto A, Harada S, Soga T, Tomita M, Takebayashi T (2015) Effects of processing and storage conditions on charged metabolomic profiles in blood. Electrophoresis 36(18):2148–2155. https://doi.org/10.1002/elps.201400600
11. Sugimoto M (2020) Salivary metabolomics for cancer detection. Expert Rev Proteomics 17(9):639–648. https://doi.org/10.1080/14789450.2020.1846524
12. Liebal UW, Phan ANT, Sudhakar M, Raman K, Blank LM (2020) Machine learning applications for mass spectrometry-based metabolomics. Meta 10(6). https://doi.org/10.3390/metabo10060243
13. Patti GJ, Yanes O, Siuzdak G (2012) Innovation: metabolomics: the apogee of the omics trilogy. Nat Rev Mol Cell Biol 13(4):263–269. https://doi.org/10.1038/nrm3314
14. Dunn WB, Wilson ID, Nicholls AW, Broadhurst D (2012) The importance of experimental design and QC samples in large-scale and MS-driven untargeted metabolomic studies of humans. Bioanalysis 4(18):2249–2264
15. Shimizu H, Usui Y, Asakage M, Nezu N, Wakita R, Tsubota K, Sugimoto M, Goto H (2020) Serum metabolomic profiling of patients with non-infectious uveitis. J Clin Med 9(12). https://doi.org/10.3390/jcm9123955
16. Nam SL, Mata AP, Dias RP, Harynuk JJ (2020) Towards standardization of data normalization strategies to improve urinary metabolomics studies by GC×GC-TOFMS. Meta 10(9). https://doi.org/10.3390/metabo10090376
17. Misra BB (2021) New software tools, databases, and resources in metabolomics: updates from 2020. Metab Off J Metab Soc 17(5):49. https://doi.org/10.1007/s11306-021-01796-1
18. Ren S, Hinzman AA, Kang EL, Szczesniak RD, Lu LJ (2015) Computational and statistical

analysis of metabolomics data. Metab Off J Metab Soc 11(6):1492–1513

19. Pang Z, Chong J, Zhou G, de Lima Morais DA, Chang L, Barrette M, Gauthier C, Jacques P, Li S, Xia J (2021) MetaboAnalyst 5.0: narrowing the gap between raw spectra and functional insights. Nucleic Acids Res. https://doi.org/10.1093/nar/gkab382
20. Saigusa D, Okamura Y, Motoike IN, Katoh Y, Kurosawa Y, Saijyo R, Koshiba S, Yasuda J, Motohashi H, Sugawara J, Tanabe O, Kinoshita K, Yamamoto M (2016) Establishment of protocols for global metabolomics by LC-MS for biomarker discovery. PLoS One 11(8):e0160555. https://doi.org/10.1371/journal.pone.0160555
21. Saito R, Sugimoto M, Hirayama A, Soga T, Tomita M, Takebayashi T (2021) Quality assessment of untargeted analytical data in a large-scale Metabolomic study. J Clin Med 10(9). https://doi.org/10.3390/jcm10091826
22. Yamamoto H, Suzuki M, Matsuta R, Sasaki K, Kang M-I, Kami K, Tatara Y, Itoh K, Nakaji S (2021) Capillary electrophoresis mass spectrometry-based metabolomics of plasma samples from healthy subjects in a cross-sectional Japanese population study. Meta 11(5):314

Chapter 22

Instrument-Agnostizing Methodology for Liquid Chromatography–Mass Spectrometry Systems

Rosalía López-Ruíz, Sandra Martín-Torres, Ana M. Jiménez-Carvelo, Roberto Romero-González, and Luis Cuadros-Rodríguez

Abstract

Mass spectrometry is a powerful analytical technique used to identify unknown compounds, to quantify known compounds, and to elucidate the structure and chemical properties of molecules. Nevertheless, the transfer of data from one instrument to another is one of the main problems, and obtaining the same or similar information from an analogous instrument but from a different manufacturer or even with the same instrument after carrying out the analyses in different times spacing is not possible. Hence, a general methodology to provide a chromatographic signal (or chromatogram) independent of the instrument is needed. In this sense, this book chapter describes the standardization procedure of chromatographic signals obtained from mass spectrometry platforms to obtain instrument-agnostic chromatographic signals for the determination of standard retention scores. This parameter may be used for the quantification of compounds when different mass spectrometry platforms coupled to ultrahigh-performance liquid chromatography are employed.

Key words Instrument-agnostizing methodology, Standardized chromatogram, Instrument-independent analytical signals, Mass spectrometry

1 Introduction

Nowadays, the use of pesticides to reduce crop losses due to pests and keep food quality is still increasing. Their use has increased worldwide because of their rapid and efficient action despite the fact that organic farming is also expanding. Nevertheless, the presence of pesticide residues in fruits and vegetables, among other foods, has become a major concern for consumer's health, especially in developing countries [1].

To solve this problem, liquid chromatography coupled to mass spectrometry is the technique widely used for the determination of pesticide residues in food [2]. In this field, laboratory routine analysis must have the capacity to analyze a wide range of pesticides

Raúl González-Domínguez (ed.), *Mass Spectrometry for Metabolomics*,
Methods in Molecular Biology, vol. 2571, https://doi.org/10.1007/978-1-0716-2699-3_22,
© The Author(s), under exclusive license to Springer Science+Business Media, LLC, part of Springer Nature 2023

(more than 500) to ensure food safety and quality. This becomes a great concern due to the necessity to obtain analytical standards of each pesticide to detect them by chromatographic techniques coupled to mass spectrometry. In addition, parameters such as retention time or peak area can be difficult to reproduce in different equipment or laboratories [3], and therefore, it is difficult to compare these relevant parameters among different analytical platforms used by the same or different laboratories. Considerable efforts have been made to develop more reproducible chromatographic systems, which have not always succeeded.

To date, dozens of correspondence algorithms have been proposed to deal with retention time alignment and signal normalization of LC fingerprints, especially when LC is coupled to MS systems [4]. Nevertheless, most of them are analyte dependent, based on inner reference variables and none of these approaches can be considered as suitable procedures to obtain reproducible absolute responses to solve the universal standardization of data.

In this field, instrument-agnostizing methodology takes part as the ability to provide standardized and comparable chromatographic fingerprints regardless of the instrument, the particular instrumental conditions or the time of analysis, in order to ensure better reproducibility. This new methodology is described based on two main stages: a single stage for setting up standard retention scores, which is only applied once and referred to an external reference; and the stage of chromatogram agnostizing in which specific normalizations, of both intensities and retention scores, are carried out using the previously established scores. A general description of the whole procedure is given in recently published articles [5, 6].

2 Materials

2.1 External Standard Solutions (ES)

1. Purchase analytical standards of solid contaminants, parabens, and drugs. Prepare several stock solutions:
   1.1. Standard solutions of each analyte (Table 1) (10 mL) at 1000 mg/L in MeOH.
   1.2. Standard mix solution (Table 1) (10 mL) at 10 mg/L in MeOH.
   1.3. Standard mix solution (Table 1) (10 mL) at 1 mg/L in MeOH (working solution).
2. Store solutions 1, 2, and 3 in the freezer at −21 °C.

2.2 Internal Standard Solutions (IS)

1. Purchase solid analytical standards and prepare several stock solutions:
   1.1. Standard of caffeine and triphenylphosphate (TPP) (10 mL) at 1000 mg/L in MeOH.

Table 1
Chemical compounds chosen to constitute the standard mixture

Compound name	CAS	Molecular formula	Monitored ion (*m/z*)	SRS
Trimethoprim	738–70-5	$C_{14}H_{18}N_4O_3$	291.14505	1.000
3,5-Dihydroxybenzoic	99–10-5	$C_7H_6O_4$	153.01933	1.114
Syringic acid	530–57-4	$C_9H_{10}O_5$	197.04555	1.805
Sulfadimethoxine	122–11-2	$C_{12}H_{14}N_4O_4S$	311.08081	2.750
Methylparaben	99–76-3	C_7H_6O	151.04007	3.750
Benzoic acid	65–85-0	$C_7H_6O_2$	121.02950	4.140
Ethylparaben	120–47-8	$C_9H_{10}O_3$	165.05572	6.417
Propiophenone	202–257-6	$C_9H_{10}O$	135.08044	8.034
Benzaldehyde	100–52-7	C_7H_6O	107.04914	8.725
Butanophenone (*Butyrophenone*)	207–799-7	$C_{10}H_{12}O$	149.09609	11.113
Benzophenone	119–61-9	$C_{13}H_{10}O$	183.08044	12.234
Pentanophenone (*Valerophenone*)	213–767-3	$C_{11}H_{14}O$	163.11174	13.415
Benzyl benzoate	204–402-9	$C_{14}H_{12}O_2$	213.09101	13.668
Hexanophenone	213–394-6	$C_{12}H_{16}O$	177.12739	15.101
Heptanophenone	216–802-0	$C_{13}H_{18}O$	191.14113	16.391
Octanophenone (*Caprylophenone*)	216–817-2	$C_{14}H_{20}O$	205.15869	17.429
Nonanophenone	227–861-7	$C_{15}H_{22}O$	219.17434	18.294
Decanophenone	227–946-9	$C_{16}H_{24}O$	233.18999	19.018

1.2. Standard mix solution of caffeine and TPP (10 mL) at 10 mg/L in MeOH.

2. Store solutions 1 and 2 in the freezer at −21 °C.

2.3 Mass Calibration

Accurate the mass calibration of the LC-Q-Orbitrap and LC-Orbitrap mass analyzers using a mixture of acetic acid, caffeine, Met-Arg-Phe-Ala-acetate salt and Ultramark 1621, and a mixture of acetic acid, sodium dodecyl sulfate, taurocholic acid sodium salt hydrate, and Ultramark 1621 (fluorinated phosphazines). Calibration mix solutions were provided by the commercial brand.

2.4 Separation Systems

1. Vanquish flex Quaternary LC pump (Thermo Scientific Transcend™, Thermo Fisher Scientific, San Jose, CA, USA).
2. Thermo Fisher Scientific Transcend 600 LC (Thermo Scientific Transcend™, Thermo Fisher Scientific, San Jose, CA, USA).
3. Column: Zorbax Eclipse Plus C_{18} column (100 mm × 2.1 mm, 1.8 μm particle size).

2.5 Mass Spectrometry Systems

2.5.1 MS System #1

1. Q-Exactive Orbitrap analyzer (Thermo Fisher Scientific, Bremen, Germany) using an electrospray interface (ESI) (HESI-II, Thermo Fisher Scientific, San Jose, CA, USA) in positive and negative modes.

2.5.2 MS System #2

1. Exactive Orbitrap analyzer (Thermo Fisher Scientific, Bremen, Germany) using an electrospray interface (ESI) (HESI-II, Thermo Fisher Scientific, San Jose, CA, USA) in positive and negative modes.

3 Methods

3.1 Sample Treatment

1. Select cucumber as a representative matrix of vegetables with high water content.
2. Use extraction method well-known QuEChERS procedure [7], commonly used for pesticide extraction.
3. Crush and homogenize 1 kg of cucumber.
4. Place 10 g of sample in a 50-mL Falcon® tube.
5. Add 10 mL of ACN. Shake for 1 min in a vortex.
6. Add 1 g of NaCl and 4 g of $MgSO_4$. Shake 1 min in a vortex.
7. Centrifuge the sample for 10 min at 5000 rpm (RCF = 4480 *g*).
8. Collect the supernatant to prepare the samples for the instrument-agnostizing procedure.
9. Select two internal standards (IS) and add them at a well-defined and constant amount to each test solution from samples. Ideally, the concentration of the IS in the test solution may provide the highest signal intensity (peak height) in the sample chromatogram, and they should elute into each of the two halves of the sample chromatogram.

3.2 LC Conditions

1. Mobile phase A: Water solution of 0.1% formic acid. Mobile phase B: MeOH.
2. Elution mode gradient: 0–1 min, 50% of A, decreased to 0% of A in 10 min, kept 2 min and returned to 50% of A in 0.5 min, finally from 13.5 to 20 min kept constant to equilibrate the analytical column.
3. Column temperature set at 30 °C.
4. Injection volume of 10 μL.
5. Flow rate of 0.2 mL/min.
6. The total running time is 20 min.

3.3 MS Parameters

3.3.1 MS System #1

1. ESI parameters: spray voltage, 4 kV; sheath gas (N_2, >95%), 35 (arbitrary units); auxiliary gas (N_2, >95%), 10 (arbitrary units); heater temperature, 300 °C; capillary temperature, 300 °C and S-lens RF level, 50 (arbitrary units).

3.3.2 MS System #2

1. ESI parameters: spray voltage, 4 kV; sheath gas (N_2, >95%), 35 (arbitrary units); auxiliary gas (N_2, >95%), 10 (arbitrary units); skimmer voltage, 18 V; capillary voltage, 35 V; tube lens voltage, 95 V; heater temperature, 305 °C; capillary temperature, 300 °C.

3.4 Step 1: Setting up Standard Retention Scores (SRS)

1. Select a reference chromatographic method based on simple linear gradient (or isocratic) mode and represent all those methods with the same mobile phases and the same solvent composition mixture at the beginning and at the end of the chromatographic run (*see* **Note 1**).
2. Establish an invariant reference chemical system: select a proper standard mixture of suitable chemical compounds. Ideally, the compounds should cover all the chromatographic run time; show a regular elution profile under the selected experiment conditions; be well characterized and be pure enough; commercially available and (if possible) inexpensive; and have a chemical behavior similar to endogenous components of interest in the sample.
3. Analyze the standard mix at least 10 times under nearly reproducible instrumental conditions to obtain representative retention time values.
4. Apply a robust statistic based on the median to the retention times of each i-th chemical compound along analyses (median RT_i) to remove possible outliers (*see* **Note 2**).
5. Sort out chemical constituents according to their elution order, represented by integers from 1 to *n*. A nearly linear trend must be found between RT values and elution order.
6. Calculate standard retention scores (SRSs), understood as empirical values related to retention time and elution order:

$$SRS_i = \frac{\Delta(\text{medianRT})_{i,i-1}}{\text{median}\left(\Delta(\text{medianRT})_{i,i-1}\right)} + SRS_{i-1}$$

   where $\Delta(\text{medianRT})_{i,i\text{-}1}$ denotes absolute differences on retention times between adjacent chemical constituents and (median ($\Delta(\text{medianRT})_{i,i\text{-}1}$) overall median value from them.
7. Assign an SRS to each standard: an SRS of 1 is always assigned to the first eluted compound of the standard mix (*see* **Note 3**).

Note SRSs are invariant as long as the mobile phase does not change. Thus, they remain as a fixed reference system for all analytical methods that are the same or similar to the one used as the reference chromatographic method. That is, step 1 is carried out only once before the implementation of the analytical method.

3.5 Step 2: Liquid Chromatography–Mass Spectrometry Agnostizing

3.5.1 Chromatographic Analyses of Samples

1. Design the chromatographic batch: each day, it must include the analysis of the standard mix at the beginning and at the end of the batch.
2. Analyze the external standard mix solution applying the reference method.
3. Analyze the sample solutions, which include the internal standards, applying the specific method (*see* **Notes 4** and **5**).
4. Export chromatographic signals to working format and obtain an intensity data vector for each chromatogram.

3.5.2 Normalization of Signal Intensities

1. Clean up data vector by applying a suitable preprocessing step: truncation and cutting data, de-noising and smoothing for removing/minimizing noise; baseline correction for drifting or rungs along the chromatogram; or resampling to fix the number of elements in the data vector to a given number.
2. Calculate a relative intensity for each i-th element of the intensity data vector from the samples ($\mathrm{INT_i^{rel}}$)

$$\mathrm{INT_i^{rel}} = \frac{\mathrm{INT_i}}{\mathrm{INT_i^{ref}}}$$

where $\mathrm{INT_i}$ denotes the original intensity value; $\mathrm{INT_{i\ ref}}$ is the reference intensity value of the internal standard.

3. Pooled intensity from both internal standards is calculated as

$$\begin{aligned}\mathrm{INT_i^{ref}} &= \mathrm{INT_i^{pooled}}(\mathrm{IS1}, \mathrm{IS2}) \\ &= \frac{\left(\mathrm{INT_{IS_1}} \times |\Delta t|_{\mathrm{IS_{2,i}}}\right) + \left(\mathrm{INT_{IS2}} \times |\Delta t|_{\mathrm{IS_{1,i}}}\right)}{\left|\Delta t_{\mathrm{IS_2,IS_1}}\right|}\end{aligned}$$

where $\mathrm{INT_{IS_1}}$ and $\mathrm{INT_{IS_2}}$ are the intensity values (peak heights) of both internal standards $\mathrm{IS_1}$ and $\mathrm{IS_2}$, and $|\Delta t|_{\mathrm{IS_{2,i}}}$, $|\Delta t|_{\mathrm{IS_{1,i}}}$, and $\left|\Delta t_{\mathrm{IS_2,IS_1}}\right|$ are the absolute values of the differences between the retention times for $\mathrm{IS_1}$ and the i-th element, $\mathrm{IS_2}$ and the i-th element, and $\mathrm{IS_1}$ and $\mathrm{IS_2}$, respectively.

3.5.3 Normalization of Retention Scores: Transfer to SRS Domain

1. Record and list the experimental retention times for each chemical standard of the two analyses (beginning and ending times) each day.
2. Calculate the average between the RTs obtained in the initial and final analyses a day.

3. Plot the average RTs of all the chemical compounds making up the standard mix against predetermined invariant SRSs.
4. Establish a piecewise function that is transferred by linear spline interpolation between the pairs to each section of the function.

$$= FORECAST\left(t_i; INDEX\left(x_{know}; MATCH\left(t_i; y_{know}; 1\right)\right) : INDEX\left(x_{know}; MATCH\left(t_i; y_{know}; 1\right) + 1\right); INDEX\left(y_{know}; MACTH\left(t_i; y_{know}; 1\right)\right); INDEX\left(y_{know}; MATCH\left(t_i; y_{know}; 1\right) + 1\right)\right)$$

Indeed, the transferring for retention time normalization is carried out using three specific Excel functions:

- *FORECAST:* It predicts a value within a data set previously available using the whole data set for mathematical computing. Although this function performs linear interpolations, it is also applicable when the data sets show nonlinear trends if a linear behavior is assumed in each interpolation range.
- *MACHT:* It searches for a given element within a range of cells and returns the relative position of that element.
- *INDEX:* It returns the value of a specific position in the cell range.

Further information can be found in [6].

5. Assign and record an SRS vector to each sample.
6. Apply a resampling algorithm to fix the number of elements, which constitutes each instrument-agnostic data vector (*see* **Notes 6** and **7**).

3.6 Instrument-Agnostizing of Samples Chromatograms: General Overview

1. Set up standard retention scores (SRS).
2. LC-MS signal acquisition of samples.
3. Export signal to working format.
4. Carry out preprocessing operations.
5. Normalize signal intensities by scaling the chromatographic data vector using internal standard intensities.
6. Normalize retention scores by transferring the retention time scale to SRS domain.
7. Resampling.
8. Verify results (*see* **Notes 8–19**).

4 Notes

1. Select the analytical standard mixture (Table 1) according to the criterion explained in Subheading 3.4.
2. Obtain representative retention times after the analysis of the standard mix 15 times in the Q-Exactive mass spectrometer (MS system #1).

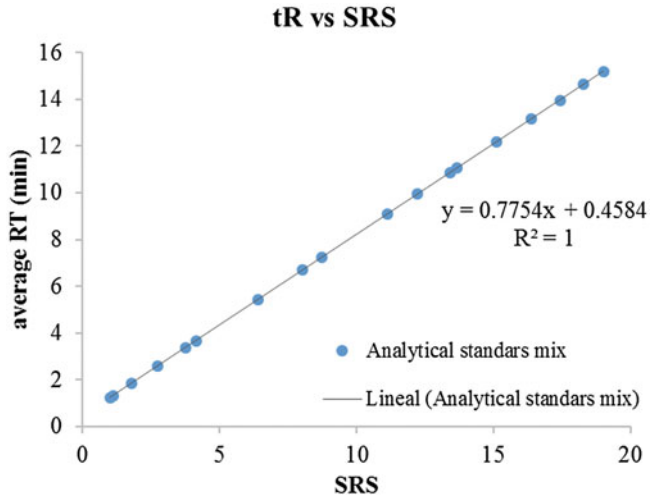

Fig. 1 Average RT (min) vs SRS values for the analytical standard mixture

Table 2
Compounds selected to test the instrument-agnostizing methodology

Compound name	Molecular formula	Monitored ion (*m/z*)
Benalaxil	$C_{20}H_{23}NO_3$	326.17507
Desmedipham	$C_{16}H_{16}N_2O_4$	301.11828
Epoxiconazole	$C_{17}H_{13}ClFN_3O$	330.08039
Fenbuconazole	$C_{19}H_{17}ClN_4$	337.12145
Fluquinconazole	$C_{16}H_8Cl_2FN_5O$	376.01627
Imazapyr	$C_{13}H_{15}N_3O_3$	262.11862
Metconazole	$C_{17}H_{22}ClN_3O$	320.15241
Penconazole	$C_{13}H_{15}Cl_2N_3$	284.07157
Pyriproxyfen	$C_{20}H_{19}NO_3$	322.14377
Triticonazole	$C_{17}H_{20}ClN_3O$	318.13676

3. Calculate SRS according to the procedure described in **steps 5–7** (Subheading 3.4) to obtain a linear equation (Table 1 and Figure 1).
4. Pesticides were chosen (see Table 2) as an example of target analytes in order to verify the advantages of the instrument-agnostizing methodology.
5. Analyze the mix of pesticides and IS in the Q-Exactive mass spectrometer (MS system #1), selecting retention time and peak area at 0.06 and 0.15 mg/L (n = 20) (Table 3).

Table 3
Mean of all measurements for RT and peak area in Q-Exactive (MS system *#1*)

Compound name	RT	Peak area	
		0.06 mg/L	0.15 mg/L
Benalaxil	13.31	599,101,141	1,433,235,309
Desmedipham	8.94	165,665,040	690,995,903
Epoxiconazole	11.29	519,462,959	1,154,567,356
Fenbuconazole	11.44	388,352,804	870,524,681
Fluquinconazole	10.96	111,765,021	214,692,592
Imazapyr	2.19	577,931,822	218,545,375
Metconazole	12.44	773,035,186	177,327,019
Penconazole	11.96	662,118,942	669,196,703
Pyriproxyfen	13.83	1,465,751,430	396,817,976
Triticonazole	11.13	722,781,016	404,038,253

Table 4
SRS employed for the instrument-agnostizing methodology in both MS systems

Compound name	SRS from MS system #1	SRS from MS system #2
Benalaxil	16.57	16.31
Desmedipham	10.93	10.76
Epoxiconazole	13.97	13.46
Fenbuconazole	14.16	13.49
Fluquinconazole	13.55	13.20
Imazapyr	2.24	2.29
Metconazole	15.46	15.19
Penconazole	14.83	13.60
Pyriproxyfen	17.25	16.99
Triticonazole	13.76	13.40

6. Once the RT vector of each sample is collected, an SRS vector should be calculated for each pesticide by applying the methodology described in Subheading 3.6 (Table 4).
7. Follow the same procedure previously described in the Exactive mass spectrometer (MS system #2), establishing SRS (Table 4).

Table 5
SRS deviation between MS system #1 and MS system #2 (Q-Exactive and Exactive mass spectrometers)

Compound name	SRS deviation
Benalaxil	−1.56
Desmedipham	−1.61
Epoxiconazole	−3.70
Fenbuconazole	−4.70
Fluquinconazole	−2.57
Imazapyr	2.48
Metconazole	−1.79
Penconazole	**−8.33**
Pyriproxyfen	−1.45
Triticonazole	−2.67

8. Once pesticides are modeled according to the instrument-agnostizing methodology, information of both mass spectrometers is compared to determine the SRS deviation and the possibilities of the instrument-agnostizing (Table 5).
9. Criteria used to classify pesticides are the following:

   SRS | > 5% ⇒ not valid for instrument-agnostizing methodology

   SRS | < 5% ⇒ valid for instrument-agnostizing methodology
10. For pesticide studies and according to the criteria exposed, it is possible to detect and determine five out six between the two mass spectrometers (compounds in bold, Table 5).
11. Quantification is also studied, using the caffeine and TPP internal standards to normalize the peak area in both mass spectrometers using the procedure described in Subheading 3.6.
12. Evaluate two concentrations to determine the possibilities of quantification (0.06 mg/L and 0.15 mg/L) (Table 6).
13. To test the normalized values, the quantification errors are calculated using the Q-Exactive mass spectrometer as reference (MS system #1), and testing concentrations in the Exactive mass spectrometer (MS system #2) (Table 7).

Table 6
Peak area normalized

	Peak area normalized (PAN) in MS system #1		Peak area normalized (PAN) in MS system #2	
Compound name	**0.06 mg/L**	**0.15 mg/L**	**0.06 mg/L**	**0.15 mg/L**
Benalaxil	0.125	0.316	0.109	0.255
Desmedipham	0.048	0.112	0.040	0.117
Epoxiconazole	0.134	0.314	0.117	0.281
Fenbuconazole	0.100	0.235	0.095	0.222
Fluquinconazole	0.029	0.059	0.026	0.060
Imazapyr	0.28	0.577	0.210	0.480
Metconazole	0.044	0.415	0.152	0.365
Penconazole	0.165	0.374	0.137	0.315
Pyriproxyfen	0.309	0.640	0.248	0.599
Triticonazole	0.176	0.110	0.133	0.309

Table 7
Concentration estimated and errors calculated

	Ce		%Error	
Compound name	**0.06 mg/L**	**0.15 mg/L**	**0.06 mg/L**	**0.15 mg/L**
Benalaxil	0.053	0.121	−12.58	−19.48
Desmedipham	0.049	0.157	−17.45	4.83
Epoxiconazole	0.052	0.134	−12.71	−10.40
Fenbuconazole	0.052	0.142	−4.53	−5.53
Fluquinconazole	0.054	0.153	−9.22	2.15
Imazapyr	0.049	0.157	−17.75	4.83
Metconazole	0.206	0.132	**243.57**	−11.97
Penconazole	0.050	0.127	−16.71	−15.50
Pyriproxyfen	0.048	0.140	−19.81	−6.68
Triticonazole	0.046	0.418	−24.23	**178.37**

14. Calculate the concentrations estimated (C_e) from Exactive using the following formula:

$$C_e = \frac{\text{PAN Exactive} \times C_T}{\text{PAN QExactive}}$$

where PAN is peak area normalized and C_T is theoretical concentration.

15. Estimate errors, using

$$\%\text{Error} = \frac{(C_e - C_T) \times 100}{C_T}$$

16. Criteria used to quantify pesticides are the following:

    Error | > 40% ⇒ no quantifiable

    Error | < 40% ⇒ quantifiable

17. According to the criteria, only two pesticides are not quantifiable at lower concentrations (metconazole) and at higher concentrations (triticonazole) (compounds in bold, Table 7).
18. Finally, combining concepts SRS and quantification, seven compounds of ten are valid for the instrumental-agnostizing methodology.
19. With that seven, it is possible for the detection and quantification only using the instrument-agnostizing analytical standards (Table 1) and the IS. Analytical standards of the pesticides are not required.

Acknowledgments

RLR acknowledges "Plan Propio of Investigation" of the University of Almería, co-funded by CAJAMAR and the Operational Program Funds Europeans of Regional Development of Andalusia (2014–2020) (FEDER), for financial support.
SMR acknowledges the PhD grant supported by the Ministry of Universities for the Training of University Teachers (FPU19/02078).

References

1. López-Ruiz R, Romero-González R, Serra B, Garrido-Frenich A (2019) Dissipation kinetic studies of fenamidone and propamocarb in vegetables under greenhouse conditions using liquid and gas chromatography coupled to high-resolution mass spectrometry. Chemosphere 226:36–46. https://doi.org/10.1016/j.chemosphere.2019.03.118
2. López-Ruiz R, Romero-González R, Garrido-Frenich A (2019) Ultrahigh-pressure liquid chromatography-mass spectrometry: an overview of the last decade. Trends Anal Chem 118:170–181. https://doi.org/10.1016/j.trac.2019.05.044
3. Brunius C, Shi L, Landberg R (2016) Large-scale untargeted LC-MS metabolomics data correction using between-batch feature alignment and cluster-based within-batch signal intensity drift correction. Metabolomics 12(173):1–13. https://doi.org/10.1007/s11306-016-1124-4

4. Smith R, Ventura D, Prince JT (2013) LC-MS alignment in theory and practice: a comprehensive algorithmic rewiew. Brief Bioinform 16: 104–107. https://doi.org/10.1093/bib/bbt080
5. Cuadros-Rodríguez L, Ortega-Gavilán F, Martín-Torres S, Medina-rodríguez S, Jiménez-Carvelo A, González-Casado A, Bagur-González MG (2021) Standardization of instrumental fingerprints 1 – part I: towards an instrumental agnosticism in gas chromatography. J Chromatogr A 1641(461983):1–11. https://doi.org/10.1016/j.chroma.2021.461983
6. Cuadros-Rodríguez L, Martín-Torres S, Ortega-Gavilán F, Jiménez-Carvelo A, López-Ruiz R, Garrido-Frenich A, Bagur-González MG, González-Casado A (2021) Standardization of chromatographic signals –part II: expanding instrument-agnostic fingerprints to reverse phase liquid chromatography. J Chromatogr A 1641(461973):1–14. https://doi.org/10.1016/j.chroma.2021.461973
7. Lehotay SJ, Son KA, Kwon H, Koesukwiwat U, Fu W, Mastovska K, Hoh E, Leepipatpiboon N (2010) Comparison of QuEChERS sample preparation methods for the analysis of pesticide residues in fruits and vegetables. J Chromatogr A 1217:2548–2560. https://doi.org/10.1016/j.chroma.2010.01.044

Index

A

Antiproliferative activity 45–54
Aqueous humor 149–154

B

Bioactive compounds 45–53, 96
Biomarkers of food intake 33
Brain microdialysates 96, 101

C

Capillary electrophoresis (CE) 46, 58, 87, 95, 96, 98, 99, 102, 103, 105–113, 116, 125, 134, 174, 211, 212, 242
Cell culture extracts 189–205
Childhood obesity 115–120, 123–130
C13-histidine 169–175
Cholesterol 143–148
Colon cancer 45–53, 83, 84

D

Data processing 38, 45, 46, 49, 51, 52, 80, 98, 108, 128, 183, 184, 213, 216–223, 229–231, 242, 243
Demyelinating diseases 169
Derivatization 96, 105–113, 241
Direct analysis 95–103

E

Erythrocyte 115–120, 126, 127
Extracellular vesicles 177–187
Extraction 3, 8, 13–16, 19, 21, 22, 30, 46–50, 59, 60, 67, 72–74, 77, 79, 112, 117–120, 134–137, 141, 144–147, 150, 152, 154, 182, 183, 185, 190, 191, 196–198, 200, 204, 205, 209, 213, 214, 218, 230, 260

F

Food by-products 45–54

G

Gas chromatography (GC) 46, 58, 116, 134, 143, 144, 147, 210, 212

H

HepaRG cells 191, 192, 194–197, 204
HepG2 cells 108–110, 112
High resolution mass spectrometry (HRMS) 34, 35, 38, 40, 51, 58, 62, 190
Human milk 177–187
Hydrophilic interaction liquid chromatography (HILIC) 36, 37, 41, 46, 117, 119, 120, 134, 135, 137–139, 141, 150, 190, 192, 199–203, 205

I

Imaging mass spectrometry (IMS) 34, 35, 38, 42, 157–167, 210
Instrument-agnostizing methodology 258, 264–266
Ion mobility 34, 35, 38, 190, 212, 213
Isobaric labelling 190, 201

L

L-DOPA 157–167
Lipidomics 177–187, 189–205
Lipids 42, 62, 64, 112, 124, 143–149, 166, 178, 182, 183, 190, 227, 230, 231
Liquid chromatography (LC) 2, 3, 8, 9, 34, 35, 37, 42, 46–48, 58, 59, 75, 117–119, 135, 143, 152, 153, 171, 173, 178, 179, 190, 198–200, 205, 210–212, 217, 220, 222, 229–232, 241–253, 257–260
Liver 108, 143–148, 191

Raúl González-Domínguez (ed.), *Mass Spectrometry for Metabolomics*,
Methods in Molecular Biology, vol. 2571, https://doi.org/10.1007/978-1-0716-2699-3,

© The Editor(s) (if applicable) and The Author(s), under exclusive license to Springer Science+Business Media, LLC, part of Springer Nature 2023

M

Mass spectrometry (MS) 2, 9, 11, 13–19, 21–31, 34, 38, 40, 46–48, 50–53, 57–68, 72, 75, 83, 85, 86, 95–101, 105–109, 116, 117, 119, 125, 128, 133–135, 137–139, 141, 143–147, 149–154, 157–159, 162, 163, 170, 171, 173, 174, 178, 179, 182–187, 190–192, 198–201, 210–214, 216, 217, 219, 220, 222, 226, 227, 229–232, 241–244, 250, 252, 253, 257, 258, 260, 261, 263–267
Metabolic profiling 209
Metabolite localization 157
Metabolomics 2, 33–42, 45–47, 51–54, 57–68, 71–80, 83, 84, 88, 105, 106, 115–120, 123, 133, 134, 137, 150, 185, 186, 190, 191, 205, 207–233, 241–253
Metals 123–130, 199, 205, 212
Multivariate analysis 40, 224, 225, 230, 231

P

Phenolic compounds 1–11
Plasma 36, 41, 58, 60, 62, 67, 73, 74, 79, 84, 85, 88, 106, 115–120, 125–128, 134, 177

Q

Quality assurance (QA) 134, 190, 191, 213, 218
Quality control (QC) 1, 37–41, 51, 60, 61, 67, 68, 75, 76, 80, 84, 120, 128–130, 134, 136, 137, 141, 152, 154, 182–185, 187, 190, 191, 198, 201, 205, 214–219, 222, 223, 230, 242–244, 249, 253

R

Remyelination 169–175
Reversed-phase liquid chromatography 35, 117, 190

S

Serotonin 157–167
Size fractionation 123–130
Solid phase microextraction (SPME) 14, 15, 19, 22
Spectroscopy 2
Statistical analysis 35, 38, 39, 45, 52, 59, 61, 80, 88, 213, 220, 229–231, 243, 244, 246, 253
Stir bar sorptive extraction (SBSE) 22, 25

U

Ultra-high performance liquid chromatography (UHPLC) 7, 35, 37, 117, 119, 135–139, 151, 212
Untargeted metabolomics 33–35, 39, 40, 42, 58, 72, 79, 133–135, 150, 189–205, 210, 214, 217, 218, 223, 226, 229
Urine 34–36, 41, 45, 58, 83–92, 116, 134, 177, 243, 249

V

Vinegar 13–19, 21–31
Volatile compounds 13, 16, 19, 22, 25, 26, 30, 31, 46, 144