Methods in Molecular Biology 1719

Springer Protocols

Michael Schrader
Lloyd Fricker *Editors*

# Peptidomics

## Methods and Strategies

Humana Press

# METHODS IN MOLECULAR BIOLOGY

*Series Editor*
**John M. Walker**
**School of Life and Medical Sciences**
**University of Hertfordshire**
**Hatfield, Hertfordshire, AL10 9AB, UK**

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

# Peptidomics

## Methods and Strategies

Edited by

**Michael Schrader**

*Department of Biotechnology and Bioinformatics, Weihenstephan-Tr. University of Applied Sciences, Freising, Bayern, Germany*

**Lloyd Fricker**

*Department of Molecular Pharmacology, Albert Einstein College of Medicine, Bronx, NY, USA*
*Department of Neuroscience, Albert Einstein College of Medicine, Bronx, NY, USA*

Humana Press

*Editors*
Michael Schrader
Department of Biotechnology and Bioinformatics
Weihenstephan-Tr. University of Applied Sciences
Freising, Bayern, Germany

Lloyd Fricker
Department of Molecular Pharmacology
Albert Einstein College of Medicine
Bronx, NY, USA

Department of Neuroscience
Albert Einstein College of Medicine
Bronx, NY, USA

ISSN 1064-3745 ISSN 1940-6029 (electronic)
Methods in Molecular Biology
ISBN 978-1-4939-7536-5 ISBN 978-1-4939-7537-2 (eBook)
https://doi.org/10.1007/978-1-4939-7537-2

Library of Congress Control Number: 2018930520



Printed on acid-free paper

This Humana Press imprint is published by Springer Nature
The registered company is Springer Science+Business Media, LLC
The registered company address is: 233 Spring Street, New York, NY 10013, U.S.A.

# Preface

This volume of *Methods in Molecular Biology* covers the field of peptidomics. Although peptides have been studied for over 100 years, and over 2.6 million articles on PubMed include this term, the vast majority of these studies are focused on a single peptide. In contrast, the field of peptidomics is much younger and not focused on a small number of peptides. Instead, peptidomics is aimed at a broad array of the peptides present in a biological sample—the peptidome. The identity of each peptide is given by its amino acid sequence and post-translational modifications. Sequence information is directly linked to the protein precursor of the peptide, and thus to the genome of the organism. Due to a great variability in arrangements, there are millions of potential peptides that could theoretically be produced from all proteins in one organism, although in practice the actual number of relatively abundant peptides is much lower. Ideally, peptidomics should detect all peptides present in a sample. However, this is not possible due to limitations in the techniques and/or the relative levels of some peptides in the sample. Peptides exhibit the same dynamic range as proteins, occurring from single copies up to being main components of a biological sample, and as with proteomics, it is difficult to detect the low abundance peptides. Also, with peptidomics studies aimed at the native forms of peptides in biological samples, there is no opportunity to enzymatically cut out a small part that generates intense and reliable signals, which is the typical approach in proteomics. Thus, peptidomics is the analysis of the detectable peptides in a sample, with newer instrumentation in chromatography and mass spectrometry pushing down the detection limits and improving the accuracy of the identification of peptides.

Part I introduces the field of peptidomics and describes many of the basic techniques used to detect and identify peptides. Chapter 1 describes the history of the field, starting before the term peptidomics was used, up to the present, thus giving an overview for scientists who are new to the field. Chapters 2, 3, and 4 describe important steps of sample preparation that are a prerequisite for high quality peptidomics results. For scientists coming into the field of peptidomics from the related field of proteomics, these chapters are essential in that the sample preparation and mass spectrometry conditions for endogenous peptides are different than those for proteins and mass spectrometry of tryptic peptides (commonly used for proteomic studies). Chapter 5 is focused on a subset of the peptidome—those peptides that function as intercellular messengers such as neuropeptides and peptide hormones. Chapter 6 is also focused on a subset of the peptidome—peptides that bind to (and are presumably substrates) of a peptidase. The example used in this chapter is for thimet oligopeptidase, although the basic technique can be applied to enrich substrates of any peptidase. The final chapter in Part I describes an approach to identify D-amino acid-containing peptides. Although only a small number of such peptides are currently known, this is presumably because the standard mass spectrometry-based methods for detecting and identifying peptides cannot distinguish between L- and D-amino acids.

While peptidomics is simply the identification of peptides in a biological sample, the field of quantitative peptidomics aims to measure levels of peptides as well as identify them. Various methods have been adapted from the field of quantitative proteomics and modified

for quantitative peptidomics. In contrast to quantitative proteomics, which typically averages the levels of multiple tryptic peptides to derive the overall level of the corresponding protein in each biological sample, quantitative peptidomics relies on the measurement of a single peptide in each sample. Part II of this book begins with an overview of the most commonly used methods and strategies for peptide quantitation, both absolute and relative (i.e., comparing levels between two or more samples). Additional Chapters 9–12 in this section describe specific methods for the relative quantitation of peptides between samples using isotopic labels (Chapters 9–11) or label-free approaches (Chapter 12).

Part III, with Chapters 13–28, includes a wide variety of protocols describing specific applications of peptidomics methods to study scientific questions. These chapters cover a broad range of biological species as well as sample types. The species under scrutiny include humans, cattle, rodents, frogs, fish, snakes, crustaceans, *Drosophila*, ants, spiders, *C. elegans,* and plants, and the sample types range from whole animals to single cells and from tissue extracts to body fluids or secretions from cells. Each species and sample type presents different challenges for peptide extraction, purification, and identification. Many of the protocols described in these chapters can be modified to examine other biological systems. Also, while specific equipment, software, and materials are described in each protocol and these are well established in the field, most can be substituted with other comparable equipment, software, and materials depending on what is available to the project. The aims of the studies described in this volume also cover a wide range. Several of the chapters address the search for neuropeptides or bioactive peptides in venoms or other samples. Some describe product-driven goals, such as cancer neo-antigens, antidiabetic agents, and biomarkers in body fluids or other samples such as dairy products. The final chapter describes some of the challenges and future directions for the field of peptidomics.

Collectively, the protocols in this volume describe relevant state-of-the-art approaches that will be of use for many years, even though the instruments used for peptide separation and mass spectrometry are rapidly improving. All of the basic approaches described in this volume are independent of the instruments used for analysis and can easily be used with newer equipment, providing larger numbers of peptides and higher confidence in the sequence identifications.

*Freising, Germany* — *Michael Schrader*
*Bronx, NY, USA* — *Lloyd Fricker*

# Acknowledgments

This volume has been a large effort from about 90 authors, all experts in the field of peptidomics. We would like to express our deepest gratitude to them for sharing their techniques to foster this field. The realization of this book would not have been possible without their efforts concerning the methods and protocols published here.

# Contents

*Preface* . . . . . *v*
*Contributors* . . . . . *xiii*

PART I OVERVIEW AND BASIC TECHNIQUES

1 Origins, Technological Development, and Applications of Peptidomics . . . . . 3
***Michael Schrader***

2 Brain Tissue Sample Stabilization and Extraction Strategies for Neuropeptidomics . . . . . 41
***Elva Fridjonsdottir, Anna Nilsson, Henrik Wadensten, and Per E. Andrén***

3 Isolation of Endogenous Peptides from Cultured Cell Conditioned Media for Mass Spectrometry . . . . . 51
***Kazuki Sasaki, Takashi Tsuchiya, and Tsukasa Osaki***

4 Mass Spectrometric Identification of Endogenous Peptides . . . . . 59
***Mikel Azkargorta, Iraide Escobes, Ibon Iloro, and Felix Elortza***

5 Bioinformatics for Prohormone and Neuropeptide Discovery . . . . . 71
***Bruce R. Southey, Elena V. Romanova, Sandra L. Rodriguez-Zas, and Jonathan V. Sweedler***

6 Substrate Capture Assay Using Inactive Oligopeptidases to Identify Novel Peptides . . . . . 97
***Vanessa Rioli and Emer S. Ferro***

7 Non-targeted Identification of D-Amino Acid-Containing Peptides Through Enzymatic Screening, Chiral Amino Acid Analysis, and LC-MS . . . . . 107
***Hua-Chia Tai, James W. Checco, and Jonathan V. Sweedler***

PART II QUANTITATIVE METHODS IN PEPTIDOMICS

8 Quantitative Peptidomics: General Considerations . . . . . 121
***Lloyd Fricker***

9 Quantitative Peptidomics with Isotopic and Isobaric Tags . . . . . 141
***Kurt Boonen, Wouter De Haes, Joris Van Houtven, Rik Verdonck, Geert Baggerman, Dirk Valkenborg, and Liliane Schoofs***

10 Quantitative Peptidomics Using Reductive Methylation of Amines . . . . . 161
***Sayani Dasgupta, Leandro M. Castro, Alexandre K. Tashima, and Lloyd Fricker***

11 Metabolic Labeling to Quantify *Drosophila* Neuropeptides and Peptide Hormones . . . . . 175
***Thomas Otto Kunz, Jiangtian Chen, Megha, and Christian Wegener***

12 Data Preprocessing, Visualization, and Statistical Analyses of Nontargeted Peptidomics Data from MALDI-MS . . . . . 187
*Harald Tammen and Rüdiger Hess*

Part III Diverse Applied Protocols and Outlook

13 Affinity Purification of Neuropeptide Precursors from Mice Lacking Carboxypeptidase E Activity . . . . . 199
*Lloyd Fricker*

14 Mass Spectrometry Based Immunopeptidomics for the Discovery of Cancer Neoantigens . . . . . 209
*Michal Bassani-Sternberg*

15 Milk Peptidomics to Identify Functional Peptides and for Quality Control of Dairy Products . . . . . 223
*David Dallas and Søren Drud Nielsen*

16 Neuropeptidomic Analysis of Zebrafish Brain . . . . . 241
*Kristien A. Van Camp, Geert Baggerman, Ronny Blust, and Steven J. Husson*

17 Identification, Quantitation, and Imaging of the Crustacean Peptidome . . . . . 247
*Kellen DeLaney, Amanda Buchberger, and Lingjun Li*

18 Identification of Endogenous Neuropeptides in the Nematode *C. elegans* Using Mass Spectrometry . . . . . 271
*Sven Van Bael, Samantha L. Edwards, Steven J. Husson, and Liesbet Temmerman*

19 EndoProteoFASP as a Tool to Unveil the Peptidome-Protease Profile: Application to Salivary Diagnostics . . . . . 293
*Fábio Trindade, Inês Falcão-Pires, Adelino Leite-Moreira, Pedro S. Gomes, Julie Klein, Rita Ferreira, and Rui Vitorino*

20 Methodology for Urine Peptidome Analysis Based on Nano-HPLC Coupled to Fourier Transform Ion Cyclotron Resonance Mass Spectrometry . . . . . 311
*Alexey S. Kononikhin, Victoria A. Sergeeva, Anna E. Bugrova, Maria I. Indeykina, Natalia L. Starodubtseva, Vitaliy V. Chagovets, Igor A. Popov, Vladimir E. Frankevich, Patrick Pedrioli, Gennady T. Sukhikh, and Eugene N. Nikolaev*

21 Identification of Components in Frog Skin Secretions with Therapeutic Potential as Antidiabetic Agents . . . . . 319
*J. Michael Conlon, R. Charlotte Moffett, Jérôme Leprince, and Peter R. Flatt*

22 High-Accuracy Mass Spectrometry Based Screening Method for the Discovery of Cysteine Containing Peptides in Animal Venoms and Toxins . . . . . 335
*Luuk N. van Oosten, Martijn W. H. Pinkse, Mervin Pieterse, Pierre Escoubas, and Peter D. E. M. Verhaert*

23 Analysis of the Snake Venom Peptidome . . . . . 349
*Solange M. T. Serrano, André Zelanis, Eduardo S. Kitano, and Alexandre K. Tashima*

24 Identification of Peptides in Spider Venom Using Mass Spectrometry . . . . . 359
***Rafael L. Lomazi, Erika S. Nishiduka, Pedro I. Silva Jr, and Alexandre K. Tashima***

25 Single Cell Peptidomics: Approach for Peptide Identification by N-Terminal Peptide Derivatization . . . . . 369
***Susanne Neupert***

26 Peptidomic Identification of Cysteine-Rich Peptides from Plants . . . . . 379
***Xinya Hemu, Aida Serra, Dina A. Darwis, Tobias Cornvik, Siu Kwan Sze, and James P. Tam***

27 Analysis of Endogenous Peptide Pools of *Physcomitrella patens* Moss . . . . . 395
***Igor Fesenko, Regina Khazigaleeva, Vadim Govorun, and Vadim Ivanov***

28 The Bright Future of Peptidomics . . . . . 407
***Peter D. E. M. Verhaert***

*Index* . . . . . *417*

# Contributors


Per E. Andrén • *Biomolecular Imaging and Proteomics, Department of Pharmaceutical Biosciences, Uppsala University, Uppsala, Sweden*

Mikel Azkargorta • *Proteomics Platform, ProteoRed-ISCIII, CIBERehd, CIC bioGUNE, Science and Technology Park of Bizkaia, Derio, Spain*

Geert Baggerman • *Department of Biology, Systemic Physiological & Ecotoxicological Research, University of Antwerp, Antwerp, Belgium; Flemish Institute for Technological Research (VITO), Mol, Belgium; Center for Proteomics (CFP), University of Antwerp, Antwerp, Belgium*

Michal Bassani-Sternberg • *Ludwig Centre for Cancer Research, University of Lausanne, Epalinges, Switzerland; Department of Fundamental Oncology, University Hospital of Lausanne, Lausanne, Switzerland*

Ronny Blust • *Department of Biology, Systemic Physiological and Ecotoxicological Research, University of Antwerp, Antwerp, Belgium*

Kurt Boonen • *Research Group of Functional Genomics and Proteomics, Department of Biology, KU Leuven, Leuven, Belgium*

Amanda Buchberger • *Department of Chemistry, University of Wisconsin—Madison, Madison, WI, USA*

Anna E. Bugrova • *V.I. Kulakov Research Center for Obstetrics, Gynecology and Perinatology, Ministry of Healthcare of the Russian Federation, Moscow, Russia; Emanuel Institute for Biochemical Physics, Russian Academy of Sciences, Moscow, Russia*

Leandro M. Castro • *Bioscience Institute, São Paulo State University, São Vicente, SP, Brazil*

Vitaliy V. Chagovets • *V.I. Kulakov Research Center for Obstetrics, Gynecology and Perinatology, Ministry of Healthcare of the Russian Federation, Moscow, Russia*

James W. Checco • *Department of Chemistry, Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana—Champaign, Urbana, IL, USA*

Jiangtian Chen • *Neurobiology and Genetics, Theodor-Boveri-Institute, Biocenter, University of Würzburg, Würzburg, Germany*

J. Michael Conlon • *SAAD Centre for Pharmacy and Diabetes, School of Biomedical Sciences, University of Ulster, Coleraine, UK*

Tobias Cornvik • *School of Biological Sciences, Nanyang Technological University, Singapore, Singapore*

David Dallas • *College of Public Health and Human Sciences, Oregon State University, Corvallis, OR, USA*

Dina A. Darwis • *School of Biological Sciences, Nanyang Technological University, Singapore, Singapore*

Sayani Dasgupta • *Department of Molecular Pharmacology, Albert Einstein College of Medicine, Bronx, NY, USA*

Wouter De Haes • *Research Groups of Functional Genomics and Proteomics, Department of Biology, KU Leuven, Leuven, Belgium; Research Group of Molecular and Functional Neurobiology, Department of Biology, KU Leuven, Leuven, Belgium*

Kellen DeLaney • *Department of Chemistry, University of Wisconsin—Madison, Madison, WI, USA*
Samantha L. Edwards • *Animal Physiology and Neurobiology, Department of Biology, KU Leuven (University of Leuven), Leuven, Belgium*
Felix Elortza • *Proteomics Platform, ProteoRed-ISCIII, CIBERehd, CIC bioGUNE, Science and Technology Park of Bizkaia, Derio, Spain*
Iraide Escobes • *Proteomics Platform, ProteoRed-ISCIII, CIBERehd, CIC bioGUNE, Science and Technology Park of Bizkaia, Derio, Spain*
Pierre Escoubas • *VenomeTech, Valbonne, France*
Inês Falcão-Pires • *Departamento de Cirurgia e Fisiologia, Faculdade de Medicina UnIC, Unidade de Investigação Cardiovascular, Universidade do Porto, Porto, Portugal*
Rita Ferreira • *Department of Chemistry, QOPNA, Mass Spectrometry Center, University of Aveiro, Aveiro, Portugal*
Emer S. Ferro • *Department of Pharmacology, Biomedical Science Institute, University of São Paulo, São Paulo, SP, Brazil*
Igor Fesenko • *Laboratory of Proteomics, Shemyakin and Ovchinnikov Institute of Bioorganic Chemistry, Russian Academy of Sciences, Moscow, Russia*
Peter R. Flatt • *SAAD Centre for Pharmacy and Diabetes, School of Biomedical Sciences, University of Ulster, Coleraine, UK*
Vladimir E. Frankevich • *V.I. Kulakov Research Center for Obstetrics, Gynecology and Perinatology, Ministry of Healthcare of the Russian Federation, Moscow, Russia*
Lloyd Fricker • *Department of Molecular Pharmacology, Albert Einstein College of Medicine, Bronx, NY, USA; Department of Neuroscience, Albert Einstein College of Medicine, Bronx, NY, USA*
Elva Fridjonsdottir • *Biomolecular Imaging and Proteomics, Department of Pharmaceutical Biosciences, Uppsala University, Uppsala, Sweden*
Pedro S. Gomes • *Faculty of Dental Medicine, Laboratory for Bone Metabolism and Regeneration, University of Porto, Porto, Portugal; REQUIMTE/LAQV, University of Porto, Porto, Portugal*
Vadim Govorun • *Laboratory of Proteomics, Shemyakin and Ovchinnikov Institute of Bioorganic Chemistry, Russian Academy of Sciences, Moscow, Russia; Laboratory of the Proteomic Analysis, Research Institute for Physico-Chemical Medicine, Moscow, Russia*
Xinya Hemu • *School of Biological Sciences, Nanyang Technological University, Singapore, Singapore*
Rüdiger Hess • *PXBioVisioN GmbH, Hannover, Germany*
Joris Van Houtven • *Department of Electrical Engineering (ESAT), STADIUS Center for Dynamical Systems, Signal Processing and Data Analytics Department, KU Leuven, Leuven, Belgium*
Steven J. Husson • *Department of Biology, University of Antwerp, Antwerp, Belgium*
Ibon Iloro • *Proteomics Platform, ProteoRed-ISCIII, CIBERehd, CIC bioGUNE, Science and Technology Park of Bizkaia, Derio, Spain*
Maria I. Indeykina • *Moscow Institute of Physics and Technology, Moscow, Russia; Emanuel Institute for Biochemical Physics, Russian Academy of Sciences, Moscow, Russia*
Vadim Ivanov • *Laboratory of Proteomics, Shemyakin and Ovchinnikov Institute of Bioorganic Chemistry, Russian Academy of Sciences, Moscow, Russia*
Regina Khazigaleeva • *Laboratory of Proteomics, Shemyakin and Ovchinnikov Institute of Bioorganic Chemistry, Russian Academy of Sciences, Moscow, Russia*

Eduardo S. Kitano • *Laboratório Especial de Toxinologia Aplicada, Center of Toxins, Immune-Response and Cell Signaling (CeTICS), Instituto Butantan, São Paulo, SP, Brazil*

Julie Klein • *INSERM U1048, Institute of Cardiovascular and Metabolic Disease, Toulouse, France; Université Toulouse III Paul-Sabatier, Toulouse, France*

Alexey S. Kononikhin • *Moscow Institute of Physics and Technology, Moscow, Russia; V.L. Talrose Institute for Energy Problems of Chemical Physics, Russian Academy of Sciences, Moscow, Russia; V.I. Kulakov Research Center for Obstetrics, Gynecology and Perinatology, Ministry of Healthcare of the Russian Federation, Moscow, Russia*

Thomas Otto Kunz • *Neurobiology and Genetics, Theodor-Boveri-Institute, Biocenter, University of Würzburg, Würzburg, Germany*

Adelino Leite-Moreira • *Departamento de Cirurgia e Fisiologia, Faculdade de Medicina, UnIC, Unidade de Investigação Cardiovascular, Universidade do Porto, Porto, Portugal*

Jérôme Leprince • *Normandie University, UNIROUEN, INSERM, DC2N, Rouen, France*

Lingjun Li • *Department of Chemistry, University of Wisconsin—Madison, Madison, WI, USA; School of Pharmacy, University of Wisconsin—Madison, Madison, WI, USA*

Rafael L. Lomazi • *Departamento de Bioquímica, Escola Paulista de Medicina, Universidade Federal de São Paulo, São Paulo, SP, Brazil*

Megha • *National Center for Biological Sciences, Tata Institute for Fundamental Research, Bangalore, India*

R. Charlotte Moffett • *SAAD Centre for Pharmacy and Diabetes, School of Biomedical Sciences, University of Ulster, Coleraine, UK*

Susanne Neupert • *Department for Biology, Zoological Institute, University of Cologne, Cologne, Germany*

Søren Drud Nielsen • *College of Public Health and Human Sciences, Oregon State University, Corvallis, OR, USA*

Eugene N. Nikolaev • *Moscow Institute of Physics and Technology, Moscow, Russia; V.L. Talrose Institute for Energy Problems of Chemical Physics, Russian Academy of Sciences, Moscow, Russia; Emanuel Institute for Biochemical Physics, Russian Academy of Sciences, Moscow, Russia; Skolkovo Institute of Science and Technology, Skolkovo, Moscow Region, Russia*

Anna Nilsson • *Biomolecular Imaging and Proteomics, Department of Pharmaceutical Biosciences, Uppsala University, Uppsala, Sweden*

Erika S. Nishiduka • *Departamento de Bioquímica, Escola Paulista de Medicina, Universidade Federal de São Paulo, São Paulo, SP, Brazil*

Mervin Pieterse • *Section for Analytical Biotechnology and Innovative Peptide Biology, Delft University of Technology, Delft, The Netherlands*

Martijn W.H. Pinkse • *Section for Analytical Biotechnology and Innovative Peptide Biology, Delft University of Technology, Delft, The Netherlands*

Igor A. Popov • *Moscow Institute of Physics and Technology, Moscow, Russia; V.L. Talrose Institute for Energy Problems of Chemical Physics, Russian Academy of Sciences, Moscow, Russia; V.I. Kulakov Research Center for Obstetrics, Gynecology and Perinatology, Ministry of Healthcare of the Russian Federation, Moscow, Russia*

Vanessa Rioli • *Special Laboratory of Applied Toxinology (LETA), Center of Toxins, Immune Response and Cell Signaling (CETICS), Butantan Institute, São Paulo, SP, Brazil*

Sandra L. Rodriguez-Zas • *Department of Animal Sciences, University of Illinois at Urbana—Champaign, Urbana, IL, USA*
Elena V. Romanova • *Department of Chemistry and Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana—Champaign, Urbana, IL, USA*
Kazuki Sasaki • *National Cerebral and Cardiovascular Center, Osaka, Japan*
Michael Schrader • *Department of Biotechnology and Bioinformatics, Weihenstephan-Tr. University of Applied Sciences, Freising, Germany*
Liliane Schoofs • *Research Group of Functional Genomics and Proteomics, Department of Biology, KU Leuven, Leuven, Belgium*
Victoria A. Sergeeva • *Moscow Institute of Physics and Technology, Moscow, Russia; V.L. Talrose Institute for Energy Problems of Chemical Physics, Russian Academy of Sciences, Moscow, Russia; Emanuel Institute for Biochemical Physics, Russian Academy of Sciences, Moscow, Russia*
Aida Serra • *School of Biological Sciences, Nanyang Technological University, Singapore, Singapore*
Solange M.T. Serrano • *Laboratório Especial de Toxinologia Aplicada, Center of Toxins, Immune-Response and Cell Signaling (CeTICS), Instituto Butantan, São Paulo, SP, Brazil*
Pedro I. Silva Jr • *Laboratório Especial de Toxinologia Aplicada, Center of Toxins, Immune-Response and Cell Signaling (CeTICS), Instituto Butantan, São Paulo, SP, Brazil*
Bruce R. Southey • *Department of Animal Sciences, University of Illinois at Urbana—Champaign, Urbana, IL, USA*
Natalia L. Starodubtseva • *Moscow Institute of Physics and Technology, Moscow, Russia; V.I. Kulakov Research Center for Obstetrics, Gynecology and Perinatology, Ministry of Healthcare of the Russian Federation, Moscow, Russia*
Gennady T. Sukhikh • *V.I. Kulakov Research Center for Obstetrics, Gynecology and Perinatology, Ministry of Healthcare of the Russian Federation, Moscow, Russia*
Jonathan V. Sweedler • *Department of Chemistry and Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana—Champaign, Urbana, IL, USA*
Siu Kwan Sze • *School of Biological Sciences, Nanyang Technological University, Singapore, Singapore*
Hua-Chia Tai • *Department of Chemistry, Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana—Champaign, Urbana, IL, USA*
James P. Tam • *School of Biological Sciences, Nanyang Technological University, Singapore, Singapore*
Harald Tammen • *PXBioVisioN GmbH, Hannover, Germany*
Alexandre K. Tashima • *Department of Biochemistry, Escola Paulista de Medicina, Federal University of São Paulo, São Paulo, SP, Brazil*
Liesbet Temmerman • *Animal Physiology and Neurobiology, Department of Biology, KU Leuven (University of Leuven), Leuven, Belgium*
Fábio Trindade • *Department of Medical Sciences, iBiMED, Institute for Biomedicine, University of Aveiro, Aveiro, Portugal; Departamento de Cirurgia e Fisiologia, Faculdade de Medicina, UnIC, Unidade de Investigação Cardiovascular, Universidade do Porto, Porto, Portugal*
Dirk Valkenborg • *Center for Proteomics, University of Antwerp, Antwerp, Belgium; Interuniversity Institute for Biostatistics and Statistical Bioinformatics, Hasselt University, Diepenbeek, Belgium*

Sven Van Bael • *Animal Physiology and Neurobiology, Department of Biology, KU Leuven (University of Leuven), Leuven, Belgium*
Kristien A. Van Camp • *Department of Biology, Systemic Physiological and Ecotoxicological Research, University of Antwerp, Antwerp, Belgium*
Luuk N. van Oosten • *Section for Analytical Biotechnology and Innovative Peptide Biology, Delft University of Technology, Delft, The Netherlands; Institute of Pharmacy and Molecular Biotechnology, Heidelberg University, Heidelberg, Germany*
Rik Verdonck • *Research Group of Molecular Developmental Physiology and Signal Transduction, Department of Biology, KU Leuven, Leuven, Belgium*
Peter D.E.M. Verhaert • *University of Maastricht Multimodal Molecular Imaging Institute (M4i), Faculty of Health, Medicine and Life Sciences, Maastricht, Netherlands; ProteoFormiX, Beerse, Belgium*
Rui Vitorino • *Department of Medical Sciences, iBiMED, Institute for Biomedicine, University of Aveiro, Aveiro, Portugal; Departamento de Cirurgia e Fisiologia, Faculdade de Medicina, UnIC, Unidade de Investigação Cardiovascular, Universidade do Porto, Porto, Portugal*
Henrik Wadensten • *Biomolecular Imaging and Proteomics, Department of Pharmaceutical Biosciences, Uppsala University, Uppsala, Sweden*
Christian Wegener • *Neurobiology and Genetics, Theodor-Boveri-Institute, Biocenter, University of Würzburg, Würzburg, Germany*
André Zelanis • *Instituto de Ciência e Tecnologia, Universidade Federal de São Paulo, São José dos Campos, SP, Brazil*

# Part I

## Overview and Basic Techniques

# Chapter 1

# Origins, Technological Development, and Applications of Peptidomics

Michael Schrader

## Abstract

Peptidomics is the comprehensive characterization of peptides from biological sources mainly by HPLC and mass spectrometry. Mass spectrometry allows the detection of a multitude of single peptides in complex mixtures. The term first appeared in full papers in the year 2001, after over 100 years of peptide research with a main focus on one or a few specific peptides. Within the last 15 years, this new field has grown to over 1200 publications. Mass spectrometry techniques, in combination with other analytical methods, were developed for the fast and comprehensive analysis of peptides in proteomics and specifically adjusted to implement peptidomics technologies. Although peptidomics is closely linked to proteomics, there are fundamental differences with conventional bottom-up proteomics. The development of peptidomics is described, including the most important implementations for its technological basis. Different strategies are covered which are applied to several important applications, such as neuropeptidomics and discovery of bioactive peptides or biomarkers. This overview includes links to all other chapters in the book as well as recent developments of separation, mass spectrometric, and data processing technologies. Additionally, some new applications in food and plant peptidomics as well as immunopeptidomics are introduced.

**Key words** Peptidomic, Peptidome, Peptide research, Peptide analysis, Mass spectrometry

## 1 Introduction

Peptidomics is the study of the peptidome, which is defined as the peptides present in a biological sample. Currently, peptidomics studies involve the comprehensive analysis of native peptides by HPLC and mass spectrometry (MS). This analytical methodology works without prior knowledge of biological activity. MS provides information about single peptides, even in very complex mixtures. The first attempt of such an approach had been taken by Desiderio in 1981 [1]. It allowed to detect and quantify picomole amounts of intact endogenous, chemically underivatized oligopeptides extracted from biological tissue. However, it took 20 more years to define this specific field of research with the new term "peptidomics" for the comprehensive characterization of peptides present in a biological sample. Consistently, four groups independently

Michael Schrader and Lloyd Fricker (eds.), *Peptidomics: Methods and Strategies*, Methods in Molecular Biology, vol. 1719, https://doi.org/10.1007/978-1-4939-7537-2_1, 

published the first full scientific papers in 2001 using "peptidomics" or "peptidome" in their titles [2–6] whereas a Swedish group introduced "peptidomics" in another abstract [7]. Shortly before, it had already been preliminarily coined at a meeting in the UK organized by Micromass [8] and almost simultaneously by a trademark of the biotech company BioVisioN (Hannover, Germany). From the latter, an abstract of a talk at the ABRF conference "From Singular to Global Analyses of Biological Systems" in February 2000 [9] was first to generally publish the concept of peptidomics. Technologically, it is closely related to proteomics which is more widely used, but although principally based on the same instrumentation peptidomics requires different analytical approaches from those used in conventional bottom-up proteomics and also needs different know-how [10]. Such kind of know-how was developed intensively in parallel to proteomics activities, after the breakthrough in biological MS given by the inventions of electrospray [11] and and matrix-assisted laser desorption/ionization (MALDI-) MS [12] in the late 1980s. In the following, this has included the ability of MS to largely replace N-terminal Edman sequencing [13] by tandem MS methods [2, 3], similar to proteomics [14].

It took until 2005 that the term peptidomics (or peptidomic or peptidome), developed in Europe and Japan, received a wider acceptance (Fig. 1). Since then, these terms have been used in more and more publications. PubMed searches reveal a total of around 1200 publications until the end of 2016. A few conferences, special issues, and books have been dedicated to peptidomics (e.g., [15–18]). For 2010 and every year since 2012, at least 100 full papers are listed on PubMed with continuous increase of the average output (Fig. 1). Despite the current increase in interest, there has been a lack for an updated volume covering the specific experimental needs of this new research area especially concerning the adjusted analytical strategies and methods and the development in different areas of applications.

Peptides are present in large numbers and in varying amounts in all human body fluids, cells, and within tissues. They have many physiological functions, for example, as hormonal messengers, cytokines, antimicrobial agents, and protease inhibitors. The molecular form of these bioactive peptides usually cannot be directly predicted from the genome sequence as these are liberated by specific multistep processing pathways (covered in depth in several monographs, e.g., [19–22]). Moreover, many more peptides serve as "information carriers" reflecting the status of parts of or even an entire organism since they are generated as a result of such metabolism, or more unspecific proteolytic degradation of larger proteins by different types of peptidases. The last aspect gave rise to their name (from greek *peptos, peptein*-digested, to digest). Peptides themselves usually show no enzymatic activity. There is still no clear-cut definition

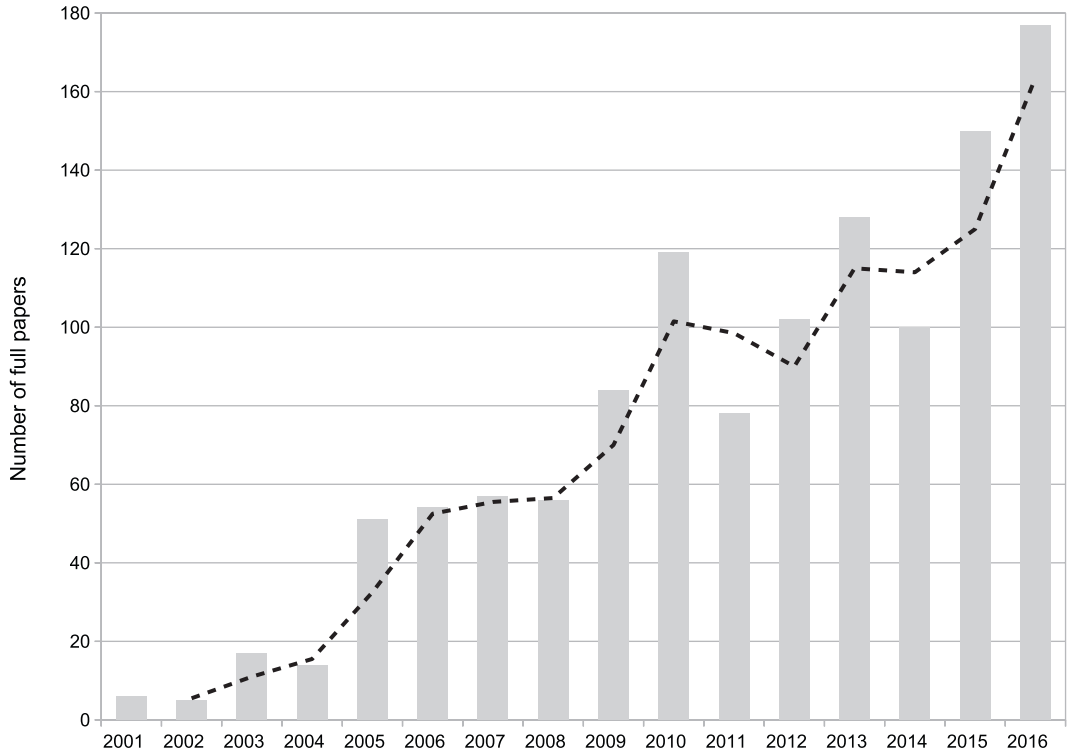



**Fig. 1** Number of full papers in peptidomics cited in PubMed per year and gliding average for two consecutive years, respectively (2001–2016). Search terms (peptidome* or peptidomic* or neuropeptidomic* or neuropeptidome* or immunopeptidome* or immunopeptidomic*) not peptidomet*

for a peptide if compared to a protein. The current IUPAC definition for peptides [23] states "Peptides are amides derived from two or more amino carboxylic acid molecules (the same or different) by formation of a covalent bond …," in principle including proteins as large peptides. The former IUPAC definition named a boundary for proteins: "Naturally occurring and synthetic polypeptides having molecular weights greater than about 10,000 (the limit is not precise)" [24]. This rather diffuse boundary results from the inherent overlap in molecular behavior of big peptides and small proteins. A somewhat fitting characteristic is the tertiary structure of proteins that typically consists of several secondary structure elements that allows for a specific biological function and the possibility to denaturate to a diversity of stable but inactive isoforms. For this review, the term peptide shall be applied in the range of 0.1–10 kDa and may sometimes include entities up to 15 kDa. In this respect, peptidomics complements the analysis of proteomes and other biochemical repertoires (Table 1) with several special characteristics due to the biochemical properties and biological conditions for peptidomes.

**Table 1**
**Comparison of main molecular repertoires addressed by "-omics" technologies with respect to their analytical characterization (with approximations for human biochemistry)**

| Biochemical repertoire and molecular basis | Diversity in humans (estimated number of molecular entities) | Current analytical coverage of molecular identities | Approximate half-life of such molecules in human body | Specificity of a single analyte to a single gene |
|---|---|---|---|---|
| Genome (DNA) | Intermediate ($2 \times 10^4$) | Completed [319, 320] | Months to years | Highest |
| Transcriptome (RNA) | High (>several $10^4$, incl. splicing) | Practically complete [319, 321] | Hours | Very high |
| Proteome (proteins) | High (>many $10^4$, incl. isoforms and PTMs) | Almost complete [320, 321] | Hours to weeks | Very high |
| Peptidome (peptides) | Very high (»many $10^4$, incl. PTMs and truncations) | Mainly incomplete (own estimate) | Minutes to days | Intermediate to very high, depending on number of amino acids |
| Metabolome (heterogeneous repository of compounds with low molecular mass) | Intermediate(>several $10^4$, with very diverse composition) | Substantial part completed [322, 323] | Diverse, depending on chemical nature | Low to no link, depending on type of molecule |
| Lipidome (lipids) | High (>several $10^4$) | Mainly incomplete [324] | Minutes to days | No link |
| Glycome (carbohydrates, glycans) | Very high (>several $10^4$) | Mainly incomplete [325] | Hours to days | No link |

Peptides are thus one of the most complex and challenging classes of biomolecules, still mainly a *terra incognita*, even in the human body. The peptidome is rapidly changing in time, with single peptides comprising relatively high specificity, being thus ideal signaling molecules as well as potential biomarkers which are directly linked to a protein precursor and subsequent gene [4]
*PTM* posttranslational modification

## 2 Peptidomics as a New Research Area

### *2.1 Historical Origin Is the Discovery of Peptide Hormones*

Bioactive peptides have been studied for over 100 years, ever since Bayliss and Starling found secretin in intestinal extracts [25]. This discovery led to a search for other members of this new class of signaling molecules named peptide hormones that when released from one tissue could excite or stimulate organ function in a different location. Many other peptide hormones were discovered using

a similar approach in which a biological activity was traced to purify a new substance. After that, amino acid sequencing revealed the identity of the underlying bioactive peptide. Several of these discoveries resulted in Nobel Prizes; for example one was awarded in 1923 in Physiology or Medicine, divided between Banting and MacLeod, for the discovery of insulin [26]. The respective novel scientific insights of how the body communicates internally, locally between cells, and especially via the bloodstream between distant cells and entire organ systems were groundbreaking. Similar methods were used to discover peptide neurotransmitters and peptidergic neuromodulators, collectively termed neuropeptides.

Despite knowledge of its existence, it took more than 50 years to determine the exact molecular nature of secretin and publication of its sequence containing 27 amino acids (3 kDa), arranged as a linear chain [27]. The much more complex sequence of insulin (6 kDa, 51 amino acids, 3 disulfide bonds) could be characterized even one decade earlier [28, 29] because of the availability of sufficient amounts of insulin isolated from animals for pharmaceutical purposes; however, it took about 10 years to resolve. Optimized analytical methods developed at that time were amino acid analysis [30] and N-terminal sequencing [13]. The main problem was the necessity for purification of substantial amounts of a pure molecular compound to perform these analyses. Therefore, up to tons of tissues had to be extracted and further fractionated over several steps followed by laborious bioassays to subsequently localize the hormones of interest [31]. Proctolin, the first insect neuropeptide for example, was purified from 125 kg of whole cockroaches to yield approximately 180 μg of this pentapeptide [32]. The outstanding prerequisite for isolation of almost all early discoveries of peptide hormones was the availability of enough starting material to pass through several purification steps. This can be attributed to the generally small concentrations of single-peptide hormones because of their rapid turnover and the overwhelming amounts of surrounding housekeeping proteins with similar chemical nature and huge numbers of other peptides.

Several researchers characterized peptides present in biological samples using upcoming analytical tools without primarily focusing on biological activity. In the early 1980s, Mutt, Tatemoto, and colleagues purified peptides from pig intestine. They replaced bioassays for the first time by screening for the presence of a C-terminal amide group, recognizing at that time that this posttranslational modification is a common feature of peptide hormones [33]. Using this approach, several new neuropeptides and/or peptide hormones were identified. Such “peptide-first” approaches were forerunners to peptidomics studies, but cannot be designated as peptidomics because they followed a single structural feature.

### 2.2 Technological Origins from Instrumental Analytics

Peptidomic technology has been substantially driven by innovations in analytical chemistry. As bioactive peptides occur in very low concentrations within complex biological matrices, analytical methods have to be very sensitive and rely on specific sample preparation strategies depending on the biological source. The complexity of vertebrate peptidomes, for example, is extremely high and samples usually contain many peptidases, too. Peptide extraction from biological sources thus needs fast steps to avoid regular degradation ([34, 35] and chapter 2 in this volume). Most of the more basic information is spread over a multitude of publications. This book thus aims to give detailed current protocols for different species and specimens within its further chapters.

Importantly, the conceptual breakthrough in peptidomics would not have been possible without the inventions of ESI-MS [11] and MALDI-MS [12]. They almost immediately ruled out the formerly applied ionization by fast atom bombardment (FAB) [36] as well as field and plasma desorption [37, 38]. FAB-MS could be already applied to complex biochemical questions [39, 40]; however, all these ionization methods still needed nanomolar amounts of sample and/or long measurement times [41, 42]. Typically, MS gets more difficult the bigger the molecular analyte is. This depends on the difficulty in ionizing and vaporizing a big molecule without breaking its molecular bonds at the same time. Because of this, bottom-up proteomics studies cut proteins into many small peptides prior to MS analysis. Also, the early mass spectrometers had a rather low resolution (less than 1000) and a low mass accuracy compared to modern instrumentation (Table 2). Exact masses using isotopic patterns could be generated for smaller peptides only. These advantages of peptides being readily accessible to MS [43] rapidly became of interest to a few groups analyzing peptide hormones or neuropeptides that had the additionally necessary know-how in analytical chemistry at their command (e.g., [39, 40, 44]).

The development of a large variety of high-resolution chromatographic media in liquid chromatography (LC), especially reversed phase (RP) or ion exchange (IEX), enabled improved purification of low-concentrated compounds from complex matrices [45–50]. Even though all methods are available for long, this is still a critical step [51] requiring sufficient practical knowledge. Successful use of a two-dimensional combination of IEX-chromatography and RP-HPLC in peptide research includes a systematic isolation of peptides from adrenal chromaffin vesicles using FAB-MS or ESI-MS to determine the number of peptide species in individual chromatographic fractions [47]. Gel electrophoresis, still a gold standard for protein purification, was of little need to separate peptides as their molecular size is too small to be separated with high resolution. On the other hand, the development of two-dimensional gel electrophoresis became an important part in the

**Table 2**
**Development steps in the mass spectrometric (MS) analysis of peptides**

| Approximate time frame | 1980–1989 | 1990–1999 | 2000–2005 | 2006–2016 |
|---|---|---|---|---|
| MS ionization methods mostly applied | Fast atom bombardment (FAB), plasma and field desorption | MALDI (1 Hz) + ESI + microESI + LC-ESI | nanoESI + DE-MALDI (20 Hz) + microLC-ESI | nanoLC-ESI + DE-MALDI (1000 Hz) |
| Sample amount needed for MS | Nanomoles | Picomoles | Femtomoles | Attomoles |
| Time needed for a good spectrum | Minutes to hours | Minutes | Seconds | Milliseconds |
| Routine MS resolution | Up to 1000 | 1000–15,000 | 5000–20,000 | 10,000–100,000 |
| Typical mass accuracy | <1000 ppm | <200 ppm | <20 ppm | <5 ppm |
| Identification methods | Chemical Edman sequencing | Edman sequencing + CID- or PSD-MS/MS | Database searching after CID- or PSD-MS/MS | CID- and ETD-MS/MS, incl. automated de novo MS/MS |
| Favored MS analyzers | Sector fields and quadrupole (Q) | Triple quadrupole and time-of-flight (TOF) | Ion trap, Q-TOF, TOF/TOF and FT-ICR | hybrid multi-analyzers, often with Orbitrap |
| MS quantification methods | Too elaborate by MS | Differential MALDI- and LC-ESI-MS | First isotopic labels applied | Labeling reagent kits introducing multiple isotopes |

Introduction of matrix-assisted laser desorption/ionization (MALDI) in 1988 [12] and electrospray ionization (ESI) in 1989 [11] resulted in ideal detection methods for peptides. Further optimizations laid the ground for the current detection principles in peptidomics
*DE* delayed extraction, *CID* collision-induced dissociation, *PSD* post-source decay, *FT-ICR* Fourier transform ion cyclotron resonance, *ETD* Electron transfer dissociation

idea to foster a comprehensive analysis of all proteins in a biological sample which transformed traditional protein chemistry into the research area of proteomics [52]. The combination of technical advances in separation techniques, MS, and bioinformatics made this possible. Potential use in the new biopharmaceutical industry further fostered this field enormously [53].

### 2.3 MS as the Central Tool for Peptide Identification

After many successes, biological MS was broadly used to detect peptides since the 1990s, soon after first publications of both new ionization methods. MS sensitivity now was high enough, yielding results within minutes while needing only picomole amounts of peptides [43] (Table 2). First approaches of ESI-MS combined the classical identification by N-terminal chemical sequencing with MS, especially for large peptides, such as the 10 kDa precursor of human guanylin [54], a 48-mer insect peptide [55], or in the identification of special sequence motifs from locusts [56]. A drastic increase in sensitivity of ESI-MS was achieved by the invention of micro- [57] and nanoelectrospray [58], allowing to generate mass spectra from femtomolar amounts of peptides. The latter method thus became a standard in peptidomics. MALDI-MS in delayed extraction (DE) mode [59] offered a similar sensitivity. The capability of MALDI-MS to deliver a very fast screening of hundreds of samples allowed a new way of assaying peptides. Its application for complex mixtures needed an optimization of its manually simple but complex combination of parallel physicochemical processes within sample preparation [60]. This has been applied on, for example, profiling of peptide moieties in marine invertebrates [61], spider venoms [62], or single neurons [63, 64].

In the 1990s, the analysis of purified peptide hormones and neuropeptides often involved MS, but final identification relied on N-terminal chemical sequencing. Another major breakthrough was necessary to propel the ideas of peptidomics: the use of tandem MS (MS/MS). Gas-phase fragmentation is an inherent phenomenon during energy uptake of molecules in vacuum, which can be ideally useful while analyzed in two-stage mass spectrometers. After selection of peptide ions by their mass-to-charge ratio in the first part of the mass spectrometer and subsequent fragmentation of these, a second mass analysis results in a fragment mass spectrum that is related to the sequence of the peptide. First examples came up long before peptidomics appeared, as it was possible to chemically modify peptides and interpret these directly in cases of short or partly known sequences [65–68]. To overcome a necessary chemical derivatization of the peptides, more suitable new mass spectrometric methods were applied using collision-induced dissociation (CID) for sequence analysis [69]. Moreover, CID in combination with ESI-MS/MS delivered much better data without the need for chemical sample preparation and a substantially enhanced throughput and speed of the fragmentation and sequenc-

ing processes as well as straightforward interpretation [43, 70, 71]. These methods developed into the standard protein and peptide sequencing tool, in peptide research for example applied on proteolytic degradation of a peptide hormone [72] or in neuropeptide discovery [73], sometimes still complemented by N-terminal chemical sequencing [74].

Especially the availability of MS for nonspecialists in instrument development [75] and its widespread application in biological sciences were the origins for the core peptidomic technology. Genomics activities in parallel generated high-quality protein sequence databases with SWISS-PROT as the best curated protagonist in the field (*see* [76] for a review). This led to the development of dedicated software allowing automated high-precision identification [77–79]. Since then, proteins are thus digested into smaller peptides to perform proteomics experiments of tens, later hundreds, and today thousands of peptides per hour.

### 2.4 Peptide Profiling as Forerunner of Peptidomics

An ideal combination is the online coupling of sophisticated purification tools to mass spectrometric detection. Early analyses by liquid chromatography coupled with mass spectrometry (LC-MS) demonstrated the usefulness of combining retention time and mass spectrometric sequence information for mixtures of peptides [80] in a much better way than using gas chromatography, which required small chemically modified peptides [81]. In 1981, data obtained with field desorption MS provided, for the first time, measurement of intact, chemically underivatized peptides extracted from biological matrices [1]. To my knowledge, this was the first attempt that could actually be claimed as peptidomics. Retrospectively, earliest quasi-peptidomic technology furthermore was used to characterize frog secretions [39] or to determine the amino acid sequence information of opioid peptides [40] as well as the sequence determination of peptides generated from proteins by enzymatic or chemical cleavage [82]. The liquid inlet of ESI-MS at atmospheric pressure has the clear advantage of its use coupled online to separations in aqueous phase. Application of the corresponding LC-MS methods allowed for the analysis of subpicomolar amounts of shorter peptides binding to major histocompatibility complex class I and II of the immune system [83–85]. These studies included quantification of hundreds of peptides as well as identification of many immunopeptides by MS/MS and Edman sequencing and thus were in advance to immunopeptidomics introduced later. Other complex biological sources for peptides could also be resolved by ESI-LC-MS [72]. Later studies used microbore ESI-LC-MS methods to quantify neuropeptides [86] and to detect and map peptides in human urine [87] as well as several thousands of circulating human blood peptides [88, 49]. Similar approaches disclosed the peptide repertoire of human cerebrospinal fluid (CSF) [89] or

compared different clinical samples of CSF [90]. Another study used ESI-LC-MS to identify known and novel neuropeptide precursors in brain extracts purified from mutant mice that lacked a critical neuropeptide processing enzyme [91, 73]. The LC-MS combination was then further developed to be operated in an automated mode [92].

The apparent alternative for the very sophisticated instrumental combination in LC-MS was an analysis by MALDI-MS. The high sensitivity of MALDI-MS linked with the ability to simultaneously display up to about hundred peptides in a single spectrum allows for a fast profiling of complex mixtures [93]. Together with the much lower interference by salts and other components compared to ESI-MS, peptide hormones can be detected from complex biological mixtures with reduced sample preparation and measurement efforts. It thus became possible to analyze tissue preparations [94], single cells [61, 95, 96], and even organelles [97]. MALDI-MS requires a vacuum and has to be applied off-line after LC separations. However, it is a fast screening tool that can be used to examine many fractions [62, 64, 98, 99]. Therefore, it was the main basis in the comprehensive mapping of a large natural peptide library from blood filtrate [48, 88, 100]. Comprehensive analyses of body fluids as well as from tissue extracts regularly reported thousands of peptides to be present in these samples [4]. Profiling activities by MALDI-MS were further developed to allow for relative quantification. First applications were the detection of molecules in spider venoms [62, 101]. Also, attempts had been taken to replace laborious bioassays for known entities by a mass spectrometric assay [98]. Differential studies analyzed changes induced during the immune response of Drosophila [102], comparison of spectra from neurointermediate lobes of individual rats [103], or screening for disulfide-rich peptides after derivatization by carboxyamidomethylation in human urine [104] as well as blood filtrate [105].

Most of the studies described in this section did not use the term peptidomics, but most are forerunners because they described unbiased searches for peptides in a biological sample. While these early studies identified some of the peptides present in the samples, they were limited by the sensitivity, speed, and accuracy of the available mass spectrometers, and by the computers, software, and protein databases available at the time. Developments in these areas were initially made for early proteomic applications, and the principal concepts were adopted for use in peptidomics. Thus, a virtual cooperation started combining a diverse field of applications, united by a rather homogeneous set of methods.

### *2.5 Precision Adjustments of Peptidomic Technology*

Combining the first approaches for identification and relative quantification by MS, the tools for complete peptidomic analysis were basically ready for application [2–4, 6]. The process chain of peptidomics contains a combination of several steps, starting with the extraction of peptides and finally leading to their identification (Fig. 2). Their sequence could be the basis for further quantitative

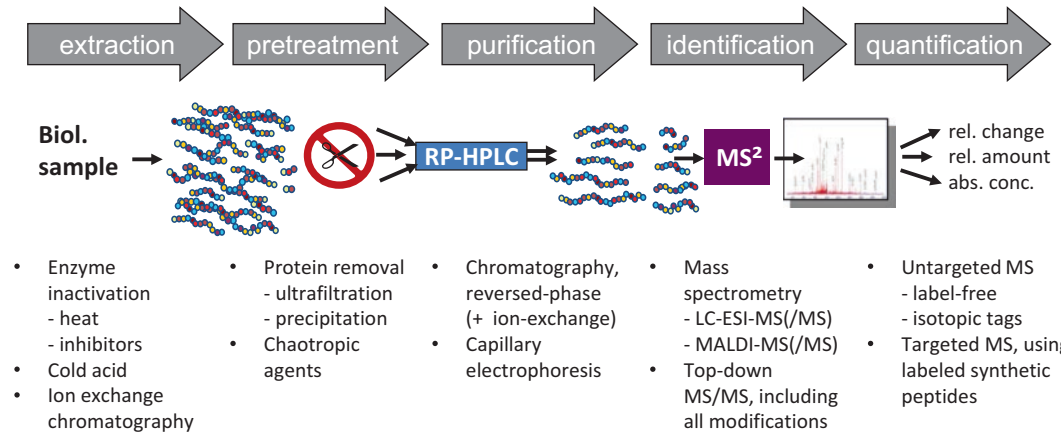



**Fig. 2** Workflow of the peptidomics analytical process which is divided into five main steps involving different types of methods used (a comparison to proteomics is given elsewhere [10]). More experienced groups change the order by quantifying first to identify peptides only with relevant changes of their concentration. Mass spectrometric quantification is described in detail in further chapters (*see* chapter 8 and part II of this book)

experiments. This has also been the order for the chapters in the first part of this book. It took up to the year 2001 to develop all comprised methods in a reproducible manner, and integrate them to be ready for automation, including proper data interpretation (for a contemporary review *see* [106]). A few of the most methodologically mature groups in peptidomics changed this order of the experiments in the process chain. Before identification, quantifications of a vast number of peptides are performed by different high-throughput methods, all based on MS (*see* respective overview chapter 8 and related chapters in part II of this book). This has been due to the results of early applications of peptidomics on body fluids and also tissue extracts revealing that the vast majority of the identified peptides were protein degradation fragments [89, 100, 107, 108]. It had led to an uncertainty about the scientific value of just "mapping" thousands of peptides in biological samples which is a general question in all -omics technologies (compare Table 1). On the other hand, some proteolytic cleavages play very important roles in the activation of neuropeptides and peptide hormones as well as other bioactive peptides (*see* for example [109–114]). Proteolytic processing by peptidases might be the most underestimated posttranslational processing step, at least for the generation of peptides [115–119].

Beyond generating lists of molecular inventories, peptidomics expanded to include targeted approaches examining posttranslational modifications (PTM) related to molecular function such as mentioned later for disulfides or cleavage sites of specific peptidases. Alternatively, single neurons were targeted for well-directed discovery of neuropeptides (*see* [95, 120–122] and chapter 25 in this book). Another major improvement was the early development of quantification methods for peptides [123, 124], leaned on techniques used for quantitation in proteomics [14]. The ability to

perform relative comparisons of levels of hundreds of peptides among two or more samples has intensively been further developed and has provided many hints as to the function of specific peptides, based on their regulation (*see* overview in chapter 8 and following chapters in this volume).

In peptidomics, each sequence identification has to rely on MS/MS data of a single peptide which therefore has to be very carefully interpreted, most often by manual validation of the results from automated software [125] which is different to identification of tryptic peptides in proteomics [126–129]. It includes mass shifts by PTMs as well as amino acid mutations requiring a combination of de novo sequencing and well-directed database searches (*see* Subheading 4.2). In quantitative proteomics it is well known that different features of the amino acid sequence may impede or act as positive discriminators for successful identifications [130]. Moreover, charge distribution, especially for positively charged hydrophilic residues, is a classifier for peptides being analyzed through the electrospray process. This option in proteomics to focus on a limited subset of peptides that lead to easier identification of a protein and offer a better behavior in quantification is not given in peptidomics. Any single peptide of interest typically does not adhere to several of the favored features. Mass spectrometric analysis of endogenous peptides thus presents a number of drawbacks, especially hardly predictable ionization and non-predictable fragmentation patterns ([131–133] and chapter 4 in this book).

Posttranslational modifications are often important for biological function and need additional efforts in MS analysis [43, 134, 135]. Disulfide bonds are one important modification (*see* Subheading 3.2). Direct sequencing of such peptides by MS/MS or analysis of their disulfide connectivity is much more complicated [136–140] and correct quantification of these also might be an issue [141]. Disulfide bonds as well as N-terminal pyroglutamylation or C-terminal amidation [91, 142] prevent peptides from premature proteolytic degradation and thus improve the half-life of many peptides with biological function. Several other PTMs such as glycosylation and phosphorylation are rather seldom described in combination with peptidomic studies [120]. Glycosylations are more often found for larger proteins, and phosphorylation is typical for intracellular content which is less frequently addressed by peptidomics, despite for some neuropeptides [35, 120]. The correct identification of phosphorylated peptides is rather difficult, although recent improvements of mass spectrometric technology are helping to solve this problem [143]. Terminal modifications with amides or pyroglutamylation are much easier to handle in this respect [91, 144, 145]. On the other hand, some less common PTMs difficult to analyze like palmitoylations or other lipid modifications are even seldom looked at until recent development of new MS methods [146]. Another very specific PTM is the isomerization of L- to D-amino acids which does not change the peptide's mass and is thus difficult to detect by MS (*see* chapter 7 in this book).

Quantification of hits in peptidomics relies on a perfect interplay of chromatographic separation and subsequent mass spectrometric quantification [106, 124, 147]. To increase specificity for a given analyte, additional MS/MS information is often used to select single peptides from complex samples like saliva [148]. Even more specific and accurate is the use of isotopic labeling where signals are partly shifted because of heavier isotopes [149, 150]. Several chapters in this book give further explanations about strategies and protocols used (*see* especially chapters 8 and 9). This concept is driven by the idea that only peptides with altered expression or release rates are of further interest. Identification afterwards is focused on the sole set of these target peptides [149, 151–153]. The throughput of the process chain is then not determined by MS/MS identification, but by MS quantification (which could still involve MS/MS, compare Fig. 2).

## 3 Applications of Peptidomics

### *3.1 Studies of Peptide Hormones, Neuropeptides, and Other Bioactive Peptides*

Many of the earliest peptidomics studies were focused on peptide hormones and neuropeptides, and each of these "peptidomic schools" started with their own specific focus and technology. They applied the new technology to very different sources, like insects [154, 155], frog skin secretions [156], vertebrate tissues [124, 157] or cancer cell lines [158] and body fluids like blood plasma and cerebrospinal fluid [4]. The approaches moreover differ being targeted for specific analytes or biological functions or displaying completely untargeted analyses which are then driven by new analytical technologies. Nevertheless, it has persisted a small niche compared to proteomics [10]. Studies of neuropeptides are such an important part of peptidomics that neuropeptidomics became the most prominent subdivision introduced in 2003 [159] which was soon adopted by many others (reviewed by [160–164]). About a quarter of the chapters in this book thus address current methods in this specific field, namely chapters 2, 5, 11, 13, 16–18 and 25.

Main targets for biological functions next to hormones and neuropeptides include the already mentioned antimicrobial peptides as well as peptide animal toxins which occur in various species, for example in many insects. Honeybee [165, 166], ants [167, 168], and other anthropods [169–173] are examples (*see* also chapters 22 and 24 in this book). Furthermore, frog venoms contain such peptides [174], and partly snake venoms, too (*see* [175–177] and chapter 23 in this volume). Marine species like conus snails [178, 179] or sea anemones [180, 181] also secrete peptide toxins.

### 3.2 Biochemical Assays for the Discovery of Bioactive Peptides

Peptidomic techniques have resulted in numerous deeper insights into biological processes, much like genomics and proteomics have revolutionized modern biochemical science. Disulfide bonds are an especially important PTM that among other reasons stabilize three-dimensional structures of linear amino acid chains [182]. Although the presence of several cysteines in small peptides is statistically unlikely, many bioactive peptides contain several disulfide bridges to prevent them from enzymatic degradation, whereas free cysteines are usually very seldom. There are many examples, for instance several peptide hormones (as for insulin [29], endothelin [183], family of natriuretic peptides [184]), CC-chemokines [185, 186], and antimicrobial peptides (*see* [187–190] and chapters 21 and 26 in this book) as well as most animal toxins (*see* [175, 179, 191] and chapters 22 and 24 in this book) and moreover plant defense peptides (*see* Subheading 4.5). An assay using chemical modification and subsequent liquid chromatography was developed to detect cysteine-rich peptides in tissue extracts before the appearance of peptidomics [192]. In a first attempt using peptidomics technology to screen biological fluids for such peptides, labeled cysteines were selectively detected by differential MALDI-MS to search for pairs of cysteines [105]. Using this approach, it was possible to identify several new truncated variants of a known beta-defensin in blood filtrate as well as in urine [104]. Moreover, a new peptide with four disulfide bridges within only 25 amino acids was discovered and named liver-expressed antimicrobial peptide [193]. However, this also pinpoints a disadvantage of the approach. Although a bioactivity was expected right after sequence analysis, it took years of additional research in different groups to elucidate the major role of this peptide in iron metabolism, now called hepcidin [194]. Assay developments to quantify this biomarker, interestingly by LC-MS, are still ongoing [195, 196]. Recently, further approaches have been presented to determine cysteine-rich peptides in animal venoms ([168, 174], and chapter 22 in this book), rat brain [197], as well as plants (*see* Subheading 4.5).

As mentioned above, Mutt and Tatemoto were first with a chemical determination of a C-terminal amide group in peptide hormones instead of using a bioassay [33]. In a current peptidomic study, 19 C-terminally amidated peptides were identified from thyroid carcinoma cells, including novel bioactive peptides derived from the neurosecretory protein named VGF [198]. A more recently published approach combined an enzymatically catalyzed cleavage of a C-terminal glycine specifically resulting in a truncated and amidated C-terminus with a peptidomic analysis of the coexisting α-amidated peptides and their precursors [142]. In a differential display by LC-MS the peptides exhibit similar RP-HPLC properties and MS/MS fragmentation patterns, but a mass difference of 58 Da. Thirteen α-amidated peptides in a mouse pituitary tumor cell line could thus be identified by MS/MS.

Another elegant solution was to purify biologically meaningful peptides by enriching with an affinity column that specifically bound neuropeptide processing intermediates ([73, 91], and chapter 13 in this book). Most neuropeptides and peptide hormones are released from precursor proteins by endopeptidases cleaving at sites containing Lys and/or Arg residues followed by exopeptidase removal of these amino acids from the C-terminus of the processing intermediates [29, 115, 199, 200]. In mice that lack the active form of the major exopeptidase (carboxypeptidase E), the processing intermediates accumulate and could be purified by affinity chromatography. Peptidomic analysis of the purified peptides resulted in the identification of many known neuropeptides as well as several novel peptides from a precursor named proSAAS [73, 91]. However, as with other approaches given above, the finding of novel peptides in mouse brain did not provide direct clues as to the function, and only after extensive additional research has it been determined that some of the proSAAS-derived peptides play a role in feeding/body weight regulation and reward pathways [201, 202]. As an alternative, Sasaki et al. used a brief exocytotic stimulus on cultured endocrine cells which resulted in successful identification of many peptide hormones ([203] and chapter 3 in this book), later combined with a calcium assay [204] or applied to primary cultured rat cardiac fibroblasts [118]. Further approaches using the combination of peptidomic technology with the interaction of specific proteases have revealed intracellular peptides released by proteasomes or unknown pathways and preferred processing sites of peptidases [205]. The latter has been demonstrated by screening a neuropeptidome to characterize the substrate specificity of an extracellular matrix-bound metallocarboxypeptidase [206]; for applications related to metalloendopeptidases see chapter 6 in this book. Similar approaches were used in a comprehensive peptidomics analysis of dipeptidyl-peptidase 4 activity in rat plasma [207] or in kidneys from mice lacking this enzyme [208].

### 3.3 Peptide Biomarker Discovery

A biomarker is defined as "a defined characteristic that is measured as an indicator of normal biological processes, pathogenic processes, or responses to an exposure or intervention, including therapeutic interventions" [209]. Biomarker changes should be detectable in tissues and extracellular fluids, which represent the major link between cells, tissues, and organs of an organism. The peptidome as a huge source of temporary processing products is of special importance in this respect (compare Table 1). Prominent examples for peptide biomarkers are B-type natriuretic peptide (BNP) and N-terminal pro-BNP (NT-proBNP) which are used as gold standard biomarkers in determining heart failure diagnosis and prognosis [184].

Very early successful peptidomic biomarker discovery was applied on conditioned media from cancer cell lines [158, 210, 211].

In parallel, other academic peptide biomarker projects were facilitated by the introduction of a chip-based detection in a rather simple MALDI-MS system (SELDI: surface-enhanced laser desorption/ionization) by Ciphergen (USA), being applied on serum samples from cancer patients [212–214]. However, many promising results [215] soon became more and more questionable as technical issues started to compromise the claims. The main difficulties were in reproducibility [216, 217] and, typical for non-tryptic peptides, lack of identifications [218], thus publishing only peptide mass patterns [214, 215]. Both obstacles were difficult to overcome with these early all-in-one systems. By optimizing the sample preparation protocols of MALDI-MS towards low variances it later became possible to better compare the amounts of peptides in complex samples, including appropriate statistical analyses (*see*, e.g., [219] and chapter 12 in this book).

Body fluids are the most important specimens in medical research and diagnostics. Peptidomics seemed to be a perfect tool for the analysis of human body fluids [6] with proper relative or even absolute quantification. Of particular interest for diagnostic purposes are readily accessible clinical sample sources such as blood plasma and serum [108, 215, 220–223], urine (*see* [224–227] and chapter 20 in this volume), cerebrospinal fluid [228–231], and to a lesser extent lacrimal fluid (tears) (*see* [232] and chapter 4 in this book) or saliva (*see* [188] as well as chapter 19 in this book) and sweat [119]. Unfortunately, many biomarker candidates have been published from peptidomic as well as proteomic studies which could not be validated further [216, 233, 234]. One bitter example is the search for biomarkers associated with Alzheimer's disease which has been extensively investigated, but although several biomarker hits have been consistently identified from different groups over the years [218, 228, 235–238], no clinically applicable biomarker could be validated so far. Technological variances for non-targeted discovery approaches should be below about 10% to expect a good chance with sets of hundreds of well-documented clinical samples. Very often, the most crucial factor is not the detection technology, but the kind, quality, and quantity of the clinical samples under investigation as well as study design [239, 240]. The search for cancer biomarkers in blood specimens has been an aim for many years, with many interesting results, but up to now no valid clinical biomarkers [216, 222, 234].

With the beginning of peptidomic analysis, great expectations drove the field of biomarker discovery. Hannover-based BioVisioN had been the frontrunner for this [4, 239], relaunched as PxBioVisioN in 2006, after lack of financial resources. Several followers were active in parallel, e.g., IriDM in Belgium (www.iridm.com) or Ciphergen Biosystems in the USA (acquired by BioRad in 2003) and GeneProt

in Switzerland (ceased operations in 2005). They all struggled with the general problem that identified peptides and peptidomics technology could not easily be transformed into sustainable projects or products (similar to most proteomics companies). Very few companies still offer peptidomics-based services today, such as PxBioVisioN (Hannover, Germany; www.pxbiovision.com) and some proteomics service providers in Sweden (www.denator.com, Uppsala) and in the USA (www.creative-proteomics.com, Shirley, NY).

## 4 Further Technological Developments and Applications

More than 15 years have passed by after the initial launch of peptidomics and it survived until today as one of many "-omics" technologies (*see* Table 1). Additionally, several technological developments and a rising level of awareness have led to further opportunities and new research groups in the field.

### 4.1 Current Separation Technology

Principally, separation and extraction processes are mostly well established, however, being permanently under optimization. The basic methods for peptide extraction, separation, and sample preparation stem from the times of early protein science around the decades of 1950–1970, dealing mostly with enzymes. Nevertheless, several developments in analytical technologies, predominantly developed with proteomics in mind, also allow for new methods and procedures in peptidomics. The use of LC-MS has become a standard tool in proteomics as well as in peptidomics. Column diameters have been scaled down from 1 mm [49] to capillary LC columns with 75 μm in diameter and as a result flow rates and sample amount have decreased. With new chromatographic media of two micrometer or less in diameter and other developments, it became possible to perform separations with pressures above 1000 bar (UPLC) instead of around 100 bar (HPLC) at the early days 15 years ago. The separation times with comparable resolution are thus much shorter now [241], typically used to increase throughput, and are meanwhile the standard LC technology for peptide separations (*see* for example chapter 18 in this book).

An interesting alternative to the use of chromatography is the coupling of capillary electrophoresis (CE) to MS. Although this has been tried early for peptides, several practical obstacles had to be overcome [242, 243] to match for example the salt load in the electrophoretic system with an electrospray source. A few specialized groups have optimized the use of CE-MS for clinical samples of body fluids [244, 225]. Two very recent chapters in the same series like this volume give insight into the necessary protocols [245, 246].

### 4.2 Current Mass Spectrometric Technology

MS has undergone a tremendous development since the breakthroughs for biopolymer ionization, which is eminent regarding its hardware, software, or applications. All important parameters have been significantly improved (Table 2). Whatsoever, the use of the two ionization principles ESI and MALDI has been a constant to result in high-quality peptide analysis for almost 30 years now. Several new or hybrid *m/z*-analyzers have been developed to improve resolution and mass accuracy, especially for LC-ESI-MS/MS, delivering separation, quantification, and identification at the same time [126, 247]. This allows for many more samples to be measured without manual intervention during the whole run or for faster analyses. MS is nowadays mainly used in the hand of groups without substantial own method development activities, but with deep knowledge about biological context. Most of the time of a researcher in peptidomics (and proteomics) is absorbed with data interpretation and statistical analyses and validation of the intermediate results. LC-ESI-MS also has become the main tool for quantification, more and more replacing traditional approaches utilizing immunoassays that can lack sufficient specificity (e.g., [150, 195]). The necessary standardization is achieved by introduction of several isotopic labeling methods started with isotope-coded affinity tags in the year 2002 [248]. Robust quantitative mass spectrometric approaches [249, 250] and data mining algorithms are now that important and have grown so broadly [251] that several chapters in this volume introduce concepts, methods, and technical challenges thereof (*see* chapters 8–12 in part II of this volume).

Further development of the MS/MS process includes the development of new fragmentation processes to get other characteristics than in CID-MS/MS to better cover sequence information. Best complementing information delivers electron transfer dissociation (ETD) transferring a "soft" electron from singly charged anthracene anions to multiply protonated peptides to induce fragmentation [252] and not by introducing vibrational energy as in CID. Another option is the use of higher energy collisional dissociation (HCD) [253]. The application of both ETD and HCD allow for easier or de novo sequence identification of large peptides [254–256]. ETD and HCD were successfully applied for peptidomic experiments (*see* for example [132, 143, 257]). A further jump in quality of MS/MS data was introduced by the development of the orbitrap mass analyzer, most often used after chromatographic separation [258]. Its substantially increased resolution as well as mass accuracy combined with unprecedented speed allow for the rapid identification of thousands of peptides [226] or de novo sequencing analysis ([133] and chapters 4, 16, and 17 in this book). The fragmentation of ions in MALDI-MS relies on post-source decay (PSD) originating from the high energy transferred by the pulsed laser, thus generating further fragment ions not

seen in CID [259]. Instruments capable of tandem MS after MALDI need a second mass analyzer, currently often solved in a so-called TOF-TOF geometry. This MS/MS process includes high-energy fragmentations [260] with further types of fragments than in low-energy CID-MS/MS. A special application of this is the direct sequence analysis of disulfide-bonded peptides [261]. However, a general problem which is still not overcome is the almost exclusive generation of singly charged ions in the MALDI process. It makes the interpretation very easy, but leads to a lot of neutral fragments that cannot be analyzed in a mass spectrometer.

A further possibility based on MALDI-MS is the discovery of biomarkers by mapping the spatial distribution of a diverse range of molecules in tissue samples. It has been developed 20 years ago [262], but needed extensive technological optimizations to get into broader application. An early review summarized the applicability for compounds in a mass range from 1 to over 50 kDa, thus including all but the smallest peptides [263]. As the identification process with singly charged ions from the MALDI process is so difficult, peptidomic applications are rare, despite the good detectability of peptides and their respective anatomical localization. It was demonstrated for the mapping of insulin contained in an islet in a section of rat pancreas as well as peptide hormones in a small area of a section of rat pituitary [262]. Further developments of the analytical process led to enhanced sensitivity [264], automation of tissue preparation [265], and rapid acquisition [266]. Although the method of MALDI-MS imaging (MALDI-MSI) is now used for several years, only a few further reports demonstrate applicability in classical peptidomics (*see* [267–270] and chapter 17 in this volume). It seems that biomarker discovery with MALDI-MS imaging even with included option of MS/MS data is much better suited for smaller metabolites, drugs, and lipids [271]. Nonetheless, a few very recent papers may give rise to new applications of this method. One describes the successful determination of the spatial distribution of a plant peptidome [272]. Another application involves a combination of in situ peptidase histochemistry with MALDI-MS imaging to specifically analyze the conversion of neuropeptides in native brain tissue sections [273]. Experimental protocols for this method are given in the chapter 17 and also in another chapter of a recent volume in this series, focusing on MALDI-MSI of peptides in cryoconserved and formalin-fixed paraffin-embedded tissue [265].

### 4.3 Current Developments in Sequencing and Data Processing

Contrary to the field of next-generation sequencing for DNA/RNA, peptide sequencing has not basically changed its methodology over the past 20 years, after superseding Edman sequencing by MS/MS. Tandem MS is still its basis [247], however, with an increased automation and thus throughput that is incredible compared to the beginnings of peptidomics in the year 2001. This in principle allows

for a fast analysis of complete proteomes and peptidomes searching for concurrently and coordinately active biological processes. The idea for such a comprehensive statistical analysis has been developed earlier [274], but measurement times (and availability of a sufficient number of well-chosen biological or clinical samples) did not allow for a routine use. It can be expected that the understanding of biological context will rise substantially especially in peptidomics of peptide hormones and proteolytic processing of those (*see* also chapter 5 in this volume). These processes are extremely correlated and lead to complex networks [163] of many peptides simultaneously regulated by a set of proteolytic enzymes [117].

Another important area of optimization is data processing. Automated MS/MS interpretation software provides a reasonable number of identifications for peptides without any modifications from protein sequence databases, although the scoring parameters are largely based on proteomic, and not peptidomic, applications [127, 128]. Algorithms also exist for de novo sequencing [159, 275, 276] or other bioinformatics means (e.g., [254, 277, 278]), but further improvements are needed to allow for more and unambiguous identifications of peptides in organisms not listed in databases. On the other hand, next-generation sequencing of DNA/RNA provides databases that allow for more and more indirectly determined protein sequence data that de novo sequencing of peptides might soon become unnecessary [133, 254, 275, 279, 280]. Nonetheless, the localization of posttranslational modifications will stay an unresolved problem for the interpretation of MS/MS data [135]. Despite their principal occurrence in databases, these annotated entries cannot be easily searched for by existing software. This is mostly no problem in typical bottom-up proteomic studies as it may just lead to one or a few missing peptides out of a multitude generated by trypsin. On the contrary, in peptidomics there are just single peptides to be identified (*see* [10, 128]) which do not contain tryptic cleavage sites. This leads to higher failure rates in identifications [277]. Sometimes the whole information might be lost if two or more modifications are present, especially if MS/MS fragment spectra are poor or even additional mass spectrometric adducts (for example with sodium ions) are formed. This can be overcome by a combination of MS/MS fragmentations, bioinformatics tools, and proper interpretation (*see* [132, 133, 269] and chapter 4 in this book).

Further need for reliable data processing stems from quantitative peptidomic studies. As many different methods are applied, data interpretation cannot always be solved by software. Standardizations might improve the implementation of further algorithms. As most groups use proteomic technologies, these will determine this field. Generally, there is a further need for standardized and automated data analysis that is usually solved by the

peptidomics groups themselves (*see* overview in chapter 8 about this). In parallel to the enormous amount of data generated in quantitative LC-MS/MS experiments, transcriptomic data may complement these to achieve a more holistic view in a systems biology manner [281–284]. It can be expected that further adjustments, application, and implementation of statistical algorithms, data mining, and visualization will be of importance for many more years. This is especially true for future projects aiming at peptide biomarkers. One promising example is a recent study of CSF peptides of this year boosting the number of known CSF peptides by a factor of 10 to around 18,000, including several candidates potentially linked to neurodegeneration [231].

### 4.4 Food Peptidomics

Approaches of so-called food or nutritional peptidomics have been established using peptidomic and related technologies in recent years [285, 286]. Milk with its high protein and peptide content, being body fluid and food at the same time and rather easily available, became the prior source for intense food peptidomics studies (*see* [287–289] as well as the chapter 15 in this volume). The same is equally possible for all other dairy products which contain lots of proteolytic fragments of milk proteins [290]. Food processing by fermentation is also a special focus of peptidomic studies of other edibles, e.g., cheese [291, 292] and ham [293, 294]. This kind of proteolysis mostly is rather unspecific as many microorganisms are involved with different enzymatic activities leading to large analytical sets. Small peptides (2–6 amino acids) sometimes represent by far the largest category [295]. However, these may also account for important properties of the foodstuff. For example, several peptides in cheese and other dairy products are responsible for taste as many appear to be bitter [296]. On the other hand, the proteolytic generation of peptides may cause positive physiological impact as was shown years ago that bioactive peptides from milk proteins are liberated in breast-fed newborns by pepsin processing [297, 298]. Current research and marketing activities address this field of bioactive food peptides and their health effects [286, 287].

Wine, beer, and a number of nonalcoholic beverages are other sources of interest for proteolytically processed proteins after fermentation. Subsequent peptides may originate as fragments from proteins of the plant sources used. For beer, it is well known for example that rather big (several kDa) hydrophobic peptides with specific structures importantly interfere with stability of its foam [299] which is a very important feature for consumer satisfaction. Another study involving the analysis of wheat beer by LC-ESI-MS/MS characterized 167 peptides belonging to 44 proteins; a limited number of them turned out to be epitopes potentially triggering celiac disease [300]. CE-MS has also been applied successfully to peptides in beverages [301] and

is used as part of other omics-driven activities (compare Table 1) concerned with food under the new term foodomics [302]. Another specific application covers for example food authentication of processed meat products by analysis of peptides [293, 302, 303].

### 4.5 Plant Peptidomics

Next to their use in food, plants are well known as an interesting source for antimicrobial or toxic peptides such as purothionins from barley and viscotoxins from mistletoe (*see* [190, 304] and references therein). Investigation of plant peptides stating the use of peptidomics technology took until first publications about the *arabidopsis* peptidome in 2008 [305, 306]. Recent evidences of plant peptides functioning in the responses of plants to diverse environmental influences, such as stress and infection by pathogens, suggest their potential use in agrochemical and pharmaceutical applications [307]. There is a huge variety of bioactive medium-sized peptides from plants containing several disulfide bridges to resist enzymatic degradation. These include cyclotides, liberated from precursor proteins at specific processing sites ([308, 309] and chapter 26 in this book). Furthermore, moss has been extensively analyzed to link peptide occurrence with plant metabolism ([310] and chapter 27 in this book).

### 4.6 Immunopeptidomics

The recent term, comprising all peptides presented at the cell surface by class I and class II major histocompatibility complexes, is immunopeptidome. It has already been noted that mass spectrometric studies of this have been performed long before the term peptidomics was used [83, 85, 311] as the rather small peptides were interesting targets for MS/MS (*see* above). The term "immunopeptidomics" was originally introduced by a group from the company Celera Genomics [312] upon examination of predicted immunopeptidomes of a number of vertebrates and nonvertebrates after publishing the human genome. A few papers adopted the term, but it had disappeared during 2007–2009. It then reappeared in a different meaning within a comment to a paper of Bassani-Sternberg et al. [313] who were using MS to comprehensively analyze very large peptidomes of soluble human leukocyte antigens and thus first applied peptidomic technologies to this field. This has rapidly developed into a very active research area and has led to more than 45 follow-up papers, including two comprehensive reviews of early contributors [314, 315] and several studies already applying this technology to clinical samples [316–318]. Chapter 14 gives detailed protocols for this type of immunopeptidomics.

### 4.7 Concluding Remarks About the State of Peptidomics

Compared to proteomics, peptidomics is a rather small field governed by a number of groups with highly specialized experts. Its implementation has mainly been driven by technological innova-

tions that comprise analytical and biochemical knowledge from more than half a century. This has changed now as applications are primarily on focus. Peptidomics today is a general toolbox allowing the extraction, separation, quantification, and identification of many peptides of interest in principally any biological or technologically processed source. The analytical technologies used are highly automated and principally available to any biologically driven scientist interested in peptide research. However, proper results still have to rely on specific experimental protocols which are typically spread over many papers or even not mentioned in sufficient detail for nonexperts. It is hoped that the number of groups encouraged and enabled and thus the amount of impressive new data in the peptide universe will further develop with the diversity of new information covered in this volume. An outlook to the scope of these activities is given in the final chapter 28.

## Acknowledgments

I am more than grateful for all the extremely valuable comments by Lloyd Fricker that greatly improved my manuscript. Moreover, I would like to thank Manfred Raida for his substantial feedback after critical reading of the manuscript and Peter Verhaert for several important comments.

### References

1. Desiderio DM, Yamada S, Tanzer FS et al (1981) High-performance liquid chromatographic and field desorption mass spectrometric measurement of picomole amounts of endogenous neuropeptides in biologic tissue. J Chromatogr 217:437–452
2. Verhaert P, Uttenweiler-Joseph S, de Vries M et al (2001) Matrix-assisted laser desorption/ionization quadrupole time-of-flight mass spectrometry: an elegant tool for peptidomics. Proteomics 1(1):118–131. https://doi.org/10.1002/1615-9861(200101)1:1<118::AID-PROT118>3.0.CO;2-1
3. Clynen E, Baggerman G, Veelaert D et al (2001) Peptidomics of the pars intercerebralis-corpus cardiacum complex of the migratory locust, Locusta migratoria. Eur J Biochem 268(7):1929–1939
4. Schulz-Knappe P, Zucht HD, Heine G et al (2001) Peptidomics: the comprehensive analysis of peptides in complex biological mixtures. Comb Chem High Throughput Screen 4(2):207–217
5. Minamino N (2001) Peptidome: the fact-database for endogenous peptides. Tanpakushitsu Kakusan Koso 46(11 Suppl):1510–1517
6. Schrader M, Schulz-Knappe P (2001) Peptidomics technologies for human body fluids. Trends Biotechnol 19(10 Suppl):60. https://doi.org/10.1016/S0167-7799(01)01800-5
7. Bergquist J, Ekman R (2001) Future aspects of psychoneuroimmunology—lymphocyte peptides reflecting psychiatric disorders studied by mass spectrometry. Arch Physiol Biochem 109(4):369–371. https://doi.org/10.1076/apab.109.4.369.4241
8. Verhaert PDEM, Pinkse MWH, Prieto-Conaway MC et al (2008) A short history of insect (neuro)peptidomics—a personal story of the birth and youth of an excellent model for studying peptidome biology. In: Soloviev M, Shaw C, Andren P (eds) Peptidomics: methods and applications. Wiley, Hoboken, pp 25–54
9. Schrader M, Schulz-Knappe P (2000) Peptidomics for human body fluids by combination of chromatography and mass spectrometry. J Biomol Tech 11(1):27–60

10. Schrader M, Schulz-Knappe P, Fricker LD (2014) Historical perspective of peptidomics. EuPA Open Proteom 3:171–182. https://doi.org/10.1016/j.euprot.2014.02.014
11. Fenn JB, Mann M, Meng CK et al (1989) Electrospray ionization for mass spectrometry of large biomolecules. Science 246(4926):64–71
12. Karas M, Hillenkamp F (1988) Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. Anal Chem 60(20):2299–2301
13. Edman P (1959) Chemistry of amino acids and peptides. Annu Rev Biochem 28:69–96. https://doi.org/10.1146/annurev.bi.28.070159.000441
14. Aebersold R, Mann M (2003) Mass spectrometry-based proteomics. Nature 422(6928):198–207. https://doi.org/10.1038/nature01511
15. John H, Staendker L (eds) (2004) Peptide separation and analysis. J Chromatogr B Biomed Sci Appl 803(1):1–172
16. Marko-Varga GA (2005) Proteomics and peptidomics. New technology platforms elucidating biology, 1st edn. Comprehensive analytical chemistry, vol 46. Elsevier, Amsterdam
17. Soloviev M, Andrén P, Shaw C (eds) (2008) Peptidomics: methods and applications. Wiley, Hoboken
18. Soloviev M (2010) Peptidomics. Methods and protocols, Methods in molecular biology, vol 615. Humana Press, New York
19. Hook VY-HH (1998) Proteolytic and cellular mechanisms in prohormone and proprotein processing, Molecular biology intelligence unit, vol 2. Landes, Austin
20. Dalbey RE, Sigman DS (2002) Co- and post-translational proteolysis of proteins. In: The enzymes, vol 22, 3rd edn. Academic, San Diego
21. Rehfeld JF, Bundgaard JR (2010) Cellular peptide hormone synthesis and secretory pathways, Results and problems in cell differentiation, vol 50. Springer, Berlin
22. Mbikay M, Seidah NG (2011) Proprotein convertases, Springer protocols, vol 768. Humana Press, New York
23. Favre HA, Powell WH (2014) Nomenclature of organic chemistry: IUPAC recommendations and preferred names 2013. Royal Society of Chemistry, London
24. Moss GP, Smith PAS, Tavernier D (1995) Glossary of class names of organic compounds and reactivity intermediates based on structure (IUPAC Recommendations 1995). Pure Appl Chem 67(8–9). https://doi.org/10.1351/pac199567081307
25. Bayliss WM, Starling EH (1902) The mechanism of pancreatic secretion. J Physiol 28(5):325–353
26. Banting FG (1926) An address on diabetes and insulin: being the Nobel lecture delivered at Stockholm on September 15th, 1925. Can Med Assoc J 16(3):221–232
27. Mutt V, Jorpes JE, Magnusson S (1970) Structure of porcine secretin. The amino acid sequence. Eur J Biochem 15(3):513–519
28. Sanger F (1959) Chemistry of insulin; determination of the structure of insulin opens the way to greater understanding of life processes. Science 129(3359):1340–1344
29. Steiner DF (2011) On the discovery of precursor processing. Methods Mol Biol 768:3–11. https://doi.org/10.1007/978-1-61779-204-5_1
30. Spackman DH, Stein WH, Moore S (1958) Automatic recording apparatus for use in chromatography of amino acids. Anal Chem 30(7):1190–1206. https://doi.org/10.1021/ac60139a006
31. Mutt V (1980) Chemistry, isolation and purification of gastrointestinal hormones. Biochem Soc Trans 8(1):11–14
32. Brown BE, Starratt AN (1975) Isolation of proctolin, a myotropic peptide, from Periplaneta americana. J Insect Physiol 21(11):1879–1881. https://doi.org/10.1016/0022-1910(75)90257-7
33. Tatemoto K, Mutt V (1980) Isolation of two novel candidate hormones using a chemical method for finding naturally occurring polypeptides. Nature 285(5764):417–418
34. Mathe AA, Stenfors C, Brodin E et al (1990) Neuropeptides in brain: effects of microwave irradiation and decapitation. Life Sci 46(4):287–293
35. Fricker LD (2010) Analysis of mouse brain peptides using mass spectrometry-based peptidomics: implications for novel functions ranging from non-classical neuropeptides to microproteins. Mol Biosyst 6(8):1355–1365. https://doi.org/10.1039/c003317k
36. Williams DH, Bradley CV, Santikarn S et al (1982) Fast-atom-bombardment mass spectrometry. A new technique for the determination of molecular weights and amino acid sequences of peptides. Biochem J 201(1):105–117
37. Winkler HU, Beckey HD (1972) Field desorption mass spectrometry of peptides. Biochem Biophys Res Commun 46(2):391–398
38. Macfarlane RD, Torgerson DF (1976) Californium-252 plasma desorption mass spectroscopy. Science 191(4230):920–925
39. Gibson BW, Poulter L, Williams DH (1985) A mass spectrometric assay for novel peptides:

application to Xenopus laevis skin secretions. Peptides 6(Suppl 3):23–27
40. Dass C, Desiderio DM (1987) Fast atom bombardment mass spectrometry analysis of opioid peptides. Anal Biochem 163(1):52–66
41. Fohlman J, Peterson PA, Roepstorff P et al (1985) Comparison of 252californium plasma desorption and fast atom bombardment mass spectrometry for analysis of small peptides. Biomed Mass Spectrom 12(8):380–387
42. Feistner GJ, Højrup P, Evans CJ et al (1989) Mass spectrometric charting of bovine posterior/intermediate pituitary peptides. Proc Natl Acad Sci U S A 86(16):6013–6017
43. Carr SA, Hemling ME, Bean MF et al (1991) Integration of mass spectrometry in analytical biotechnology. Anal Chem 63(24):2802–2824
44. Baratte B, Gras-Masse H, Ricart G et al (1991) Isolation and characterization of authentic Phe-Met-Arg-Phe-NH2 and the novel Phe-Thr-Arg-Phe-NH2 peptide from Nereis diversicolor. Eur J Biochem 198(3):627–633
45. Regnier FE (1983) High-performance liquid chromatography of biopolymers. Science 222(4621):245–252
46. Simpson RJ, Moritz RL, Begg GS et al (1989) Micropreparative procedures for high sensitivity sequencing of peptides and proteins. Anal Biochem 177(2):221–236
47. Sigafoos J, Chestnut WG, Merrill BM et al (1993) Novel peptides from adrenomedullary chromaffin vesicles. J Anat 183(Pt 2):253–264
48. Schulz-Knappe P, Schrader M, Standker L et al (1997) Peptide bank generated by large-scale preparation of circulating human peptides. J Chromatogr A 776(1):125–132
49. Raida M, Schulz-Knappe P, Heine G et al (1999) Liquid chromatography and electrospray mass spectrometric mapping of peptides from human plasma filtrate. J Am Soc Mass Spectrom 10(1):45–54. https://doi.org/10.1016/S1044-0305(98)00117-2
50. Issaq HJ, Chan KC, Blonder J et al (2009) Separation, detection and quantitation of peptides by liquid chromatography and capillary electrochromatography. J Chromatogr A 1216(10):1825–1837. https://doi.org/10.1016/j.chroma.2008.12.052
51. Finoulst I, Pinkse M, van Dongen W et al. (2011) Sample preparation techniques for the untargeted LC-MS-based discovery of peptides in complex biological matrices. J Biomed Biotechnol 20112011, 245291. doi: https://doi.org/10.1155/2011/245291
52. Wilkins MR, Sanchez JC, Williams KL et al (1996) Current challenges and future applications for protein maps and post-translational vector maps in proteome projects. Electrophoresis 17(5):830–838. https://doi.org/10.1002/elps.1150170504
53. Cunningham MJ (2000) Genomics and proteomics: the new millennium of drug discovery and development. J Pharmacol Toxicol Methods 44(1):291–300
54. Kuhn M, Raida M, Adermann K et al (1993) The circulating bioactive form of human guanylin is a high molecular weight peptide (10.3 kDa). FEBS Lett 318(2):205–209
55. Spittaels K, Devreese B, Schoofs L et al (1996) Isolation and identification of a cAMP generating peptide from the flesh fly, Neobellieria bullata (Diptera: Sarcophagidae). Arch Insect Biochem Physiol 31(2):135–147. https://doi.org/10.1002/(SICI)1520-6327(1996)31:2<135:AID-ARCH2>3.0.CO;2-Z
56. Schoofs L, Hamdaoui A, Devreese B et al (1998) The ovary of the desert locust Schistocerca gregaria contains a glycine- and proline-rich peptide that displays sequence similarities with a new class of GPRP proteins from plants. Biochem Biophys Res Commun 243(2):390–394. https://doi.org/10.1006/bbrc.1998.8104
57. Andren PE, Emmett MR, Caprioli RM (1994) Micro-electrospray. Zeptomole/attomole per microliter sensitivity for peptides. J Am Soc Mass Spectrom 5(9):867–869. https://doi.org/10.1016/1044-0305(94)87010-1
58. Wilm M, Mann M (1996) Analytical properties of the nanoelectrospray ion source. Anal Chem 68(1):1–8
59. Edmondson RD, Russell DH (1996) Evaluation of matrix-assisted laser desorption ionization-time-of-flight mass measurement accuracy by using delayed extraction. J Am Soc Mass Spectrom 7(10):995–1001. https://doi.org/10.1016/1044-0305(96)00027-X
60. Cohen SL, Chait BT (1996) Influence of matrix solution conditions on the MALDI-MS analysis of peptides and proteins. Anal Chem 68(1):31–37
61. Garden RW, Moroz LL, Moroz TP et al (1996) Excess salt removal with matrix rinsing. Direct peptide profiling of neurons from marine invertebrates using matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. J Mass Spectrom 31(10):1126–1130. https://doi.org/10.1002/(SICI)1096-9888(199610)31:10<1126:AID-JMS403>3.0.CO;2-7
62. Escoubas P, Célérier M-L, Nakajima T (1997) High-performance liquid chromatography matrix-assisted laser desorption/ionization time-of-flight mass spectrometry peptide

fingerprinting of tarantula venoms in the genusBrachypelma. Chemotaxonomic and biochemical applications. Rapid Commun Mass Spectrom 11(17):1891–1899. https://doi.org/10.1002/(SICI)1097-0231(19971111)11:17<1891:AID-RCM94>3.0.CO;2-X
63. Garden RW, Shippy SA, Li L et al (1998) Proteolytic processing of the Aplysia egg-laying hormone prohormone. Proc Natl Acad Sci U S A 95(7):3972–3977
64. Hsieh S, Dreisewerd K, van der Schors RC et al (1998) Separation and identification of peptides in single neurons by microcolumn liquid chromatography-matrix-assisted laser desorption/ionization time-of-flight mass spectrometry and postsource decay analysis. Anal Chem 70(9):1847–1852
65. Shemyakin MM, Ovchinnikov YA, Kiryushkin AA et al (1966) Mass spectrometric determination of the amino-acid sequence of peptides. Nature 211(5047):361–366. https://doi.org/10.1038/211361a0
66. Agarwal KL, Kenner GW, Sheppard RC (1969) Feline gastrin. An example of peptide sequence analysis by mass spectrometry. J Am Chem Soc 91(11):3096–3097
67. Ling N, Rivier J, Burgus R et al (1973) Direct sequence determination of ovine luteinizing hormone releasing factor by mass spectrometry. Biochemistry 12(26):5305–5310
68. Wipf HK, Irving P, McCamish M et al (1973) Mass spectrometric studies of peptides. V. Determination of amino acid sequences in peptide mixtures by mass spectrometry. J Am Chem Soc 95(10):3369–3375. https://doi.org/10.1021/ja00791a048
69. Biemann K, Martin SA (1987) Mass spectrometric determination of the amino acid sequence of peptides and proteins. Mass Spectrom Rev 6(1):1–75. https://doi.org/10.1002/mas.1280060102
70. Papayannopoulos IA (1995) The interpretation of collision-induced dissociation tandem mass spectra of peptides. Mass Spectrom Rev 14:49–73. https://doi.org/10.1002/mas.1280140104
71. Paizs B, Suhai S (2005) Fragmentation pathways of protonated peptides. Mass Spectrom Rev 24(4):508–548. https://doi.org/10.1002/mas.20024
72. Eriksson U, Andren PE, Caprioli RM et al (1996) Reversed-phase high-performance liquid chromatography combined with tandem mass spectrometry in studies of a substance P-converting enzyme from human cerebrospinal fluid. J Chromatogr A 743(1):213–220
73. Che FY, Yan L, Li H et al (2001) Identification of peptides from brain and pituitary of Cpe(fat)/Cpe(fat) mice. Proc Natl Acad Sci U S A 98(17):9971–9976. https://doi.org/10.1073/pnas.161542198
74. Predel R, Kellner R, Baggerman G et al (2000) Identification of novel perivisceroki-nins from single neurohaemal release sites in insects MS/MS fragmentation complemented by Edman degradation. Eur J Biochem 267(12):3869–3873
75. Carr SA, Annan RS (2001) Overview of peptide and protein analysis by mass spectrometry. Curr Protoc Protein Sci. Chapter 16:Unit 16.1. https://doi.org/10.1002/0471140864.ps1601s04
76. Apweiler R, Bairoch A, Wu CH (2004) Protein sequence databases. Curr Opin Chem Biol 8(1):76–80. https://doi.org/10.1016/j.cbpa.2003.12.004
77. Yates JR III, Eng JK, Clauser KR et al (1996) Search of sequence databases with uninterpreted high-energy collision-induced dissociation spectra of peptides. J Am Soc Mass Spectrom 7(11):1089–1098. https://doi.org/10.1016/S1044-0305(96)00079-7
78. Clauser KR, Baker P, Burlingame AL (1999) Role of accurate mass measurement (+/− 10 ppm) in protein identification strategies employing MS or MS/MS and database searching. Anal Chem 71(14):2871–2882
79. Perkins DN, Pappin DJC, Creasy DM et al (1999) Probability-based protein identification by searching sequence databases using mass spectrometry data. Electrophoresis 20(18):3551–3567. https://doi.org/10.1002/(SICI)1522-2683(19991201)20:18<3551:AID-ELPS3551>3.0.CO;2-2
80. Dawkins BG, Arpino PJ, McLafferty FW (1978) Polypeptide sequencing by liquid chromatography mass spectrometry. Biomed Mass Spectrom 5(1):1–6. https://doi.org/10.1002/bms.1200050102
81. Priddle JD, Rose K, Offord RE (1976) The separation and sequencing of permethylated peptides by mass spectrometry directly coupled to gas-liquid chromatography. Biochem J 157(3):777–780
82. Hunt DF, Yates JR, Shabanowitz J et al (1986) Protein sequencing by tandem mass spectrometry. Proc Natl Acad Sci U S A 83(17):6233–6237
83. Hunt DF, Henderson RA, Shabanowitz J et al (1992) Characterization of peptides bound to the class I MHC molecule HLA-A2.1 by mass spectrometry. Science 255(5049):1261–1263
84. Hunt DF, Michel H, Dickinson TA et al (1992) Peptides presented to the immune system by the murine class II major histocom-

patibility complex molecule I-Ad. Science 256(5065):1817–1820
85. Chicz RM, Urban RG, Lane WS et al (1992) Predominant naturally processed peptides bound to HLA-DR1 are derived from MHC-related molecules and are heterogeneous in size. Nature 358(6389):764–768. https://doi.org/10.1038/358764a0
86. Emmett APE, Caprioli RM (1995) Specific molecular mass detection of endogenously released neuropeptides using in vivo microdialysis/mass spectrometry. J Neurosci Methods 62(1-2):141–147
87. Heine G, Raida M, Forssmann WG (1997) Mapping of peptides and protein fragments in human urine using liquid chromatography-mass spectrometry. J Chromatogr A 776(1):117–124
88. Jurgens M, Schrader M, Raida M et al (1998) Multi-dimensional mapping of human blood peptides by mass spectrometry. J Biomol Tech 9(4):24–30
89. Stark M, Danielsson O, Griffiths WJ et al (2001) Peptide repertoire of human cerebrospinal fluid: novel proteolytic fragments of neuroendocrine proteins. J Chromatogr B Biomed Sci Appl 754(2):357–367
90. Heine G, Zucht HD, Schuhmann MU et al (2002) High-resolution peptide mapping of cerebrospinal fluid: a novel concept for diagnosis and research in central nervous system diseases. J Chromatogr B Analyt Technol Biomed Life Sci 782(1-2):353–361
91. Fricker LD, McKinzie AA, Sun J et al (2000) Identification and characterization of pro-SAAS, a granin-like neuroendocrine peptide precursor that inhibits prohormone processing. J Neurosci 20(2):639–648
92. Licklider LJ, Thoreen CC, Peng J et al (2002) Automation of nanoscale microcapillary liquid chromatography-tandem mass spectrometry with a vented column. Anal Chem 74(13):3076–3083
93. Perkins JR, Smith B, Gallagher RT et al (1993) Application of electrospray mass spectrometry and matrix-assisted laser desorption ionization time-of-flight mass spectrometry for molecular weight assignment of peptides in complex mixtures. J Am Soc Mass Spectrom 4(8):670–684. https://doi.org/10.1016/1044-0305(93)85032-S
94. van Golen FA, Li KW, de Lange RP et al (1995) Mutually exclusive neuronal expression of peptides encoded by the FMRFa gene underlies a differential control of copulation in Lymnaea. J Biol Chem 270(47):28487–28493
95. Jimenez CR, Li KW, Dreisewerd K et al (1998) Direct mass spectrometric peptide profiling and sequencing of single neurons reveals differential peptide patterns in a small neuronal network. Biochemistry 37(7):2070–2076. https://doi.org/10.1021/bi971848b
96. Li L, Garden RW, Sweedler JV (2000) Single-cell MALDI: a new tool for direct peptide profiling. Trends Biotechnol 18(4):151–160
97. Rubakhin SS, Garden RW, Fuller RR et al (2000) Measuring the peptides in individual organelles with mass spectrometry. Nat Biotechnol 18(2):172–175. https://doi.org/10.1038/72622
98. Schrader M, Jurgens M, Hess R et al (1997) Matrix-assisted laser desorption/ionisation mass spectrometry guided purification of human guanylin from blood ultrafiltrate. J Chromatogr A 776(1):139–145
99. Floyd PD, Li L, Moroz TP et al (1999) Characterization of peptides from Aplysia using microbore liquid chromatography with matrix-assisted laser desorption/ionization time-of-flight mass spectrometry guided purification. J Chromatogr A 830(1):105–113
100. Richter R, Schulz-Knappe P, Schrader M et al (1999) Composition of the peptide fraction in human blood plasma: database of circulating human peptides. J Chromatogr B Biomed Sci Appl 726(1-2):25–35
101. Escoubas P, Chamot-Rooke J, Stöcklin R et al (1999) A comparison of matrix-assisted laser desorption/ionization time-of-flight and liquid chromatography electrospray ionization mass spectrometry methods for the analysis of crude tarantula venoms in the Pterinochilus group. Rapid Commun Mass Spectrom 13(18):1861–1868. https://doi.org/10.1002/(SICI)1097-0231(19990930)13:18<1861:AID-RCM730>3.0.CO;2-7
102. Uttenweiler-Joseph S, Moniatte M, Lagueux M et al (1998) Differential display of peptides induced during the immune response of Drosophila: a matrix-assisted laser desorption ionization time-of-flight mass spectrometry study. Proc Natl Acad Sci U S A 95(19):11342–11347
103. Jimenez CR, Li KW, Dreisewerd K et al (1997) Pattern changes of pituitary peptides in rat after salt-loading as detected by means of direct, semiquantitative mass spectrometric profiling. Proc Natl Acad Sci U S A 94(17):9481–9486
104. Zucht HD, Grabowsky J, Schrader M et al (1998) Human beta-defensin-1: a urinary peptide present in variant molecular forms and its putative functional implication. Eur J Med Res 3(7):315–323
105. Neitz S, Jurgens M, Kellmann M et al (2001) Screening for disulfide-rich peptides

in biological sources by carboxyamidomethylation in combination with differential matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. Rapid Commun Mass Spectrom 15(17):1586–1592. https://doi.org/10.1002/rcm.413
106. Baggerman G, Verleyen P, Clynen E et al (2004) Peptidomics. J Chromatogr B Analyt Technol Biomed Life Sci 803(1):3–16. https://doi.org/10.1016/j.jchromb.2003.07.019
107. Skold K, Svensson M, Kaplan A et al (2002) A neuroproteomic approach to targeting neuropeptides in the brain. Proteomics 2(4):447–454. https://doi.org/10.1002/1615-9861(200204)2:4<447:AID-PROT447>3.0.CO;2-A
108. Tammen H, Schulte I, Hess R et al (2005) Peptidomic analysis of human blood specimens: comparison between plasma specimens and serum by differential peptide display. Proteomics 5(13):3414–3422. https://doi.org/10.1002/pmic.200401219
109. Munch J, Standker L, Adermann K et al (2007) Discovery and optimization of a natural HIV-1 entry inhibitor targeting the gp41 fusion peptide. Cell 129(2):263–275. https://doi.org/10.1016/j.cell.2007.02.042
110. Samir P, Link AJ (2011) Analyzing the cryptome: uncovering secret sequences. AAPS J 13(2):152–158. https://doi.org/10.1208/s12248-011-9252-2
111. Osaki T, Sasaki K, Minamino N (2011) Peptidomics-based discovery of an antimicrobial peptide derived from insulin-like growth factor-binding protein 5. J Proteome Res 10(4):1870–1880. https://doi.org/10.1021/pr101114a
112. Guerrero A, Dallas DC, Contreras S et al (2014) Mechanistic peptidomics: factors that dictate specificity in the formation of endogenous peptides in human milk. Mol Cell Proteomics 13(12):3343–3351. https://doi.org/10.1074/mcp.M113.036194
113. Munch J, Standker L, Forssmann W-G et al (2014) Discovery of modulators of HIV-1 infection from the human peptidome. Nat Rev Microbiol 12(10):715–722. https://doi.org/10.1038/nrmicro3312
114. Zirafi O, Kim K-A, Standker L et al (2015) Discovery and characterization of an endogenous CXCR4 antagonist. Cell Rep 11(5):737–747. https://doi.org/10.1016/j.celrep.2015.03.061
115. Rehfeld JF, Goetze JP (2003) The posttranslational phase of gene expression: new possibilities in molecular diagnosis. Curr Mol Med 3(1):25–38
116. Rogers LD, Overall CM (2013) Proteolytic post-translational modification of proteins: proteomic tools and methodology. Mol Cell Proteomics 12(12):3532–3542. https://doi.org/10.1074/mcp.M113.031310
117. Fortelny N, Cox JH, Kappelhoff R et al (2014) Network analyses reveal pervasive functional regulation between proteases in the human protease web. PLoS Biol 12(5):e1001869. https://doi.org/10.1371/journal.pbio.1001869
118. Tsuchiya T, Osaki T, Minamino N et al (2015) Peptidomics for studying limited proteolysis. J Proteome Res 14(11):4921–4931. https://doi.org/10.1021/acs.jproteome.5b00820
119. Yu Y, Prassas I, Muytjens CMJ et al (2017) Proteomic and peptidomic analysis of human sweat with emphasis on proteolysis. J Proteomics 155:40–48. https://doi.org/10.1016/j.jprot.2017.01.005
120. Jimenez CR, Spijker S, de Schipper S et al (2006) Peptidomics of a single identified neuron reveals diversity of multiple neuropeptides with convergent actions on cellular excitability. J Neurosci 26(2):518–529. https://doi.org/10.1523/JNEUROSCI.2566-05.2006
121. Neupert S, Johard HAD, Nassel DR et al (2007) Single-cell peptidomics of drosophila melanogaster neurons identified by Gal4-driven fluorescence. Anal Chem 79(10):3690–3694. https://doi.org/10.1021/ac062411p
122. Neupert S, Rubakhin SS, Sweedler JV (2012) Targeted single-cell microchemical analysis: MS-based peptidomics of individual paraformaldehyde-fixed and immunolabeled neurons. Chem Biol 19(8):1010–1019. https://doi.org/10.1016/j.chembiol.2012.05.023
123. Che F-Y, Fricker LD (2002) Quantitation of neuropeptides in Cpe(fat)/Cpe(fat) mice using differential isotopic tags and mass spectrometry. Anal Chem 74(13):3190–3198
124. Che FY, Eipper BA, Mains RE et al (2003) Quantitative peptidomics of pituitary glands from mice deficient in copper transport. Cell Mol Biol (Noisy-le-Grand) 49(5):713–722
125. Carr S, Aebersold R, Baldwin M et al (2004) The need for guidelines in publication of peptide and protein identification data: Working Group on Publication Guidelines for Peptide and Protein Identification Data. Mol Cell Proteomics 3(6):531–533. https://doi.org/10.1074/mcp.T400006-MCP200
126. Boonen K, Landuyt B, Baggerman G et al (2008) Peptidomics: the integrated approach of MS, hyphenated techniques and bioinformatics for neuropeptide analysis. J Sep Sci 31(3):427–445. https://doi.org/10.1002/jssc.200700450
127. Akhtar MN, Southey BR, Andren PE et al (2012) Evaluation of database search pro-

grams for accurate detection of neuropeptides in tandem mass spectrometry experiments. J Proteome Res 11(12):6044–6055. https://doi.org/10.1021/pr3007123
128. Hayakawa E, Menschaert G, de Bock PJ et al (2013) Improving the identification rate of endogenous peptides using electron transfer dissociation and collision-induced dissociation. J Proteome Res 12(12): 5410–5421. https://doi.org/10.1021/pr400446z
129. Deutsch EW, Overall CM, van Eyk JE et al (2016) Human proteome project mass spectrometry data interpretation guidelines 2.1. J Proteome Res 15(11):3961–3970. https://doi.org/10.1021/acs.jproteome.6b00392
130. Webb-Robertson BJ, Cannon WR, Oehmen CS et al (2010) A support vector machine model for the prediction of proteotypic peptides for accurate mass and time proteomics. Bioinformatics 26(13):1677–1683
131. Samgina TY, Vorontsov EA, Gorshkov VA et al (2014) Mass spectrometric de novo sequencing of natural non-tryptic peptides: comparing peculiarities of collision-induced dissociation (CID) and high energy collision dissociation (HCD). Rapid Commun Mass Spectrom 28(23):2595–2604. https://doi.org/10.1002/rcm.7049
132. Azkargorta M, Soria J, Ojeda C et al (2015) Human basal tear peptidome characterization by CID, HCD, and ETD followed by in silico and in vitro analyses for antimicrobial peptide identification. J Proteome Res 14(6):2649–2658. https://doi.org/10.1021/acs.jproteome.5b00179
133. Samgina TY, Tolpina MD, Trebse P et al (2016) LTQ Orbitrap Velos in routine de novo sequencing of non-tryptic skin peptides from the frog Rana latastei with traditional and reliable manual spectra interpretation. Rapid Commun Mass Spectrom 30(2):265–276
134. Cottrell JS (2011) Protein identification using MS/MS data. J Proteomics 74(10):1842–1851. https://doi.org/10.1016/j.jprot.2011.05.014
135. Kim M-S, Zhong J, Pandey A (2016) Common errors in mass spectrometry-based analysis of post-translational modifications. Proteomics 16(5):700–714. https://doi.org/10.1002/pmic.201500355
136. Badock V, Raida M, Adermann K et al (1998) Distinction between the three disulfide isomers of guanylin 99-115 by low-energy collision-induced dissociation. Rapid Commun Mass Spectrom 12(23):1952–1956. https://doi.org/10.1002/(SICI)1097-0231(19981215)12:23<1952:AID-RCM420>3.0.CO;2-N
137. Samgina TY, Vorontsov EA, Gorshkov VA et al (2011) Novel cysteine tags for the sequencing of non-tryptic disulfide peptides of anurans: ESI-MS study of fragmentation efficiency. J Am Soc Mass Spectrom 22(12):2246–2255. https://doi.org/10.1007/s13361-011-0247-0
138. Lee M, Lee Y, Kang M et al (2011) Disulfide bond cleavage in TEMPO-free radical initiated peptide sequencing mass spectrometry. J Mass Spectrom 46(8):830–839. https://doi.org/10.1002/jms.1955
139. Bhatia S, Kil YJ, Ueberheide B et al (2012) Constrained de novo sequencing of conotoxins. J Proteome Res 11(8):4191–4200. https://doi.org/10.1021/pr300312h
140. Bhattacharyya M, Gupta K, Gowd KH et al (2013) Rapid mass spectrometric determination of disulfide connectivity in peptides and proteins. Mol Biosyst 9(6):1340–1350. https://doi.org/10.1039/c3mb25534d
141. Fricker LD (2015) Limitations of mass spectrometry-based peptidomic approaches. J Am Soc Mass Spectrom 26(12):1981–1991. https://doi.org/10.1007/s13361-015-1231-x
142. An Z, Chen Y, Koomen JM et al (2012) A mass spectrometry-based method to screen for alpha-amidated peptides. Proteomics 12(2):173–182. https://doi.org/10.1002/pmic.201100327
143. Sasaki K, Osaki T, Minamino N (2013) Large-scale identification of endogenous secretory peptides using electron transfer dissociation mass spectrometry. Mol Cell Proteomics 12(3):700–709. https://doi.org/10.1074/mcp.M112.017400
144. Ianzer D, Konno K, Marques-Porto R et al (2004) Identification of five new bradykinin potentiating peptides (BPPs) from Bothrops jararaca crude venom by using electrospray ionization tandem mass spectrometry after a two-step liquid chromatography. Peptides 25(7):1085–1092. https://doi.org/10.1016/j.peptides.2004.04.006
145. Chen T, Scott C, Tang L et al (2005) The structural organization of aurein precursor cDNAs from the skin secretion of the Australian green and golden bell frog, Litoria aurea. Regul Pept 128(1):75–83. https://doi.org/10.1016/j.regpep.2004.12.022
146. Ji Y, Bachschmid MM, Costello CE et al (2016) S- to N-palmitoyl transfer during proteomic sample preparation. J Am Soc Mass Spectrom 27(4):677–685. https://doi.org/10.1007/s13361-015-1319-3
147. John H, Walden M, Schafer S et al (2004) Analytical procedures for quantification of peptides in pharmaceutical research by liquid

chromatography-mass spectrometry. Anal Bioanal Chem 378(4):883–897. https://doi.org/10.1007/s00216-003-2298-y
148. Gardner MS, Rowland MD, Siu AY et al (2009) A comprehensive defensin assay for saliva. Anal Chem 81(2):557–566. https://doi.org/10.1021/ac801609r
149. Che F-Y, Fricker LD (2005) Quantitative peptidomics of mouse pituitary: comparison of different stable isotopic tags. J Mass Spectrom 40(2):238–249. https://doi.org/10.1002/jms.743
150. Mesmin C, Dubois M, Becher F et al (2010) Liquid chromatography/tandem mass spectrometry assay for the absolute quantification of the expected circulating apelin peptides in human plasma. Rapid Commun Mass Spectrom 24(19):2875–2884. https://doi.org/10.1002/rcm.4718
151. Fricker LD, Lim J, Pan H et al (2006) Peptidomics: identification and quantification of endogenous peptides in neuroendocrine tissues. Mass Spectrom Rev 25(2):327–344. https://doi.org/10.1002/mas.20079
152. Berezniuk I, Sironi JJ, Wardman J et al (2013) Quantitative peptidomics of Purkinje cell degeneration mice. PLoS One 8(4):e60981. https://doi.org/10.1371/journal.pone.0060981
153. Verdonck R, de Haes W, Cardoen D et al (2016) Fast and reliable quantitative peptidomics with labelpepmatch. J Proteome Res 15(3):1080–1089. https://doi.org/10.1021/acs.jproteome.5b00845
154. Baggerman G, Cerstiaens A, de Loof A et al (2002) Peptidomics of the larval Drosophila melanogaster central nervous system. J Biol Chem 277(43):40368–40374. https://doi.org/10.1074/jbc.M206257200
155. Predel R, Wegener C, Russell WK et al (2004) Peptidomics of CNS-associated neurohemal systems of adult Drosophila melanogaster: a mass spectrometric survey of peptides from individual flies. J Comp Neurol 474(3):379–392. https://doi.org/10.1002/cne.20145
156. Conlon JM, Kim JB, Johansson A et al (2002) Comparative peptidomics of the endocrine pancreas: islet hormones from the clawed frog Xenopus laevis and the red-bellied newt Cynops pyrrhogaster. J Endocrinol 175(3):769–777
157. Svensson M, Skold K, Svenningsson P et al (2003) Peptidomics-based discovery of novel neuropeptides. J Proteome Res 2(2):213–219
158. Sasaki K, Sato K, Akiyama Y et al (2002) Peptidomics-based approach reveals the secretion of the 29-residue COOH-terminal fragment of the putative tumor suppressor protein DMBT1 from pancreatic adenocarcinoma cell lines. Cancer Res 62(17):4894–4898
159. Huybrechts J, Nusbaum MP, Bosch LV et al (2003) Neuropeptidomic analysis of the brain and thoracic ganglion from the Jonah crab, Cancer borealis. Biochem Biophys Res Commun 308(3):535–544
160. de Haes W, van Sinay E, Detienne G et al (2015) Functional neuropeptidomics in invertebrates. Biochim Biophys Acta 1854(7):812–826. https://doi.org/10.1016/j.bbapap.2014.12.011
161. Romanova EV, Sweedler JV (2015) Peptidomics for the discovery and characterization of neuropeptides and hormones. Trends Pharmacol Sci 36(9):579–586. https://doi.org/10.1016/j.tips.2015.05.009
162. OuYang C, Liang Z, Li L (2015) Mass spectrometric analysis of spatio-temporal dynamics of crustacean neuropeptides. Biochim Biophys Acta 1854(7):798–811. https://doi.org/10.1016/j.bbapap.2014.10.023
163. Hook V, Bandeira N (2015) Neuropeptidomics mass spectrometry reveals signaling networks generated by distinct protease pathways in human systems. J Am Soc Mass Spectrom 26(12):1970–1980. https://doi.org/10.1007/s13361-015-1251-6
164. Lee JE (2016) Neuropeptidomics: mass spectrometry-based identification and quantitation of neuropeptides. Genomics Inform 14(1):12–19. https://doi.org/10.5808/GI.2016.14.1.12
165. van Vaerenbergh M, Cardoen D, Formesyn EM et al (2013) Extending the honey bee venome with the antimicrobial peptide apidaecin and a protein resembling wasp antigen 5. Insect Mol Biol 22(2):199–210. https://doi.org/10.1111/imb.12013
166. Sturm S, Ramesh D, Brockmann A et al (2016) Agatoxin-like peptides in the neuroendocrine system of the honey bee and other insects. J Proteomics 132:77–84. https://doi.org/10.1016/j.jprot.2015.11.021
167. Cologna CT, Cardoso JS, Jourdan E et al (2013) Peptidomic comparison and characterization of the major components of the venom of the giant ant Dinoponera quadriceps collected in four different areas of Brazil. J Proteomics 94:413–422. https://doi.org/10.1016/j.jprot.2013.10.017
168. Aili SR, Touchard A, Koh JMS et al (2016) Comparisons of protein and peptide complexity in poneroid and formicoid ant venoms. J Proteome Res 15(9):3039–3054. https://doi.org/10.1021/acs.jproteome.6b00182
169. Favreau P, Menin L, Michalet S, Perret F, Cheneval O, Stocklin M, Bulet P, Stocklin R (2006) Mass spectrometry strategies for

venom mapping and peptide sequencing from crude venoms: case applications with single arthropod specimen. Toxicon 47(6):676–687
170. Escoubas P, Quinton L, Nicholson GM (2008) Venomics: unravelling the complexity of animal venoms with mass spectrometry. J Mass Spectrom 43(3):279–295. https://doi.org/10.1002/jms.1389
171. Schwartz EF, Mourao CB, Moreira KG et al (2012) Arthropod venoms: a vast arsenal of insecticidal neuropeptides. Biopolymers 98(4):385–405
172. Hakim MA, Yang S, Lai R (2015) Centipede venoms and their components: resources for potential therapeutic applications. Toxins (Basel) 7(11):4832–4851. https://doi.org/10.3390/toxins7114832
173. Kuzmenkov AI, Sachkova MY, Kovalchuk SI et al (2016) Lachesana tarabaevi, an expert in membrane-active toxins. Biochem J 473(16):2495–2506. https://doi.org/10.1042/BCJ20160436
174. Evaristo GPC, Verhaert PDEM, Pinkse MWH (2012) PTM-driven differential peptide display: survey of peptides containing inter/intra-molecular disulfide bridges in frog venoms. J Proteomics 77:215–224. https://doi.org/10.1016/j.jprot.2012.09.001
175. Calvete JJ, Schrader M, Raida M et al (1997) The disulphide bond pattern of bitistatin, a disintegrin isolated from the venom of the viper Bitis arietans. FEBS Lett 416(2):197–202
176. Munawar A, Trusch M, Georgieva D et al (2011) Venom peptide analysis of Vipera ammodytes meridionalis (Viperinae) and Bothrops jararacussu (Crotalinae) demonstrates subfamily-specificity of the peptidome in the family Viperidae. Mol Biosyst 7(12):3298–3307. https://doi.org/10.1039/c1mb05309d
177. Tashima AK, Zelanis A, Kitano ES et al (2012) Peptidomics of three Bothrops snake venoms: insights into the molecular diversification of proteomes and peptidomes. Mol Cell Proteomics 11(11):1245–1262. https://doi.org/10.1074/mcp.M112.019331
178. Dutertre S, Jin A-H, Kaas Q et al (2013) Deep venomics reveals the mechanism for expanded peptide diversity in cone snail venom. Mol Cell Proteomics 12(2):312–329. https://doi.org/10.1074/mcp.M112.021469
179. Rodriguez AM, Dutertre S, Lewis RJ et al (2015) Intraspecific variations in Conus purpurascens injected venom using LC/MALDI-TOF-MS and LC-ESI-TripleTOF-MS. Anal Bioanal Chem 407(20):6105–6116. https://doi.org/10.1007/s00216-015-8787-y
180. Frazao B, Vasconcelos V, Antunes A (2012) Sea anemone (Cnidaria, Anthozoa, Actiniaria) toxins: an overview. Mar Drugs 10(8):1812–1851. https://doi.org/10.3390/md10081812
181. Rodriguez AA, Standker L, Zaharenko AJ et al (2012) Combining multidimensional liquid chromatography and MALDI-TOF-MS for the fingerprint analysis of secreted peptides from the unexplored sea anemone species Phymanthus crucifer. J Chromatogr B Analyt Technol Biomed Life Sci 903:30–39. https://doi.org/10.1016/j.jchromb.2012.06.034
182. Wong JWH, Ho SYW, Hogg PJ (2011) Disulfide bond acquisition through eukaryotic protein evolution. Mol Biol Evol 28(1):327–334. https://doi.org/10.1093/molbev/msq194
183. Vignon-Zellweger N, Heiden S, Miyauchi T et al (2012) Endothelin and endothelin receptors in the renal and cardiovascular systems. Life Sci 91(13–14):490–500. https://doi.org/10.1016/j.lfs.2012.03.026
184. Gaggin HK, Januzzi JL Jr (2014) Natriuretic peptides in heart failure and acute coronary syndrome. Clin Lab Med 34(1):43–58., vi. https://doi.org/10.1016/j.cll.2013.11.007
185. Schulz-Knappe P, Magert HJ, Dewald B et al (1996) HCC-1, a novel chemokine from human plasma. J Exp Med 183(1):295–299
186. Thiele S, Rosenkilde MM (2014) Interaction of chemokines with their receptors—from initial chemokine binding to receptor activating steps. Curr Med Chem 21(31):3594–3614
187. Bulet P, Stocklin R, Menin L (2004) Antimicrobial peptides: from invertebrates to vertebrates. Immunol Rev 198:169–184
188. Trindade F, Amado F, Pinto da Costa J et al (2015) Salivary peptidomic as a tool to disclose new potential antimicrobial peptides. J Proteomics 115:49–57. https://doi.org/10.1016/j.jprot.2014.12.004
189. Conlon JM (2015) Host-defense peptides of the skin with therapeutic potential: From hagfish to human. Peptides 67:29–38. https://doi.org/10.1016/j.peptides.2015.03.005
190. Tam JP, Wang S, Wong KH et al (2015) Antimicrobial Peptides from Plants. Pharmaceuticals (Basel) 8(4):711–757. https://doi.org/10.3390/ph8040711
191. Lavergne V, Alewood PF, Mobli M et al (2015) The structural universe of disulfide-rich venom peptides. In: King GF (ed) Venoms to drugs. Venom as a source for the development of human therapeutics. Royal Society of Chemistry, Cambridge, pp 37–79
192. Sillard R, Jornvall H, Carlquist M et al (1993) Chemical assay for cyst(e)ine-rich peptides detects a novel intestinal peptide ZF-1, homologous to a single zinc-finger motif. Eur J Biochem 211(1–2):377–380

193. Krause A, Neitz S, Magert HJ et al (2000) LEAP-1, a novel highly disulfide-bonded human peptide, exhibits antimicrobial activity. FEBS Lett 480(2–3):147–150
194. Ganz T (2011) Hepcidin and iron regulation, 10 years later. Blood 117(17):4425–4433. https://doi.org/10.1182/blood-2011-01-258467
195. Lefebvre T, Dessendier N, Houamel D et al (2015) LC-MS/MS method for hepcidin-25 measurement in human and mouse serum: clinical and research implications in iron disorders. Clin Chem Lab Med 53(10):1557–1567. https://doi.org/10.1515/cclm-2014-1093
196. Swensen AC, Finnell JG, Matias C et al (2016) Whole blood and urine bioactive Hepcidin-25 determination using liquid chromatography mass spectrometry. Anal Biochem 517:23–30. https://doi.org/10.1016/j.ab.2016.10.023
197. Yu X, Khani A, Ye X et al (2015) High-efficiency recognition and identification of disulfide bonded peptides in rat neuropeptidome using targeted electron transfer dissociation tandem mass spectrometry. Anal Chem 87(23):11646–11651. https://doi.org/10.1021/ac504872z
198. Yamaguchi H, Sasaki K, Satomi Y et al (2007) Peptidomic identification and biological validation of neuroendocrine regulatory peptide-1 and -2. J Biol Chem 282(36):26354–26360. https://doi.org/10.1074/jbc.M701665200
199. Hook V, Funkelstein L, Lu D et al (2008) Proteases for processing proneuropeptides into peptide neurotransmitters and hormones. Annu Rev Pharmacol Toxicol 48:393–423. https://doi.org/10.1146/annurev.pharmtox.48.113006.094812
200. Sapio MR, Fricker LD (2014) Carboxypeptidases in disease: insights from peptidomic studies. Proteomics Clin Appl 8(5-6):327–337. https://doi.org/10.1002/prca.201300090
201. Morgan DJ, Wei S, Gomes I et al (2010) The propeptide precursor proSAAS is involved in fetal neuropeptide processing and body weight regulation. J Neurochem 113(5):1275–1284. https://doi.org/10.1111/j.1471-4159.2010.06706.x
202. Wardman JH, Berezniuk I, Di S et al (2011) ProSAAS-derived peptides are colocalized with neuropeptide Y and function as neuropeptides in the regulation of food intake. PLoS One 6(12):e28152. https://doi.org/10.1371/journal.pone.0028152
203. Sasaki K, Satomi Y, Takao T et al (2009) Snapshot peptidomics of the regulated secretory pathway. Mol Cell Proteomics 8(7):1638–1647. https://doi.org/10.1074/mcp.M900044-MCP200
204. Sasaki K, Takahashi N, Satoh M et al (2010) A peptidomics strategy for discovering endogenous bioactive peptides. J Proteome Res 9(10):5047–5052. https://doi.org/10.1021/pr1003455
205. Ferro ES, Rioli V, Castro LM et al (2014) Intracellular peptides: from discovery to function. EuPA Open Proteom 3:143–151. https://doi.org/10.1016/j.euprot.2014.02.009
206. Lyons PJ, Fricker LD (2010) Substrate specificity of human carboxypeptidase A6. J Biol Chem 285(49):38234–38242. https://doi.org/10.1074/jbc.M110.158626
207. Jost MM, Lamerz J, Tammen H et al (2009) In vivo profiling of DPP4 inhibitors reveals alterations in collagen metabolism and accumulation of an amyloid peptide in rat plasma. Biochem Pharmacol 77(2):228–237. https://doi.org/10.1016/j.bcp.2008.09.032
208. Tinoco AD, Tagore DM, Saghatelian A (2010) Expanding the dipeptidyl peptidase 4-regulated peptidome via an optimized peptidomics platform. J Am Chem Soc 132(11):3819–3830. https://doi.org/10.1021/ja909524e
209. FDA-NIH Biomarker Working Group (2016) BEST (Biomarkers, EndpointS, and other Tools) Resource. https://www.ncbi.nlm.nih.gov/books/NBK338449/
210. Sato K, Sasaki K, Akiyama Y et al (2001) Mass spectrometric high-throughput analysis of serum-free conditioned medium from cancer cell lines. Cancer Lett 170(2):153–159
211. Sato K, Sasaki K, Tsao M-S et al (2002) Peptide differential display of serum-free conditioned medium from cancer cell lines. Cancer Lett 176(2):199–203
212. Petricoin EF, Ardekani AM, Hitt BA et al (2002) Use of proteomic patterns in serum to identify ovarian cancer. Lancet 359(9306):572–577. https://doi.org/10.1016/S0140-6736(02)07746-2
213. Petricoin EF, Liotta LA (2004) SELDI-TOF-based serum proteomic pattern diagnostics for early detection of cancer. Curr Opin Biotechnol 15(1):24–30. https://doi.org/10.1016/j.copbio.2004.01.005
214. Villanueva J, Philip J, Entenberg D et al (2004) Serum peptide profiling by magnetic particle-assisted, automated sample processing and MALDI-TOF mass spectrometry. Anal Chem 76(6):1560–1570. https://doi.org/10.1021/ac0352171
215. Ornstein DK, Rayford W, Fusaro VA et al (2004) Serum proteomic profiling can discriminate prostate cancer from benign pros-

tates in men with total prostate specific antigen levels between 2.5 and 15.0 ng/ml. J Urol 172(4 Pt 1):1302–1305
216. Diamandis EP (2003) Point: Proteomic patterns in biological fluids: do they represent the future of cancer diagnostics? Clin Chem 49(8):1272–1275
217. Albrethsen J (2007) Reproducibility in protein profiling by MALDI-TOF mass spectrometry. Clin Chem 53(5):852–858. https://doi.org/10.1373/clinchem.2006.082644
218. Carrette O, Demalte I, Scherl A et al (2003) A panel of cerebrospinal fluid potential biomarkers for the diagnosis of Alzheimer's disease. Proteomics 3(8):1486–1494. https://doi.org/10.1002/pmic.200300470
219. Duncan MW, Roder H, Hunsucker SW (2008) Quantitative matrix-assisted laser desorption/ionization mass spectrometry. Brief Funct Genomic Proteomic 7(5):355–370. https://doi.org/10.1093/bfgp/eln041
220. Hashiguchi T, Tanaka K, Lee L-J et al (2009) Diagnostic value of serum peptidome analyses for protease activated pathological conditions beyond cancer diagnosis. Med Hypotheses 73(5):760–763. https://doi.org/10.1016/j.mehy.2009.04.026
221. Zhu P, Bowden P, Zhang D et al (2011) Mass spectrometry of peptides and proteins from human blood. Mass Spectrom Rev 30(5):685–732. https://doi.org/10.1002/mas.20291
222. Labots M, Schutte LM, van der Mijn JC et al (2014) Mass spectrometry-based serum and plasma peptidome profiling for prediction of treatment outcome in patients with solid malignancies. Oncologist 19(10):1028–1039. https://doi.org/10.1634/theoncologist.2014-0101
223. Mahboob S, Mohamedali A, Ahn SB et al (2015) Is isolation of comprehensive human plasma peptidomes an achievable quest? J Proteomics 127(Pt B):300–309. https://doi.org/10.1016/j.jprot.2015.05.010
224. Jurgens M, Appel A, Heine G et al (2005) Towards characterization of the human urinary peptidome. Comb Chem High Throughput Screen 8(8):757–765
225. Good DM, Zurbig P, Argiles A et al (2010) Naturally occurring human urinary peptides for use in diagnosis of chronic kidney disease. Mol Cell Proteomics 9(11):2424–2437. https://doi.org/10.1074/mcp.M110.001917
226. Smith CR, Batruch I, Bauca JM et al (2014) Deciphering the peptidome of urine from ovarian cancer patients and healthy controls. Clin Proteomics 11(1):23. https://doi.org/10.1186/1559-0275-11-23
227. Kononikhin AS, Starodubtseva NL, Bugrova AE et al (2016) An untargeted approach for the analysis of the urine peptidome of women with preeclampsia. J Proteomics 149:38–43. https://doi.org/10.1016/j.jprot.2016.04.024
228. Selle H, Lamerz J, Buerger K et al (2005) Identification of novel biomarker candidates by differential peptidomics analysis of cerebrospinal fluid in Alzheimer's disease. Comb Chem High Throughput Screen 8(8):801–806
229. Yuan X, Desiderio DM (2005) Human cerebrospinal fluid peptidomics. J Mass Spectrom 40(2):176–181. https://doi.org/10.1002/jms.737
230. Holtta M, Zetterberg H, Mirgorodskaya E et al (2012) Peptidome analysis of cerebrospinal fluid by LC-MALDI MS. PLoS One 7(8):e42555. https://doi.org/10.1371/journal.pone.0042555
231. Hansson KT, Skillback T, Pernevik E et al (2017) Expanding the cerebrospinal fluid endopeptidome. Proteomics. https://doi.org/10.1002/pmic.201600384
232. Azkargorta M, Soria J, Acera A et al (2016) Human tear proteomics and peptidomics in ophthalmology: toward the translation of proteomic biomarkers into clinical practice. J Proteomics. https://doi.org/10.1016/j.jprot.2016.05.006
233. Wilkins MR, Appel RD, van Eyk JE et al (2006) Guidelines for the next 10 years of proteomics. Proteomics 6(1):4–8. https://doi.org/10.1002/pmic.200500856
234. Issaq HJ, Waybright TJ, Veenstra TD (2011) Cancer biomarker discovery: opportunities and pitfalls in analytical methods. Electrophoresis 32(9):967–975. https://doi.org/10.1002/elps.201000588
235. Jahn H, Wittke S, Zurbig P et al (2011) Peptide fingerprinting of Alzheimer's disease in cerebrospinal fluid: identification and prospective evaluation of new synaptic biomarkers. PLoS One 6(10):e26540. https://doi.org/10.1371/journal.pone.0026540
236. Wijte D, McDonnell L, Balog CI et al (2012) A novel peptidomics approach to detect markers of Alzheimer's disease in cerebrospinal fluid. Methods 56(4):500–507. https://doi.org/10.1016/j.ymeth.2012.03.018
237. Holtta M, Minthon L, Hansson O et al (2015) An integrated workflow for multiplex CSF proteomics and peptidomics-identification of candidate cerebrospinal fluid biomarkers of Alzheimer's disease. J Proteome Res 14(2):654–663. https://doi.org/10.1021/pr501076j

238. Hendrickson RC, Lee AY, Song Q et al (2015) High resolution discovery proteomics reveals candidate disease progression markers of Alzheimer's disease in human cerebrospinal fluid. PLoS One 10(8):e0135365. https://doi.org/10.1371/journal.pone.0135365
239. Schrader M, Selle H (2006) The process chain for peptidomic biomarker discovery. Dis Markers 22(1-2):27–37
240. Bauca JM, Martinez-Morillo E, Diamandis EP (2014) Peptidomics of urine and other biofluids for cancer diagnostics. Clin Chem 60(8):1052–1061. https://doi.org/10.1373/clinchem.2013.211714
241. Jorgenson JW (2010) Capillary liquid chromatography at ultrahigh pressures. Annu Rev Anal Chem (Palo Alto Calif) 3:129–150. https://doi.org/10.1146/annurev.anchem.1.031207.113014
242. Figeys D, Aebersold R (1998) High sensitivity analysis of proteins and peptides by capillary electrophoresis-tandem mass spectrometry: recent developments in technology and applications. Electrophoresis 19(6):885–892. https://doi.org/10.1002/elps.1150190603
243. Hernandez-Borges J, Neususs C, Cifuentes A et al (2004) On-line capillary electrophoresis-mass spectrometry for the analysis of biomolecules. Electrophoresis 25(14):2257–2281. https://doi.org/10.1002/elps.200405954
244. Schiffer E, Mischak H, Novak J (2006) High resolution proteome/peptidome analysis of body fluids by capillary electrophoresis coupled with MS. Proteomics 6(20):5615–5627. https://doi.org/10.1002/pmic.200600230
245. Simo C, Cifuentes A, Kasicka V (2013) Capillary electrophoresis-mass spectrometry for Peptide analysis: target-based approaches and proteomics/peptidomics strategies. Methods Mol Biol 984:139–151. https://doi.org/10.1007/978-1-62703-296-4_11
246. Scriba GK (2016) Separation of peptides by capillary electrophoresis. Methods Mol Biol 1483:365–391. https://doi.org/10.1007/978-1-4939-6403-1_18
247. Aebersold R, Mann M (2016) Mass-spectrometric exploration of proteome structure and function. Nature 537(7620):347–355. https://doi.org/10.1038/nature19949
248. Gygi SP, Rist B, Griffin TJ et al (2002) Proteome analysis of low-abundance proteins using multidimensional chromatography and isotope-coded affinity tags. J Proteome Res 1(1):47–54
249. Ewles M, Goodwin L (2011) Bioanalytical approaches to analyzing peptides and proteins by LC–MS/MS. Bioanalysis 3(12):1379–1397. https://doi.org/10.4155/bio.11.112
250. Pailleux F, Beaudry F (2012) Internal standard strategies for relative and absolute quantitation of peptides in biological matrices by liquid chromatography tandem mass spectrometry. Biomed Chromatogr 26(8):881–891. https://doi.org/10.1002/bmc.2757
251. Romanova EV, Dowd SE, Sweedler JV (2013) Quantitation of endogenous peptides using mass spectrometry based methods. Curr Opin Chem Biol 17(5):801–808. https://doi.org/10.1016/j.cbpa.2013.05.030
252. Syka JEP, Coon JJ, Schroeder MJ et al (2004) Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. Proc Natl Acad Sci U S A 101(26):9528–9533. https://doi.org/10.1073/pnas.0402700101
253. Olsen JV, Macek B, Lange O et al (2007) Higher-energy C-trap dissociation for peptide modification analysis. Nat Methods 4(9):709–712. https://doi.org/10.1038/nmeth1060
254. Seidler J, Zinn N, Boehm ME et al (2010) De novo sequencing of peptides by MS/MS. Proteomics 10(4):634–649. https://doi.org/10.1002/pmic.200900459
255. Shen Y, Tolic N, Xie F et al (2011) Effectiveness of CID, HCD, and ETD with FT MS/MS for degradomic-peptidomic analysis: comparison of peptide identification methods. J Proteome Res 10(9):3929–3943. https://doi.org/10.1021/pr200052c
256. Guthals A, Clauser KR, Frank A et al (2013) Sequencing-grade de novo analysis of MS/MS triplets (CID/HCD/ETD) from overlapping peptides. J Proteome Res 12(6):2846–2857. https://doi.org/10.1021/pr400173d
257. Shen Y, Tolic N, Purvine SO et al (2012) Improving collision induced dissociation (CID), high energy collision dissociation (HCD), and electron transfer dissociation (ETD) fourier transform MS/MS degradome-peptidome identifications using high accuracy mass information. J Proteome Res 11(2):668–677. https://doi.org/10.1021/pr200597j
258. Makarov A, Scigelova M (2010) Coupling liquid chromatography to Orbitrap mass spectrometry. J Chromatogr A 1217(25):3938–3945. https://doi.org/10.1016/j.chroma.2010.02.022
259. Kaufmann R, Spengler B, Lutzenkirchen F (1993) Mass spectrometric sequencing of linear peptides by product-ion analysis in a reflectron time-of-flight mass spectrometer using matrix-assisted laser desorption ionization. Rapid Commun Mass Spectrom 7(10):902–910. https://doi.org/10.1002/rcm.1290071010

260. Vestal ML, Campbell JM (2005) Tandem time-of-flight mass spectrometry. Methods Enzymol 402:79–108. https://doi.org/10.1016/S0076-6879(05)02003-3
261. Samgina TY, Artemenko KA, Gorshkov VA et al (2008) Oxidation versus carboxamidomethylation of S-S bond in ranid frog peptides: pro and contra for de novo MALDI-MS sequencing. J Am Soc Mass Spectrom 19(4):479–487. https://doi.org/10.1016/j.jasms.2007.12.010
262. Caprioli RM, Farmer TB, Gile J (1997) Molecular imaging of biological samples: localization of peptides and proteins using MALDI-TOF MS. Anal Chem 69(23):4751–4760
263. Chaurand P, Caprioli RM (2002) Direct profiling and imaging of peptides and proteins from mammalian cells and tissue sections by mass spectrometry. Electrophoresis 23(18):3125–3135. https://doi.org/10.1002/1522-2683(200209)23:18<3125:AID-ELPS3125>3.0.CO;2-#
264. Goodwin RJ, Nilsson A, Borg D et al (2012) Conductive carbon tape used for support and mounting of both whole animal and fragile heat-treated tissue sections for MALDI MS imaging and quantitation. J Proteomics 75(16):4912–4920. https://doi.org/10.1016/j.jprot.2012.07.006
265. Beine B, Diehl HC, Meyer HE et al (2016) Tissue MALDI mass spectrometry imaging (MALDI MSI) of peptides. Methods Mol Biol 1394:129–150. https://doi.org/10.1007/978-1-4939-3341-9_10
266. 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(6):1022–1031. https://doi.org/10.1007/s13361-011-0121-0
267. Minerva L, Boonen K, Menschaert G et al (2011) Linking mass spectrometric imaging and traditional peptidomics: a validation in the obese mouse model. Anal Chem 83(20):7682–7691. https://doi.org/10.1021/ac200888j
268. Lanni EJ, Rubakhin SS, Sweedler JV (2012) Mass spectrometry imaging and profiling of single cells. J Proteomics 75(16):5036–5051. https://doi.org/10.1016/j.jprot.2012.03.017
269. Ye H, Wang J, Zhang Z et al (2015) Defining the neuropeptidome of the spiny lobster panulirus interruptus brain using a multidimensional mass spectrometry-based platform. J Proteome Res 14(11):4776–4791. https://doi.org/10.1021/acs.jproteome.5b00627
270. Hishimoto A, Nomaru H, Ye K et al (2016) Molecular histochemistry identifies peptidomic organization and reorganization along striatal projection units. Biol Psychiatry 79(5):415–420. https://doi.org/10.1016/j.biopsych.2015.09.012
271. Prentice BM, Chumbley CW, Caprioli RM (2015) High-speed MALDI MS/MS imaging mass spectrometry using continuous raster sampling. J Mass Spectrom 50(4):703–710. https://doi.org/10.1002/jms.3579
272. Gemperline E, Keller C, Jayaraman D et al (2016) Examination of endogenous peptides in medicago truncatula using mass spectrometry imaging. J Proteome Res. https://doi.org/10.1021/acs.jproteome.6b00471
273. Bivehed E, Stromvall R, Bergquist J et al (2016) Region-specific bioconversion of dynorphin neuropeptide detected by in situ histochemistry and MALDI imaging mass spectrometry. Peptides. https://doi.org/10.1016/j.peptides.2016.11.006
274. Lamerz J, Selle H, Scapozza L et al (2005) Correlation-associated peptide networks of human cerebrospinal fluid. Proteomics 5(11):2789–2798. https://doi.org/10.1002/pmic.200401192
275. Ma B, Zhang K, Hendrie C et al (2003) PEAKS: powerful software for peptide de novo sequencing by tandem mass spectrometry. Rapid Commun Mass Spectrom 17(20):2337–2342. https://doi.org/10.1002/rcm.1196
276. Clynen E, Baggerman G, Huybrechts J et al (2003) Peptidomics of the locust corpora allata: identification of novel pyrokinins (-FXPRLamides). Peptides 24(10):1493–1500
277. Menschaert G, Vandekerckhove TT, Baggerman G et al (2010) A hybrid, de novo based, genome-wide database search approach applied to the sea urchin neuropeptidome. J Proteome Res 9(2):990–996. https://doi.org/10.1021/pr900885k
278. Petruzziello F, Fouillen L, Wadensten H et al (2012) Extensive characterization of Tupaia belangeri neuropeptidome using an integrated mass spectrometric approach. J Proteome Res 11(2):886–896. https://doi.org/10.1021/pr200709j
279. Medzihradszky KF, Chalkley RJ (2015) Lessons in de novo peptide sequencing by tandem mass spectrometry. Mass Spectrom Rev 34(1):43–63
280. Caers J, Boonen K, van den Abbeele J et al (2015) Peptidomics of Neuropeptidergic Tissues of the Tsetse Fly Glossina morsitans morsitans. J Am Soc Mass Spectrom 26(12):2024–2038. https://doi.org/10.1007/s13361-015-1248-1

281. Toullec J-Y, Corre E, Bernay B et al (2013) Transcriptome and peptidome characterisation of the main neuropeptides and peptidic hormones of a euphausiid. The Ice Krill, Euphausia crystallorophias. PLoS One 8(8):e71609. https://doi.org/10.1371/journal.pone.0071609
282. Zhang S, Cao X, He Y et al (2014) Semi-quantitative analysis of changes in the plasma peptidome of Manduca sexta larvae and their correlation with the transcriptome variations upon immune challenge. Insect Biochem Mol Biol 47:46–54. https://doi.org/10.1016/j.ibmb.2014.02.002
283. Seligmann H (2015) Codon expansion and systematic transcriptional deletions produce tetra-, pentacoded mitochondrial peptides. J Theor Biol 387:154–165. https://doi.org/10.1016/j.jtbi.2015.09.030
284. Koehbach J, Clark RJ (2016) Unveiling the diversity of cyclotides by combining peptidome and transcriptome analysis. Biopolymers. https://doi.org/10.1002/bip.22858
285. Mamone G, Picariello G, Caira S et al (2009) Analysis of food proteins and peptides by mass spectrometry-based techniques. J Chromatogr A 1216(43):7130–7142. https://doi.org/10.1016/j.chroma.2009.07.052
286. Panchaud A, Affolter M, Kussmann M (2012) Mass spectrometry for nutritional peptidomics: how to analyze food bioactives and their health effects. J Proteomics 75(12):3546–3559. https://doi.org/10.1016/j.jprot.2011.12.022
287. Dallas DC, Guerrero A, Khaldi N et al (2013) Extensive in vivo human milk peptidomics reveals specific proteolysis yielding protective antimicrobial peptides. J Proteome Res 12(5):2295–2304. https://doi.org/10.1021/pr400212z
288. Baum F, Fedorova M, Ebner J et al (2013) Analysis of the endogenous peptide profile of milk: identification of 248 mainly casein-derived peptides. J Proteome Res 12(12):5447–5462. https://doi.org/10.1021/pr4003273
289. Sassi M, Arena S, Scaloni A (2015) MALDI-TOF-MS platform for integrated proteomic and peptidomic profiling of milk samples allows rapid detection of food adulterations. J Agric Food Chem 63(27):6157–6171. https://doi.org/10.1021/acs.jafc.5b02384
290. Dallas DC, Citerne F, Tian T et al (2016) Peptidomic analysis reveals proteolytic activity of kefir microorganisms on bovine milk proteins. Food Chem 197(Pt A):273–284. https://doi.org/10.1016/j.foodchem.2015.10.116
291. Guarino C, Fuselli F, La Mantia A et al (2010) Peptidomic approach, based on liquid chromatography/electrospray ionization tandem mass spectrometry, for detecting sheep's milk in goat's and cow's cheeses. Rapid Commun Mass Spectrom 24(6):705–713. https://doi.org/10.1002/rcm.4426
292. Sforza S, Cavatorta V, Lambertini F et al (2012) Cheese peptidomics: a detailed study on the evolution of the oligopeptide fraction in Parmigiano-Reggiano cheese from curd to 24 months of aging. J Dairy Sci 95(7):3514–3526. https://doi.org/10.3168/jds.2011-5046
293. Gallego M, Mora L, Aristoy MC et al (2015) Titin-derived peptides as processing time markers in dry-cured ham. Food Chem 167:326–339. https://doi.org/10.1016/j.foodchem.2014.06.088
294. Castellano P, Mora L, Escudero E et al (2016) Antilisterial peptides from Spanish dry-cured hams: purification and identification. Food Microbiol 59:133–141. https://doi.org/10.1016/j.fm.2016.05.018
295. Lahrichi SL, Affolter M, Zolezzi IS et al (2013) Food peptidomics: large scale analysis of small bioactive peptides—a pilot study. J Proteomics 88:83–91. https://doi.org/10.1016/j.jprot.2013.02.018
296. Habibi-Najafi MB, Lee BH (1996) Bitterness in cheese: a review. Crit Rev Food Sci Nutr 36(5):397–411. https://doi.org/10.1080/10408399609527733
297. Liepke C, Zucht HD, Forssmann WG et al (2001) Purification of novel peptide antibiotics from human milk. J Chromatogr B Biomed Sci Appl 752(2):369–377
298. Liepke C, Adermann K, Raida M et al (2002) Human milk provides peptides highly stimulating the growth of bifidobacteria. Eur J Biochem 269(2):712–718
299. Ferreira IM, Jorge K, Nogueira LC et al (2005) Effects of the combination of hydrophobic polypeptides, iso-alpha acids, and malto-oligosaccharides on beer foam stability. J Agric Food Chem 53(12):4976–4981. https://doi.org/10.1021/jf047796w
300. Picariello G, Mamone G, Cutignano A et al (2015) Proteomics, peptidomics, and immunogenic potential of wheat beer (Weissbier). J Agric Food Chem 63(13):3579–3586. https://doi.org/10.1021/acs.jafc.5b00631
301. Ibanez C, Simo C, Garcia-Canas V et al (2013) Metabolomics, peptidomics and proteomics applications of capillary electrophoresis-mass spectrometry in Foodomics: a review. Anal Chim Acta 802:1–13. https://doi.org/10.1016/j.aca.2013.07.042
302. Ortea I, Pascoal A, Canas B et al (2012) Food authentication of commercially-relevant shrimp and prawn species: from classical methods to Foodomics. Electrophoresis

33(15):2201–2211. https://doi.org/10.1002/elps.201100576
303. Montowska M, Alexander MR, Tucker GA et al (2015) Authentication of processed meat products by peptidomic analysis using rapid ambient mass spectrometry. Food Chem 187:297–304. https://doi.org/10.1016/j.foodchem.2015.04.078
304. Silva FA, de Sousa OM, de Souza JM et al (2017) Plant proteomics and peptidomics in host-pathogen interactions: the weapons used by each side. Curr Protein Pept Sci 18(4):400–410
305. Ohyama K, Ogawa M, Matsubayashi Y (2008) Identification of a biologically active, small, secreted peptide in Arabidopsis by in silico gene screening, followed by LC-MS-based structure analysis. Plant J 55(1):152–160. https://doi.org/10.1111/j.1365-313X.2008.03464.x
306. Farrokhi N, Whitelegge JP, Brusslan JA (2008) Plant peptides and peptidomics. Plant Biotechnol J 6(2):105–134. https://doi.org/10.1111/j.1467-7652.2007.00315.x
307. Sarethy IP (2016) Plant peptides: bioactivity, opportunities and challenges. Protein Pept Lett 24(2):102–108
308. Hellinger R, Koehbach J, Soltis DE et al (2015) Peptidomics of circular cysteine-rich plant peptides: analysis of the diversity of cyclotides from viola tricolor by transcriptome and proteome mining. J Proteome Res 14(11):4851–4862. https://doi.org/10.1021/acs.jproteome.5b00681
309. Serra A, Hemu X, Nguyen GKT et al (2016) A high-throughput peptidomic strategy to decipher the molecular diversity of cyclic cysteine-rich peptides. Sci Rep 6:23005. https://doi.org/10.1038/srep23005
310. Fesenko IA, Arapidi GP, Skripnikov AY et al (2015) Specific pools of endogenous peptides are present in gametophore, protonema, and protoplast cells of the moss Physcomitrella patens. BMC Plant Biol 15:87. https://doi.org/10.1186/s12870-015-0468-7
311. Engelhard VH, Appella E, Benjamin DC et al (1993) Mass spectrometric analysis of peptides associated with the human class I MHC molecules HLA-A2.1 and HLA-B7 and identification of structural features that determine binding. Chem Immunol 57:39–62
312. Istrail S, Florea L, Halldorsson BV et al (2004) Comparative immunopeptidomics of humans and their pathogens. Proc Natl Acad Sci U S A 101(36):13268–13272. https://doi.org/10.1073/pnas.0404740101
313. Bassani-Sternberg M, Barnea E, Beer I et al (2010) Soluble plasma HLA peptidome as a potential source for cancer biomarkers. Proc Natl Acad Sci U S A 107(44):18769–18776. https://doi.org/10.1073/pnas.1008501107
314. Caron E, Kowalewski DJ, Chiek Koh C et al (2015) Analysis of major histocompatibility complex (MHC) immunopeptidomes using mass spectrometry. Mol Cell Proteomics 14(12):3105–3117. https://doi.org/10.1074/mcp.O115.052431
315. Bassani-Sternberg M, Coukos G (2016) Mass spectrometry-based antigen discovery for cancer immunotherapy. Curr Opin Immunol 41:9–17. https://doi.org/10.1016/j.coi.2016.04.005
316. Bassani-Sternberg M, Braunlein E, Klar R et al (2016) Direct identification of clinically relevant neoepitopes presented on native human melanoma tissue by mass spectrometry. Nat Commun 7:13404. https://doi.org/10.1038/ncomms13404
317. Backert L, Kowalewski DJ, Walz S et al (2017) A meta-analysis of HLA peptidome composition in different hematological entities: entity-specific dividing lines and "pan-leukemia" antigens. Oncotarget 8(27):43915–43924. https://doi.org/10.18632/oncotarget.14918
318. Abelin JG, Keskin DB, Sarkizova S et al (2017) Mass spectrometry profiling of HLA-associated peptidomes in mono-allelic cells enables more accurate epitope prediction. Immunity 46(2):315–326. https://doi.org/10.1016/j.immuni.2017.02.007
319. Uhlen M, Hallstrom BM, Lindskog C et al (2016) Transcriptomics resources of human tissues and organs. Mol Syst Biol 12(4):862. https://doi.org/10.15252/msb.20155865
320. Kim M-S, Pinto SM, Getnet D et al (2014) A draft map of the human proteome. Nature 509(7502):575–581. https://doi.org/10.1038/nature13302
321. Wilhelm M, Schlegl J, Hahne H et al (2014) Mass-spectrometry-based draft of the human proteome. Nature 509(7502):582–587. https://doi.org/10.1038/nature13319
322. Wishart DS, Jewison T, Guo AC et al (2013) HMDB 3.0—the human metabolome database in 2013. Nucleic Acids Res 41(Database issue):D801–D807. https://doi.org/10.1093/nar/gks1065
323. Sud M, Fahy E, Cotter D et al (2016) Metabolomics workbench: an international repository for metabolomics data and metadata, metabolite standards, protocols, tutorials and training, and analysis tools. Nucleic Acids Res 44(D1):D463–D470. https://doi.org/10.1093/nar/gkv1042
324. Fahy E, Cotter D, Sud M et al (2011) Lipid classification, structures and tools. Biochim Biophys Acta 1811(11):637–647. https://doi.org/10.1016/j.bbalip.2011.06.009
325. Cummings RD, Pierce JM (2014) The challenge and promise of glycomics. Chem Biol 21(1):1–15. https://doi.org/10.1016/j.chembiol.2013.12.010

# Chapter 2

# Brain Tissue Sample Stabilization and Extraction Strategies for Neuropeptidomics

**Elva Fridjonsdottir, Anna Nilsson, Henrik Wadensten, and Per E. Andrén**

## Abstract

Neuropeptides are bioactive peptides that are synthesized and secreted by neurons in signaling pathways in the brain. Peptides and proteins are extremely vulnerable to proteolytic cleavage when their biological surrounding changes. This makes neuropeptidomics challenging due to the rapid alterations that occur to the peptidome after harvesting of brain tissue samples. For a successful neuropeptidomic study the biological tissue sample analyzed should resemble the premortem state as much as possible. Heat stabilization has been proven to inhibit postmortem degradation by denaturing proteolytic enzymes, hence increasing identification rates of neuropeptides. Here, we describe a stabilization protocol of a frozen tissue specimen that increases the number of intact mature neuropeptides identified and minimizes interference of degradation products from abundant proteins. Additionally, we present an extraction protocol that aims to extract a wide range of hydrophilic and hydrophobic neuropeptides by using both an aqueous and an organic extraction medium.

**Key words** Neuropeptides, Brain tissue, Sample preparation, Stabilization, Heat inactivation, Postmortem degradation, Extraction, Electrospray mass spectrometry

## 1 Introduction

Neuropeptides are peptides secreted by neurons through regulated routes and they commonly act on brain receptors. They are produced from prohormones after cleavage by peptidases such as prohormone convertases and carboxypeptidases. They can act as neurotransmitters directly, as modulators of neurotransmission, as paracrine or autocrine regulators, or as hormones where they are transported by the circulatory system to target distant organs to regulate physiology and behavior [1]. These molecules are of interest due to their diverse biological functions and associations in various disease states [2].

Neuropeptidomics and proteomics share many technical aspects using reversed phase nanoflow liquid chromatography electrospray ionization tandem mass spectrometry (nanoLC-ESI MS/MS) but the latter mostly focuses on the characterization of in vitro digested

Michael Schrader and Lloyd Fricker (eds.), *Peptidomics: Methods and Strategies*, Methods in Molecular Biology, vol. 1719, https://doi.org/10.1007/978-1-4939-7537-2_2, 

peptides from proteins. However, neuropeptides from the same precursor can have distinct bioactive effects; therefore it is desirable to characterize the undigested, native neuropeptides with preserved posttranslational modifications. In these analyses, it is extremely important that the analytes measured reflect the actual in vivo concentrations as closely as possible.

Following tissue sampling, significant alterations occur to the proteome and the peptidome due to a rapid release of degradation mediators [3, 4]. Within seconds after sampling, levels of analytes as well as posttranslational modification can change, so they do not reflect their in vivo state anymore and important biological information is lost [5]. The proteolytic degradation affects the peptidomic analysis either directly by proteolytic cleavage of analytes of interest (Fig. 1), or indirect by significantly increasing the complexity of the sample by degradation of highly abundant proteins.

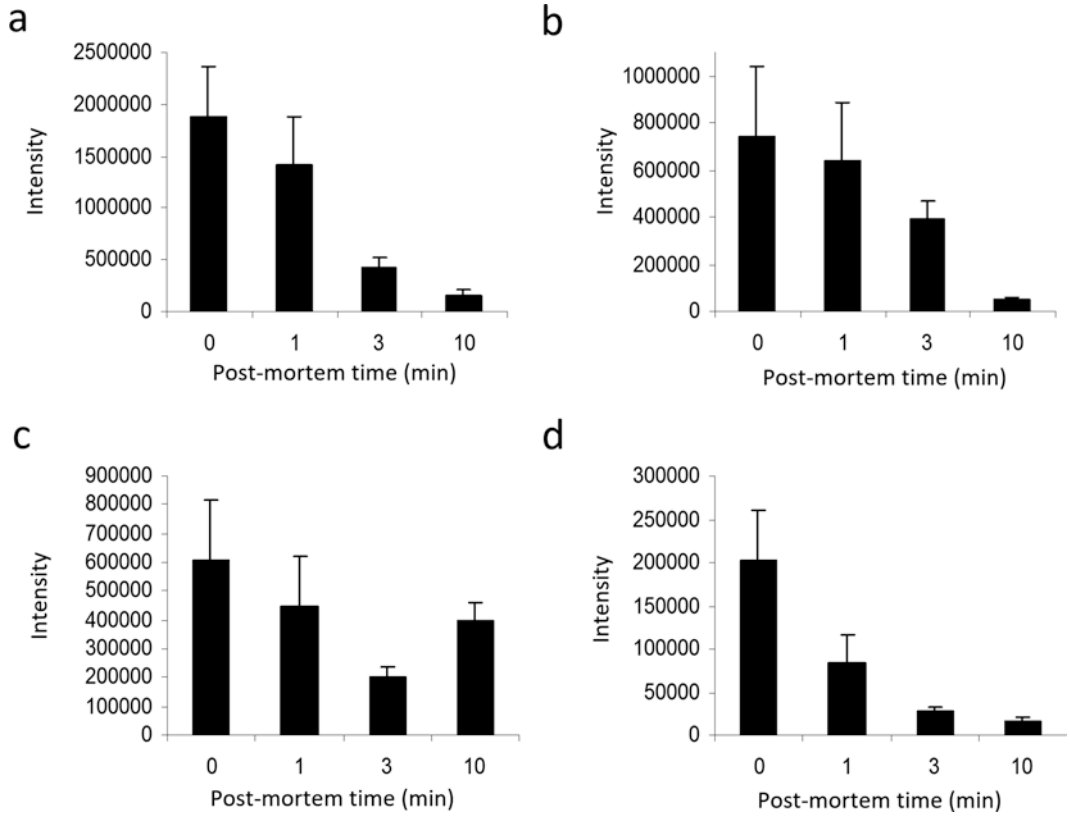



**Fig. 1** Postmortem time effect on the relative levels of neuropeptides. (**a**) Proopiomelanocortin-derived peptide, (**b**) beta-endorphin, (**c**) CLIP, and (**d**) phosphorylated CLIP. Mouse brain tissue (hypothalamus) was investigated with different postmortem intervals ($n = 4$). The control group was sacrificed by focused microwave irradiation (0 min). Three groups were sacrificed by cervical dislocation and kept in room temperature (22 °C) at different time intervals, 1, 3, and 10 min, respectively, and the brains were subsequently irradiated by focused microwaves (from reference [5])

Heat stabilization prevents post-sampling proteolytic activity by denaturating the proteolytic enzymes and is an effective tool to avoid these complications [6]. It should be performed prior to any other post-sampling handling procedure and the postmortem time should be kept to a minimum to give the most effective result. Microwave irradiation can raise the temperature to 90 °C within a few seconds, inducing the rapid denaturation of proteins and inactivation of proteases [3]. However, the morphology of the samples can change and the heat distribution can be uneven throughout the sample [7]. Another, approach is to apply heat stabilization with the Denator Stabilizor system. It is a sample inactivation instrument for instant fixation of tissue designed for neuropeptidomics [4]. This instrument provides a controlled and uniform heat transfer to the sample for the time needed to reach 90 °C and can preserve the morphology of the sample. It can be performed either directly on fresh tissue or on frozen tissue [8] and it preserves posttranslational modifications like phosphorylations [4]. When comparing denatured samples with non-denatured samples, the number of candidate neuropeptides increases in denatured samples and background interference caused by degradation of abundant proteins is minimized [4, 6] (Fig. 2).

Extraction procedures for neuropeptidomics should aim to extract a diverse range of peptides, preserve posttranslational modifications, and still be as simple and efficient as possible. An extraction in aqueous acetic acid has been used in many neuropeptidomic studies with consistent and reproducible results in peptide identification and quantitative analysis [9–13]. However, neuropeptides have diverse physicochemical properties and vary a lot in molecular weight. Shorter peptides are often hydrophilic but as their length increases they become less soluble in water, complicating the extraction of longer peptide chains in aqueous solutions. In aqueous solutions larger peptides might also adsorb to autosampler vials or LC-tubing. With a higher concentration of organic solvent these effects diminish and extraction and storage of the hydrophobic peptides is improved but the recovery of smaller peptides is negatively affected [14]. Considering these complications, a two-step extraction that starts with an aqueous extraction followed by an organic extraction seems to be an appropriate procedure that covers a wide range of peptides. Zhang et al. [15] presented an approach where the aqueous fraction was extracted first, followed by extractions with 20% and then 50% methanol. This resulted in increased numbers of identified peptides and allowed identification of hydrophobic and larger peptides previously unreported in neuropeptidomic studies [15].

Here, we describe a stabilization protocol for frozen brain tissue samples and an extraction protocol for neuropeptides. The extraction is divided into two parts; first an aqueous acidic extraction and then an extraction in methanol to increase the extraction efficacy of hydrophobic peptide chains. A centrifugal filter device is used to separate molecules larger than 10 kDa from the peptides. The aqueous and

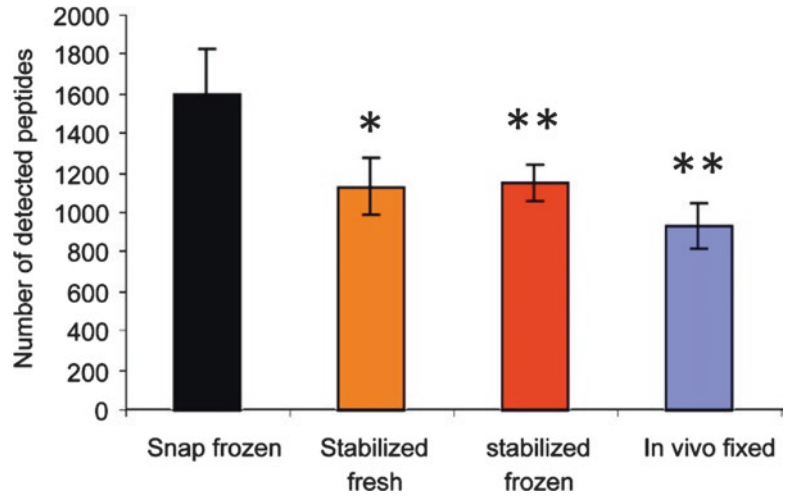



**Fig. 2** Number of detected peptides derived from degraded proteins. Comparisons were made between snap-frozen mouse brain tissue (black), stabilized fresh tissue (orange), stabilized frozen tissue (red), and in vivo fixed tissue by focused microwave radiation (blue). The analysis was carried out by nanoLC-ESI-MS ($n = 4$). Enzyme proteolytic activity was inhibited in the stabilized and in vivo fixed samples. This is shown by a significantly larger number of detected peptides in the snap-frozen samples compared to the stabilized fresh and frozen samples and in vivo fixed samples ($^*p < 0.05$, $^{**}p < 0.01$). The study demonstrated that rapid neuropeptide and protein postmortem changes occurred within minutes in brain tissue (modified from reference [4]). Hence, one should not assume that the more peptides identified, the better the results, but instead, this might be an indicator for high postmortem proteolytic activity. Van Dijck et al. compared microwave heating to two different freezing methods and concluded that freezing in liquid nitrogen was the best method due to higher numbers of identified peptides and better reproducibility [13]. However, a better assessment of sample quality and integrity would be to search for markers of sample degradation, such as peptides derived from abundant proteins, e.g., hemoglobins, tubulins, dynamins, heat shock proteins, actins [4], α-synucleins, and stathmins [5]. A low number of these peptides would be an indicator for a high quality sample that represents the true biological in vivo peptidome more closely

organic phases are collected separately. The two solutions can be analyzed in two separate nanoLC-ESI MS/MS analyses or they can be analyzed in the same analysis either by mixing the solutions prior to injection, or they can be mixed on the pre-column by injecting the organic fraction first and then the aqueous fraction.

## 2 Materials

### *2.1 Stabilization with Heat-Induced Denaturation*

1. The heat stabilizer system (Stabilizor, Denator) consists of the Stabilizor T1 instrument and Maintainor sample cards. The heat stabilizer instrument has a conductive heating unit in its core. It is a benchtop instrument that needs no special inputs. The sample cards hold the sample between two thin Teflon-based plastic films.
2. Spatula to handle tissue samples.
3. 1.5 mL microfuge tubes with low protein binding.
4. 50% (v/v) methanol in water to wash sample cards.

### 2.2 Materials for Extraction of Neuropeptides

1. Acetic acid extraction solution (AAS) 0.25% (v/v) acetic acid in water. Use LC–MS grade water and liquid acetic acid of ≥99.7% purity (*see* **Note 1**). Prepare the solution and cool on ice prior to extraction.
2. Methanol, liquid chromatography grade.
3. Homogenizer: Microtip sonication rod with about 2.2 mm diameter tip (we use Vibra-Cell, Sonics).
4. A refrigerated centrifuge that can achieve 14,000 × *g* and accommodate 1.5 mL microfuge tubes.
5. Centrifugal filter device (referred to as spin-filters): The device consists of two components, a sample reservoir with a nominal 10 kDa cutoff membrane and a filtrate vial (*see* **Note 2**). We use Microcon centrifugal filters for protein purification (EMD Millipore).
6. Methanol and distilled water to wash the sonication tip.

## 3 Methods

### 3.1 Heat Stabilization of Frozen Tissue Samples

A proper experiment should have as little variation as possible between preparations of various samples. Hence, it is important to plan the sample preparation thoroughly. When preparing a large number of samples, divide them into batches that can be completed in one working day. This protocol describes step by step how to use the heat stabilizer system for thermal stabilization of frozen brain tissue samples (*see* **Note 3**). Whole brain or dissected structures that fit into the sample card chamber can be treated according to this protocol (*see* **Note 4**).

1. Label all 1.5 mL low-bind microfuge tubes carefully, one tube per sample. Weigh them and store on dry ice while performing the stabilization.
2. Transfer frozen samples from −80 °C freezer to dry ice (*see* **Note 5**). Keep the samples frozen until they are inserted into the instrument.
3. Turn on the heat stabilizer instrument. Wait while the system starts up and the thermo plates reach 95 °C and the wait button changes to start. Select an appropriate method for frozen samples (*see* **Note 6**).
4. Open up the sample card 180° and cool on dry ice. Transfer the frozen brain tissue with a chilled forceps or spatula to the middle of the sample chamber. Close the card and press gently with fingers around the plastic edges on the sample chamber to tighten it (*see* **Note 7**). It has to be closed airtight prior to insertion into the heat stabilizer instrument.

5. Put the sample card in the sample slot in the heat stabilizer instrument.
6. Press start directly to initiate the heat treatment.
7. The sample card is ejected from the instrument at the end of stabilization.
8. Immediately after stabilization, remove the sample card from the slot and cover it with dry ice. Wait for 30 s for it to freeze. Open up the sample card and use a chilled spatula to loosen the sample and transfer it to a pre-weighted 1.5 mL low-bind microfuge tube stored on dry ice (*see* **Note 8**).
9. Weigh the tubes and record the weight of the tissue. Store the sample at −80 °C until further preparation or continue with homogenization and extraction directly.

### 3.2 Extraction of Neuropeptides from Heat Stabilized Brain Samples

It is important to plan the extraction procedure thoroughly before starting the extraction, and divide samples into batches if necessary. Calculate the amount of volume to be added to each sample, with respect to the weight of the sample (**steps 2**, **9**, and **13** show the volume of each extraction solution to be added per mg tissue (*see* **Note 9**)). Set up the equipment needed for homogenization. Set the centrifuge to 4 °C. Prepare 0.25% AAS and put it on ice to cool along with methanol. Bring the heat-stabilized samples on dry ice. All handlings should be performed on regular ice.

1. Prepare two spin-filters per sample, one for aqueous extraction (spin-filter 1) and one for organic extraction (spin-filter 2) (*see* **Note 10**). Put the sample reservoir into the filtrate vial. Add 300 μL of 0.25% AAS to the sample reservoir and centrifuge at 14,000 × *g* and 4 °C for 1 h (*see* **Note 11**).
2. Add 7.5 μL of cold 0.25% AAS per mg tissue to the low binding microfuge tube containing the sample.
3. Homogenize with a Microtip sonication rod for 30 s (*see* **Note 12**). Wash the sonication tip between each sample by rinsing first with methanol and then water.
4. Store the sample on ice until **steps 2** and **3** are completed for all samples.
5. Centrifuge the samples for 40 min at 14,000 × *g* and 4 °C to separate insoluble material.
6. Transfer the supernatant to spin-filter 1. Save the pellet and store on ice; it will be used in the organic extraction (**step 9**).
7. Place the spin-filters in the centrifuge and filter the supernatants by centrifugation for 90 min at 4 °C and 14,000 × *g* (*see* **Note 13**).
8. After filtration, the samples in the filtrate vials are ready for nanoLC-ESI MS/MS analysis and should be frozen immediately on dry ice and stored at −80 °C (*see* **Note 2**).

9. To the remaining pellet from **step 6**, add 3 μL of cold 0.25% AAS and 0.75 μL of methanol per mg tissue (20% methanol extraction).
10. Homogenize by sonication as described earlier (**step 3**).
11. Centrifuge the samples for 40 min at 14,000 × *g* and 4 °C.
12. Transfer the supernatant to spin-filter 2. Keep the spin-filter on ice, an additional portion of supernatant will be added to it in **step 16**.
13. Add 1.87 μL of cold 0.25% AAS and 1.87 μL of methanol per mg tissue to the pellet (50% methanol extraction).
14. Homogenize by sonication as described earlier.
15. Centrifuge the samples for 40 min at 14,000 × *g* and 4 °C.
16. Transfer the supernatant to spin-filter 2.
17. Place the spin-filters in the centrifuge and filter the pooled supernatants by centrifugation for 90 min at 4 °C and 14,000 × *g*.
18. After filtration, the samples in the filtrate vials are ready for nanoLC-ESI MS/MS analysis and should be frozen immediately on dry ice and stored at −80 °C (*see* **Note 2**).

## 4 Notes

1. To avoid contamination, do not pipette directly from the bottle.
2. The 10 kDa cutoff spin-filter collects all soluble material with molecular weight larger than about 10 kDa. While the flow through is used for peptidomics studies, the retained material can be used for potential proteomic studies. The >10 kDa fraction can be collected by turning the filter upside down in a fresh vial and spun for a few minutes. Store it in −80 °C until further preparation.
3. This protocol can also be used on fresh tissue directly after sampling. In that case, make sure that everything is ready for stabilization before euthanizing the animals and collecting tissue. It is important to keep the postmortem time prior to stabilization as short as possible and consistent between samples.
4. For reference, a whole rat brain fits into the sample chamber but larger tissue samples may need sectioning prior to insertion into the heat stabilizer.
5. The instrument is designed to treat samples between −78 °C and +20 °C [16]. Samples frozen in liquid nitrogen need to equilibrate on dry ice or at −20 °C before insertion into the heat stabilizer.

6. The quick frozen method will give very rapid heating by compressing the sample. It is convenient for small tissue samples but the compression affects the morphology of larger samples. Choose the frozen-structural preserve method to prevent morphology changes; it will take longer time to stabilize but the structural integrity will be preserved. When stabilizing fresh tissue select a method for fresh tissue, the same principle applies for these methods.
7. If the sample card stays on dry ice for more than 1 min it can freeze. Thaw it slightly with fingers on the plastic edges around the sample chamber and make sure it is thoroughly closed. At the same time, try to minimize the air in the cavity between the foils by pressing them inward toward the sample.
8. Each sample chamber can be used at least three times. Wash the sample chamber with 50% methanol between samples.
9. 7.5 μL of extraction solution is added per mg sample. In quantitation analysis it is important to keep this volume exact for all samples. In the aqueous extraction the whole volume (7.5 μL) is added at once. In the organic extraction the volume is added in two portions (3.75 μL + 3.75 μL). The resulting added volume would be the same per mg sample in both the aqueous and organic extractions.
10. The resulting samples will be stored in these filtrate vials, so label them carefully before starting the extraction procedure.
11. The filter membranes need to be rinsed prior to sample filtration to get rid of contaminants and ensure even filtration rates between samples. It is important that the membranes are not allowed to dry out before use.
12. Keep the sonication tip under the surface of the solution while homogenizing to avoid foaming.
13. Align the cap strap toward the center of the rotor when placing it into the centrifuge.

## References

1. Burbach JP (2011) What are neuropeptides? Methods Mol Biol. 789:1–36. https://doi.org/10.1007/978-1-61779-310-3_1
2. Zhang X, Petruzziello F, Rainer G (2014) Extending the scope of neuropeptidomics in the mammalian brain. EuPA Open Proteomics 3:273–279. https://doi.org/10.1016/j.euprot.2014.04.004
3. Svensson M, Sköld K, Svenningsson P, Andren PE (2003) Peptidomics-based discovery of novel neuropeptides. J Proteome Res 2(2):213–219
4. Svensson M, Boren M, Sköld K et al (2009) Heat stabilization of tissue proteome: a new technology for improved proteomics. J Proteome Res 8(2):974–981. https://doi.org/10.1021/pr8006446
5. Sköld K, Svensson M, Norrman M et al (2007) The significance of biochemical and molecular sample integrity in brain proteomics and peptidomics: stathmin 2-20 and peptides as sample quality indicators. Proteomics 7(24):4445–4456. https://doi.org/10.1002/pmic.200700142

6. Colgrave ML, Xi L, Lehnert SA et al (2011) Neuropeptide profiling of the bovine hypothalamus: thermal stabilization is an effective tool in inhibiting post-mortem degradation. Proteomics 11(7):1264–1276. https://doi.org/10.1002/pmic.201000423
7. Che FY, Lim J, Pan H et al (2005) Quantitative neuropeptidomics of microwave-irradiated mouse brain and pituitary. Mol Cell Proteomics 4(9):1391–1405. https://doi.org/10.1074/mcp.T500010-MCP200
8. Goodwin RJ, Lang AM, Allingham H et al (2010) Stopping the clock on proteomic degradation by heat treatment at the point of tissue excision. Proteomics 10(9):1751–1761. https://doi.org/10.1002/pmic.200900641
9. Rossbach U, Nilsson A, Fälth M et al (2009) A quantitative peptidomic analysis of peptides related to the endogenous opioid and tachykinin systems in nucleus accumbens of rats following naloxone-precipitated morphine withdrawal. J Proteome Res 8(2):1091–1098. https://doi.org/10.1021/pr800669g
10. Kultima K, Nilsson A, Scholz B et al (2009) Development and evaluation of normalization methods for label-free relative quantification of endogenous peptides. Mol Cell Proteomics 8(10):2285–2295. https://doi.org/10.1074/mcp.M800514-MCP200
11. Karlsson O, Kultima K, Wadensten H et al (2013) Neurotoxin-induced neuropeptide perturbation in striatum of neonatal rats. J Proteome Res 12(4):1678–1690. https://doi.org/10.1021/pr3010265
12. Bourdenx M, Nilsson A, Wadensten H et al (2014) Abnormal structure specific peptide transmission and processing in a primate model of Parkinson's disease and L-DOPA-induced dyskinesia. Neurobiol Dis 62:307–312. https://doi.org/10.1016/j.nbd.2013.10.016
13. Van Dijck A, Hayakawa E, Landuyt B et al (2011) Comparison of extraction methods for peptidomics analysis of mouse brain tissue. J Neurosci Methods 197(2):231–237. https://doi.org/10.1016/j.jneumeth.2011.02.023
14. Taylor SW, Andon NL, Bilakovics JM et al (2006) Efficient high-throughput discovery of large peptidic hormones and biomarkers. J Proteome Res 5(7):1776–1784. https://doi.org/10.1021/pr0600982
15. Zhang X, Petruzziello F, Zani F et al (2012) High identification rates of endogenous neuropeptides from mouse brain. J Proteome Res 11(5):2819–2827. https://doi.org/10.1021/pr3001699
16. Borén M (2015) Sample preservation through heat stabilization of proteins: principles and examples. Methods Mol Biol 1295:21–32. https://doi.org/10.1007/978-1-4939-2550-6_2

# Chapter 3

# Isolation of Endogenous Peptides from Cultured Cell Conditioned Media for Mass Spectrometry

**Kazuki Sasaki, Takashi Tsuchiya, and Tsukasa Osaki**

## Abstract

Media conditioned by cultured cells represent an excellent source rich in endogenous peptides. Unbiased mass spectrometric analysis of the constituent peptides provides an opportunity to look into proteolytic events such as bioactive peptide processing, membrane protein ectodomain shedding, or even regulated intramembrane proteolysis. If conducted on a large scale, peptidomics has the potential to pinpoint primary cleavage sites. Here a method is described for isolating peptides from cultured cell conditioned media before mass spectrometry analysis.

**Key words** Cultured cells, Endogenous peptides, Conditioned medium, Solid phase extraction, Desalting

## 1 Introduction

Conditioned media, in particular from cell lines, were a starting material for identifying novel peptide hormones, growth factors, and cytokines [1–3]. In the wake of technical advances in mass spectrometry, conditioned media again have received attention as a source for identifying primary cleavage sites for peptide hormone processing [4], ectodomain shedding, and regulated intramembrane proteolysis [5]. Consistent with endogenous peptides being prone to degradation, they are often identified in truncated forms. However, primary cleavage sites are highlighted by aligning sequenced peptides to a precursor protein sequence. In this chapter, a sample preparation procedure for isolating endogenous peptides prior to mass spectrometry analysis is provided.

Michael Schrader and Lloyd Fricker (eds.), *Peptidomics: Methods and Strategies*, Methods in Molecular Biology, vol. 1719, https://doi.org/10.1007/978-1-4939-7537-2_3, 

## 2 Materials

Cell culture is conducted with a standard technique. Use HPLC-grade water and chemicals for solid phase extraction and chromatography.

### 2.1 Conditioned Medium Preparation

1. 10-cm dishes for cell culture.
2. Hank's balanced salt solution (HBSS) without phenol red (*see* **Note 1**).
3. Secretagogue of choice (*see* **Note 2**).
4. Cell strainer (*see* **Note 3**).

### 2.2 Solid Phase Extraction

1. Acetonitrile (ACN).
2. Trifluoroacetic acid (TFA).
3. Polymer sorbent solid phase extraction cartridge (we use 6 mL bed volume, RP-1, GL Sciences, Inc., USA) (*see* **Note 4**).
4. Elution solution (40–50% ACN, 0.1% TFA).
5. Centrifuge tubes (15 and 50 mL).
6. Vacuum centrifuge.

### 2.3 Gel Permeation

1. Gel permeation chromatography column (such as G2000$SW_{XL}$, 21.5 mm × 300 mm, TOSOH, Japan) (*see* **Note 5**).
2. Chromatography system with sample loop for column above.
3. Gel permeation eluent (60% ACN, 0.1% TFA).
4. Fraction collector for fractions of 3 mL.
5. 100 μL microsyringe (we use Hamilton).
6. Formic acid.

### 2.4 Reductive Alkylation

1. TE buffer (0.5 M Tris, 1 mM EDTA, pH 8.8).
2. Dithiothreitol buffered solution; 1 M dithiothreitol dissolved in TE buffer (*see* **Note 6**).
3. Iodoacetamide buffered solution; 1 M iodoacetamide dissolved in TE buffer (*see* **Note 6**).

### 2.5 Mass Spectrometry

1. Mass spectrometer, capable of MS/MS fragmentation.
2. Mass spectrometry software for data visualization and deconvolution (usually provided by the vendor).
3. Search engine for comparison of MS/MS data with databases (*see* **Note 14**), such as Mascot (Matrix Science, London).

## 3 Methods

This protocol is for cultured cells that contain the regulated secretory pathway, in which secretory peptides like classical bioactive peptides are stored in secretory granules and await secretion in response to an exocytosis stimulus. Their contents are released to the medium within a short time period (usually in the order of minutes) so that secretory peptides are efficiently harvested [4, 6]. In contrast, this is not applicable to cells devoid of secretory granules; for these cell types the protocol needs a minor modification (*see* **Note 7**).

### 3.1 Recovering

1. Grow adherent cells, cell lines or primary culture, to near confluency in a 10-cm ***Culture Medium*** dish (*see* **Note 8**). Aspirate the spent medium and rinse the culture surface once with 5 mL of prewarmed HBSS to remove components of serum used for culture. Attach the tip of a pipette to the internal wall of the culture plate and gently add the HBSS so that adherent cells would not be detached by mechanical force (*see* **Note 9**).
2. Aspirate the wash solution and immediately add the same amount of HBSS supplemented with an equal volume of solvents used for dissolving secretagogues. Apply the solution carefully as in **step 1**.
3. Return the dish to an incubator for a specified period of time (*see* **Note 10**). Collect the medium in a centrifuge tube chilled in crushed ice, then immediately add 5 mL of prewarmed HBSS containing the secretagogue. Incubate for the same period of time. Collect the medium to another tube.
4. Let each of the recovered media go through a cell strainer and collect the run-through in a new chilled tube (*see* **Note 11**).

### 3.2 Solid Phase

1. Before collecting conditioned medium, equilibrate a solid phase extraction cartridge ***Extraction*** with 5 mL of ACN, followed by 6 mL of water. Put a three-way plug on the cartridge's outlet to prevent drying.
2. Apply the supernatant to a preconditioned cartridge column. Add 5 mL of water to rinse the cartridge and then apply 200 μL of the elution solution. Discard the solution and then add 5 mL of the same solution to elute the bound material. Eluate is collected in a 15 mL centrifuge tube pre-rinsed with the elution solution (*see* **Note 4**).
3. Cover the tube opening with parafilm, which is pierced with a needle to allow the solvent to evaporate. Reduce the content volume to less than 2 mL on a vacuum centrifuge. Freeze it at −80 °C and lyophilize (*see* **Note 12**).

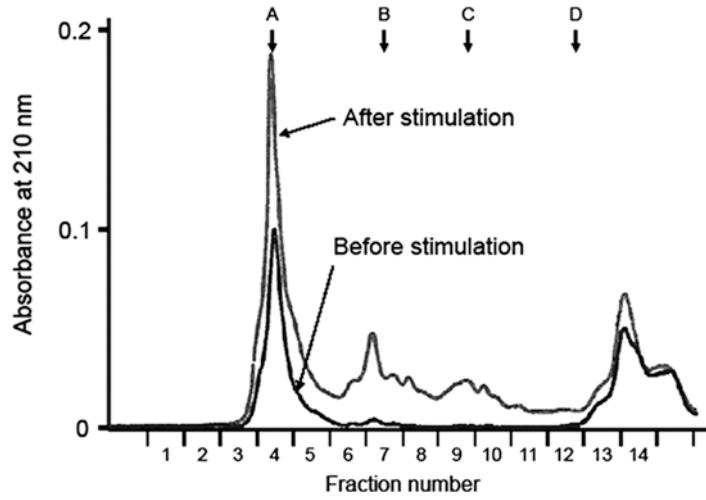



**Fig. 1** Gel filtration chromatography profiles of culture supernatant extracts from human medullary thyroid carcinoma TT cells before (lower trace) and after (upper trace) stimulation for 2 min. *Arrows* indicate molecular mass markers: (A) 66,500 Da; (B) 4271 Da; (C) 1673 Da; (D) 556 Da. Fractions numbered 7–10 were analyzed by LC-MS/MS. (Reproduced from ref. 4 with permission from the American Society for Biochemistry and Molecular Biology)

4. Meanwhile, equilibrate the gel permeation column with 60% ACN, 0.1% TFA at a flow rate of 1.5 mL/min for at least 1 h.
5. Reconstitute the lyophilized material in 100 μL of 60% ACN, 0.1% TFA. Apply the sample onto a gel permeation chromatography column via a sample loop. Twenty minutes after the loading, start collecting fractions every 2 min. Use fractions containing peptides between 1 and 10 kDa (*see* Fig. 1).
6. Reduce the content volume to less than 1 mL on a vacuum centrifuge as in **step 3**. Freeze them at −80 °C and lyophilize. They may be subjected to reductive alkylation if necessary (Proceed to Subheading 3.3).
7. Reconstitute the lyophilized material in an appropriate volume of 5% ACN, 0.1% formic acid. The sample is now ready for LC-MS/MS analysis (*see* **Note 13**).
8. In case analytes are prepared from cells of endocrine or neuronal origin, C-terminally amidated peptides are likely to be identified (*see* **Note 14**). If analytes are from cells constitutively secreting proteins and peptides, primary cleavage sites for ectodomain shedding and regulated intramembrane proteolysis are highlighted.

### 3.3 Reductive *Alkylation*

1. Dissolve the powdered sample obtained in the former **step 5** in 160 μL of ***Alkylation*** TE buffer. Ensure that the pH is not acidic (*see* **Note 15**).

2. Add 4 μL of 1 M DTT buffered solution and incubate at 37 °C for 1 h. Keep the reaction in the dark.
3. Mix well and add 8 μL of 1 M iodoacetamide buffered solution. After an incubation at room temperature for 15 min, add 6 μL of 17.4 M of acetic acid to terminate the reaction.
4. Desalt the mixture with a 1 mL solid phase extraction cartridge as described in Subheading 3.2.

## 4 Notes

1. Phenol red is abundant in culture medium and binds to a column used for peptide enrichment. Avoid using media including phenol red, as the pH indicator interferes with peptide detection by mass spectrometry. Make sure to use HBSS containing calcium and magnesium ions.
2. Secretagogues stimulate cells with secretory granules (e.g., endocrine cells) to increase peptide secretion from the compartment. In endocrine cells, chemicals causing an increase in intracellular cAMP or $Ca^{2+}$ levels can be best used as a secretagogue. Combinations of these chemicals can act synergistically or additively in secretion. However, do not use bioactive peptides, growth factors, or cytokines for stimulation, as once added to medium they are indistinguishable from endogenous peptides (unless from another species which has a slightly different sequence and mass that is distinguishable on MS). Note that recombinant proteins are usually not a single entity and are often accompanied by fragment peptides. Addition of the depolarizing agent potassium chloride to a final concentration of 50 mM is also effective on the regulated secretory pathway. As for cells devoid of secretory granules, they constitutively release proteins and processed peptides. Their secretion is not usually enhanced with secretagogues.
3. During conditioned medium retrieval, cell detachment may happen. Removing floating cells or clusters is best performed with a cell strainer. Make sure that the cleared supernatant is free from cell aggregates under a microscope. Use a cell strainer suitable for the size of target cells. In general, large cells are removed with a 20 μm cell strainer. Centrifugation does not efficiently separate cells since the conditioned medium is retrieved in culture condition essentially free from serum, which serves as a cushion in routine cell culture.
4. In general, a 6 mL RP-1 cartridge handles material from up to three 10-cm dishes. Do not add any detergents (e.g., TX-100) with an aim of enhancing peptide recovery, as they strongly interfere with mass spectrometry. For reference, a 1 mL RP-1 cartridge can hold a tissue extract from one pituitary from a rat.

5. Gel permeation chromatography should be conducted under 60% ACN, 0.1% TFA. This condition disrupts protein–peptide noncovalent interactions, thereby increasing the retrieval of peptides. Use a column compatible with ACN.
6. Dissolve the agents in TE buffer. Prepare the solutions just before use to minimize the generation of by-products during the reaction.
7. Cells lacking secretory granules need a longer incubation of several hours. Use phenol red-free RPMI or DMEM instead of HBSS. Avoid using commercially available serum-free media as they contain large amounts of insulin and transferrin to support cell growth. Endogenous peptide detection would be overwhelmingly suppressed by high levels of insulin in the medium, including both intact insulin molecules and numerous truncated fragments of insulin. The same problem would occur with transferrin.
8. We recommend starting with at least two 10-cm dishes. Only a limited number of peptide sequences (less than 200) were identified with a single dish. Primary cultured cells pose a more challenge in obtaining medium as they are vulnerable to serum-free condition. It would be advisable to add specific low-molecular-weight, water-soluble agents with the ability to support the survival of a given cell type. In general, sodium pyruvate counteracts a detrimental effect of serum-free condition. The media should be harvested within a few hours to minimize the release of intracellular peptides from degrading cells. Cells cultured in suspension or grown in aggregates, such as iPSC cultures, are difficult to work with. In such cases identified peptide species would arise from intracellular proteins, including heat shock proteins, histones, keratins, and tubulins. Note that supernatant is not completely free from these components, even if it seems cleared of cell debris under a microscope.
9. A single wash does not completely remove serum components, but extensive washings render the cells less viable and more likely to detach from the dish. Often encountered peptide species are fragments from major serum proteins, such as fibrinogens, hemoglobins, albumin, alpha-1-antiproteinase, alpha-2-antiplasmin, and alpha-2-HS-glycoprotein. Some of the $m/z$ values corresponding to the observed species could be added to a reject mass list for MS/MS.
10. For secretory granule-bearing cells, a short stimulation period (e.g., 2–15 min, depending on cell types) suffices. In general, peptides are highly prone to degradation; peptide ladders of N- or C-terminally truncated sequences are observed even after a 2-min exposure to cells. Exogenously added protease

inhibitors cocktail do not work well to prevent these cleavages, but primary cleavage sites are not difficult to identify if peptides are sequenced in large numbers [5]. The most important thing here in mass spectrometry-based peptidomics is not to introduce exogenous elements, which are likewise trapped in and eluted from a solid phase extraction cartridge and so likely to interfere with mass spectrometry of the biological samples. Protease inactivation by boiling and heat-denature tissues are well established [7, 8], but do not boil conditioned media, as it is impossible to rapidly raise internal liquid temperature.

11. It is recommended that collected media be immediately processed for peptide extraction without freezing. Longer storage of the media, either liquid or frozen, results in certain residues being chemically modified (e.g., methionine oxidation or pyroglutamylation). Such modifications of abundant peptide species will increase the complexity in LC-MS/MS, which means that MS/MS scans would not be performed on species present in trace amounts.
12. Store at −80 °C for at least 2 h and take care that the sample does not become melted before it attains a certain level of vacuum during lyophilization.
13. Dissolve the material with 20–50 μL of 5% ACN to allow for more than two LC runs. Analytes are in most cases acidic because of persisting TFA. This is inevitable and unfavorable to some physiological posttranslational modifications (e.g., the acylated group of ghrelin is removed during prolonged storage in an acidic solution).
14. C-terminal amidation is a posttranslational modification shared by many peptide hormones and bioactive peptides [9]. Supplementation with ascorbate is reported to enhance the conversion to C-terminal amidation [10]. An advantage of peptidomics over conventional methods is that it has the potential to find previously uncharacterized C-terminally amidated peptides. Some of these peptides have been found to have bioactive properties in functional assays [11–13]. Note, however, that we should carefully inspect MS/MS spectra to see if a seemingly high score reported by a search engine showed a result consistent with this C-terminal modification. First, make sure that your software correctly reported the monoisotopic precursor ion for the peptide while creating a peak list. It often happens that the software selects an incorrect isotopic peak as a monoisotopic ion, since the monoisotopic ion from a large peptide >3 kDa has a considerably lower intensity than other isotopic peaks and can be difficult to correctly identify. If the software selects an ion that is not the correct monoisotopic peak, the calculated precursor mass can differ from the real

molecular monoisotopic mass by 1 Da. Second, examine the deconvoluted MS/MS spectrum for y-ions, which retain the peptide C-terminus and so should indicate the C-terminal amide structure. If there is no match for y-ions for the putative amidated sequence, the search result is most likely to be false positive. Third, check that the precursor protein has a Gly in the position corresponding to the amide group—the enzyme peptidylglycine alpha-amidating monooxygenase requires a C-terminal Gly which is converted into the amide group [9].

15. The reconstituted solution may not turn basic because of a trace amount of TFA in the lyophilized material. If necessary, add 1 μL of unbuffered 1 M Tris solution.

## Acknowledgments

This work was supported in part by JSPS KAKENHI Grant No. 16K09404.

### References

1. Burgess AW, Metcalf D, Sparrow LG, Simpson RJ, Nice EC (1986) Granulocyte/macrophage colony-stimulating factor from mouse lung conditioned medium. Purification of multiple forms and radioiodination. Biochem J 235:805–814
2. Moseley JM, Kubota M, Diefenbach-Jagger H, Wettenhall RE, Kemp BE, Suva LJ, Rodda CP, Ebeling PR, Hudson PJ, Zajac JD et al (1987) Parathyroid hormone-related protein purified from a human lung cancer cell line. Proc Natl Acad Sci U S A 84:5048–5052
3. Mori M, Yamaguchi K, Abe K (1989) Purification of a lipoprotein lipase-inhibiting protein produced by a melanoma cell line associated with cancer cachexia. Biochem Biophys Res Commun 160:1085–1092
4. Sasaki K, Satomi Y, Takao T, Minamino N (2009) Snapshot peptidomics of the regulated secretory pathway. Mol Cell Proteomics 8:1638–1647
5. Tsuchiya T, Osaki T, Minamino N, Sasaki K (2015) Peptidomics for studying limited proteolysis. J Proteome Res 14:4921–4931
6. Jost MM, Budde P, Tammen H, Hess R, Kellmann M, Schulte I, Rose H (2005) The concept of functional peptidomics for the discovery of bioactive peptides in cell culture models. Comb Chem High Throughput Screen 8:767–773
7. Kangawa K, Matsuo H (1979) alpha-Neo-endorphin: a "big" Leu-enkephalin with potent opiate activity from porcine hypothalami. Biochem Biophys Res Commun 86:153–160
8. Svensson M, Boren M, Sköld K, Fälth M, Sjögren B, Andersson M, Svenningsson P, Andren PE (2009) Heat stabilization of the tissue proteome: a new technology for improved proteomics. J Proteome Res 8:974–981
9. Eipper BA, Bloomquist BT, Husten EJ, Milgram SL, Mains RE (1993) Peptidylglycine alpha-amidating monooxygenase and other processing enzymes in the neurointermediate pituitary. Ann N Y Acad Sci 680:147–160
10. Nikoulina SE, Andon NL, McCowen KM, Hendricks MD, Lowe C, Taylor SW (2010) A primary colonic crypt model enriched in enteroendocrine cells facilitates a peptidomic survey of regulated hormone secretion. Mol Cell Proteomics 9:728–741
11. Yamaguchi H, Sasaki K, Satomi Y, Shimbara T, Kageyama H, Mondal MS et al (2007) Peptidomic identification and biological validation of neuroendocrine regulatory peptide-1 and -2. J Biol Chem 282:26354–26360
12. Osaki T, Sasaki K, Minamino N (2011) Peptidomics-based discovery of an antimicrobial peptide derived from insulin-like growth factor-binding protein 5. J Proteome Res 10:1870–1880
13. Sasaki K, Osaki T, Minamino N (2013) Large-scale identification of endogenous secretory peptides using electron transfer dissociation mass spectrometry. Mol Cell Proteomics 12:700–709

# Chapter 4

# Mass Spectrometric Identification of Endogenous Peptides

**Mikel Azkargorta, Iraide Escobes, Ibon Iloro, and Felix Elortza**

## Abstract

Peptidomics is an emerging field focused in the analysis of endogenous peptides. Naturally occurring peptides are often endogenously produced protein fragments. Cleavage of precursor proteins by proteases generates peptides that may gain specialized functions not ascribed to their precursors, and which could reflect the state of a cell under certain physiological conditions or pathological processes.

Since peptides are found in complex matrices (e.g., serum, tear, urine, cerebrospinal fluid), they need to be isolated from the matrix and concentrated before they can be analyzed on mass spectrometry. This chapter describes methods for sample preparation prior to mass spectrometry analysis. In addition, different peptide fragmentation techniques are described which are complementary when analyzing naturally occurring peptides by liquid chromatography coupled online to tandem mass spectrometry.

**Key words** Mass-spectrometry, Fragmentation, Collision-induced dissociation, Electron-transfer dissociation, Higher energy collisional dissociation, Ethyl acetate, Peptidomics, Endogenous peptides

## 1 Introduction

Proteolytic processing is an important post-translational mechanism that increases the functional diversity of proteins, generating peptides that often gain specialized functions not ascribed to their precursors [1, 2]. These endogenous peptides have important roles in many biological processes [3]. Thus, their analysis is of great interest for a proper understanding of these events [4–6].

Mass spectrometry (MS) is the method of choice for a high-throughput analysis of peptides [7–9]. However, unlike tryptic peptides, MS analysis of endogenous peptides presents a number of drawbacks, including size heterogeneity and no predictable fragmentation patterns, among others, that hamper their proper fragmentation and detection [10]. In addition, the analysis of natural peptides coming from some biofluids, such as urine or tears, may present additional issues due to the presence of interfering substances (e.g., salts, pigments, carbohydrates, fatty acids). In this chapter we describe the general approach and cover in detail some of the methods we have used for the successful analysis of endogenous peptides (Fig. 1) [11].

Michael Schrader and Lloyd Fricker (eds.), *Peptidomics: Methods and Strategies*, Methods in Molecular Biology, vol. 1719, https://doi.org/10.1007/978-1-4939-7537-2_4, 

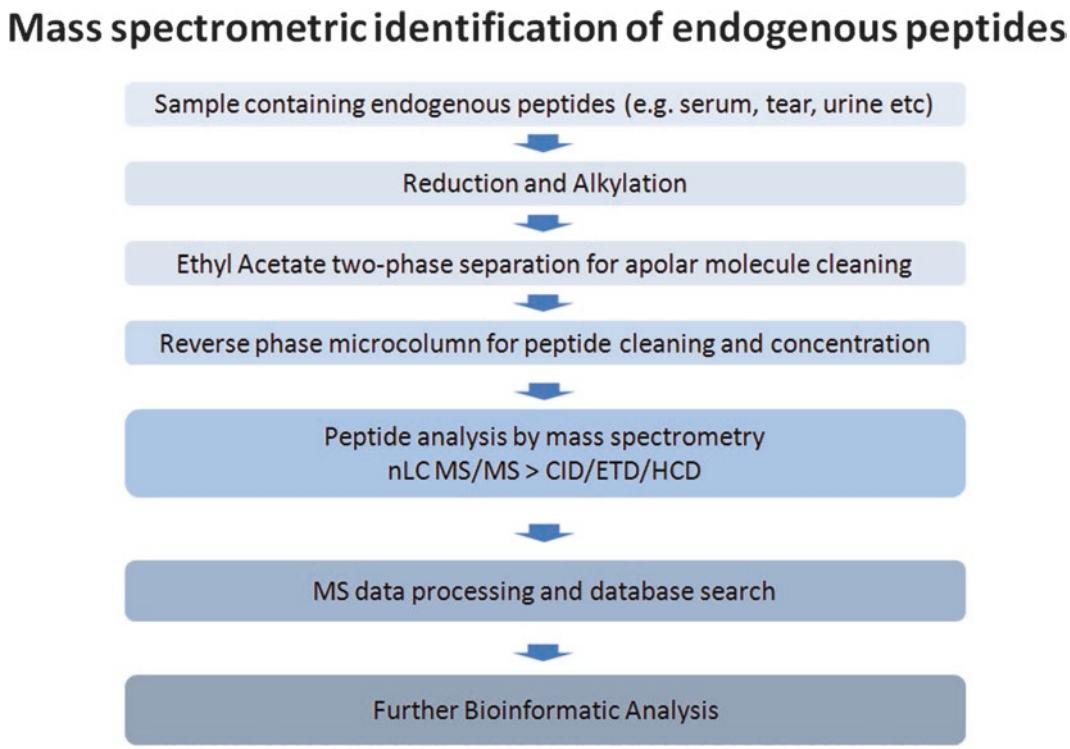



**Fig. 1** General scheme of the steps used in the mass spectrometric identification of endogenous peptides

## 2 Materials

*Reagents that are routinely used in our lab, and therefore considered as reliable and useful for the protocol provided, are recommended in the following section. Any other source for these materials might be as useful as the products mentioned. However, always check that all the solutions and buffers are prepared using MS-grade reagents.*

### *2.1 Sample Preparation*

1. DTT: Dithiothreitol (DTT) 1 M stock: Dissolve 0.154 g DTT in 1 mL Milli-Q water. Store at −20 °C in 10 μL aliquots.
2. DTT 10 mM: Dilute a 10 μL aliquot of DTT 1 M with 990 μL ammonium bicarbonate (AMBIC) 100 mM. Prepare fresh (*see* **Note 1**).
3. IAA: Iodoacetamide (IAA) 55 mM: Dissolve 5.14 mg of IAA in 1 mL AMBIC 100 mM. Prepare fresh.
4. Trifluoroacetic acid (TFA).
5. TFA 0.1%: Dissolve 1 mL TFA in 999 mL Milli-Q water.
6. Acetonitrile (ACN).
7. Speed Vac: Rotational-Vacuum-Concentrator (such as RVC 2-25, Christ).

8. Reverse Phase Microcolumns: Zip Tip with 0.6 μL resin C18 (such as Agilent OMIX C18 pipette tips, 2–10 μL).
9. Ethyl acetate LC-MS grade.
10. Optimal quality water LC/MS.
11. Ammonium bicarbonate (AMBIC) 99.5%.

### 2.2 MS Analysis

1. Formic acid (FA), product code 28905.
2. FA 0.1%: Dissolve 1 mL FA in 999 mL Milli-Q water (*see* **Note 2**).
3. Acetonitrile (ACN).
4. Glass vials.
5. Ultraperformance liquid chromatography (UPLC) system (such as NanoAcquity, Waters).
6. Reversed-phase UPLC column for peptide separation (such as Peptide BEH C18 nanoACQUITY 10K psi, 130 Å, 1.7 μm, 75 μm × 200 mm, 1/pkg, Waters).
7. Reversed-phase UPLC trapping column (such as Peptide Symmetry C18 Trap Column, 100 Å, 5 μm, 180 μm × 20 mm, 2G, V/M, 1/pkg, Waters).
8. Stainless steel mass spectrometry emitters (such as Thermo Scientific).
9. Mass spectrometer for large-scale proteomics (such as LTQ Orbitrap XL ETD Mass Spectrometer, Thermo Scientific).
10. Ultrasonic cleaning bath.

## 3 Methods

*Prepare all solutions using Milli-Q water (18 MΩ cm at 25 °C) and MS-grade reagents. Prepare and store all reagents at 4 °C (unless indicated otherwise) and follow all waste disposal regulations when disposing waste material.*

### 3.1 Sample Preparation

1. Add to the sample a solution of DTT to a final concentration of 5 mM in 100 mM AMBIC solution, and incubate at room temperature for 30 min (*see* **Note 3**).
2. Add a solution of IAA to a final concentration of 25 m. Incubate at room temperature for 30 min in the dark (*see* **Note 4**).
3. Add a solution of DTT to a final concentration of 5 mM in 100 mM AMBIC solution, and incubate at room temperature for 30 min.
4. Eventually, and depending on the origin of the sample, the use of ethyl acetate extraction can be a useful strategy for the purification

of peptides (*see* **Note 5**). For this purpose, an adapted protocol derived from the protocol described by Yeung et al. [12, 13] will be briefly described.

(a) *Ethyl acetate stock solution preparation*: Mix approximately 10 mL water and 80 mL of ethyl acetate (*see* **Note 6**). Shake the contents vigorously for 1 min so that the water breaks into tiny droplets in the ethyl acetate and let the bottle rest on the work bench for 10 min. Repeat the procedure two more times. Allow the bottle to rest until a clear layer of ethyl acetate forms on top of water (at least 30 min). Use the clear upper layer of ethyl acetate for extraction.

(b) *Ethyl acetate separation:* Bring the sample to a final volume of 100 μL using Milli-Q water. Add 1 mL of water-saturated ethyl acetate to the peptide solution. Vortex vigorously for 1 min and centrifuge at 15,600 × *g* for 15 s. Aspirate and discard the upper ethyl acetate layer in a laminar flow cabinet (*see* **Note 7**). Repeat this procedure five times.

5. Dry vacuum the samples and resuspend them in 10 μL FA 0.1%. Sonicate 5 min in the ultrasonic cleaning bath.
6. Reversed-phase separation protocol: Flush the tip with 10 μL ACN two times. Flush the tip with 10 μL FA 0.1% two times. Load the sample by taking and discarding 10 μL. Repeat this last step ten times (*see* **Note 8**). Wash the peptides once with 10 μL FA 0.1% (*see* **Note 9**). Elute the peptides using a solution containing [70:30] ACN/3% FA. For this purpose, take 10 μL of the eluting solution and release them in a clean microfuge polypropylene tube. Pipette up and down this volume ten times in order to release the sample completely from the tip.
7. Dry vacuum the samples. Samples can be stored at −20 °C.

### 3.2 Sample Load onto the Chromatographic System

1. Resuspend the sample in 10 μL 3% FA. Sonicate for 5 min in the ultrasonic cleaning bath.
2. Put the resuspended sample in a vial and load the sample into the mass spectrometer (*see* **Note 10**).
3. Peptides are separated on the UPLC system using a combination of reversed-phase trapping and separation columns.
4. The recommended chromatographic gradient is performed at a flow rate of 0.3 μL/min that includes the following steps (*see* **Note 11**):

| Time (min) | A (%) | B (%) | Flow (mL/min) |
|---|---|---|---|
| 0 | 97 | 3 | 0.3 |
| 60 | 60 | 40 | 0.3 |
| 61 | 15 | 85 | 0.3 |

| Time (min) | A (%) | B (%) | Flow (mL/min) |
|---|---|---|---|
| 70 | 15 | 85 | 0.3 |
| 72 | 97 | 3 | 0.3 |
| 90 | 97 | 3 | 0.3 |

A: FA 0.1% in $H_2O$
B: FA 0.1% in ACN

### 3.3 MS Acquisition of the Endogenous Peptides

Given the particular properties of endogenous peptides, special requirements for their fragmentation might be needed. Unlike trypsin-digested protein fragments, C-termini of naturally occurring peptides are not restricted to certain residues [10]. In addition, natural peptides might be larger than tryptic peptides and present large numbers of basic amino acids within their sequence. All this not only increases the heterogeneity of the sample, but also changes the physical properties of the endogenous peptides, affecting their fragmentation and further identification [14] (*see* **Note 12**).

Therefore, when available, alternatives to the routinely used collision-induced dissociation (CID) fragmentation, such as higher energy collisional dissociation (HCD) and electron transfer dissociation (ETD), usually provide additional information on sample composition [10, 11, 15]. Moreover, specific parameters for these methods might be necessary for the successful acquisition of data.

However, in our experience and depending on sample origin and nature, default acquisition methods routinely used for the analysis of tryptic peptides can also be successfully used for the analysis of naturally occurring peptides. Thus, a preliminary characterization of the sample using different fragmentation methods might provide very valuable information in order to define the best strategy for its analysis (*see* **Note 13**).

Detailed information on the acquisition parameters used for the identification of naturally occurring peptides in an LTQ Orbitrap XL ETD is provided in Tables 1, 2, 3, 4, and 5. Parameters regarding survey scan (Table 1), peptide selection for MS/MS analysis (Table 2), and HCD, and ETD fragmentation (Tables 3, 4, and 5, respectively) are summarized. Parameters for both the default (used for tryptic peptides) and natural peptide-specific workflows are provided, since both of them have been shown to be useful for the analysis of naturally occurring peptides in our lab (*see* **Note 14**).

*N most intense*: Selection of the n most intense peptides for triggering an MS/MS event. *NCE*: Normalized collision energy. *Act Q*: Activation Q. *Act. Time*: Activation time. *AGC target* = Automatic gain control target. *Max. IT*: Maximum injection time.

### 3.4 MS Data Analysis

Once acquired, each of the acquisition methods (default and dedicated) may require different parameters for the identification of the peptides. Table 6 compiles some of the most relevant parameters used for these database searches, such as those performed using

**Table 1**
**MS method parameters for the survey scan of tryptic peptides (default) and natural peptides (Nat pepts)**

| Acquisition method | *m/z* range | MS used | Resolution | AGC target | Max. IT (ms) |
|---|---|---|---|---|---|
| Default | 400–2000 | Orbi | 30,000 | 5.E+05 | 150 |
| Nat pepts | 400–2000 | Orbi | 100,000 | 7.E+05 | 150 |

*AGC target* Automatic gain control target, *Max. IT* Maximum injection time

**Table 2**
**Parameters for the selection of tryptic peptides (default) and natural peptides (Nat pepts) for selection to fragmentation**

| Acquisition method | Min. sign. | Ch. st. sel. | Isol. wdw. | Dyn. excl. time (s) | Dyn. excl. N |
|---|---|---|---|---|---|
| Default | 500–1000 | 2 and 3[a] | 2 *m/z* units | 60 | 1 |
| Nat pepts | 1000 | 2 or greater | 5 *m/z* units | 2400 | 2 |

*Min. sign.* minimum signal required for triggering a MS/MS event, *Ch. st. sel.* charge state selection for MS/MS events, *Isol. wdw.* isolation window, *Dyn. excl. time* dynamic exclusion time, *Dyn. excl. N* number of MS/MS acquisitions before subjection to dynamic exclusion
[a]Charge states 2 and 3 are specifically selected for CID and HCD, whereas greater charge states are also selected for ETD

**Table 3**
**CID parameters for the MS/MS scans of tryptic peptides (default) and natural peptides (Nat pepts)**

| Acquisition method | N most intense | MS used | Resolution | NCE | Act Q | Act. time (ms) | AGC target | Max. IT (ms) |
|---|---|---|---|---|---|---|---|---|
| Default | 6 | IT | – | 35 | 0.25 | 30 | 5.00E+04 | 300 |
| Nat pepts | 3 | Orbi | 100,000 | 35 | 0.25 | 30 | 2.00E+05 | 1000 |

**Table 4**
**HCD parameters for the MS/MS scans of tryptic peptides (default) and natural peptides (Nat pepts)**

| Acquisition method | N most intense | MS used | Resolution | NCE | Act Q | Act. time (ms) | AGC target | Max. IT (ms) |
|---|---|---|---|---|---|---|---|---|
| Default | 4 | Orbi | 15,000 | 35 | 0.25 | 30 | 2.00E+04 | 100 |
| Nat pepts | 3 | Orbi | 100,000 | 35 | 0.25 | 30 | 2.00E+05 | 1000 |

*N Most intense* Selection of the n most intense peptides for triggering an MS/MS event, *NCE* Normalized collision energy, *Act Q* activation Q, *Act. Time* activation time, *AGC target* Automatic gain control target, *Max. IT* Maximum injection time

**Table 5**
**ETD parameters for the MS/MS scans of tryptic peptides (default) and natural peptides (Nat pepts)**

| Acquisition method | N most intense | Suppl. Act. | Charge st. ETD time | AGC target | Max. IT (ms) | Fluor. AGC target | Fluor. Max IT (ms) |
|---|---|---|---|---|---|---|---|
| Default | 3 | Yes | Yes | 5.00E+04 | 100 | 5.00E+05 | 500 |
| Nat pepts | 2 | Yes | Yes | 5.00E+05 | 1500 | 1.00E+06 | 1000 |

*N Most intense* Selection of the n most intense peptides for triggering an MS/MS event, *Suppl. Act.* supplemental activation, *Charge st. ETD time* charge state-dependent ETD time, *AGC target* automatic gain control target, *Max. IT* Maximum injection time, *Fluor. AGC target* reagent automatic gain control target, *Max. IT* reagent maximum injection time

**Table 6**
**Parameters for the Mascot search of the acquisitions obtained using the default and dedicated methods**

| Parameter | Default | Nat pepts |
|---|---|---|
| Precursor mass tol. | 10 ppm | 10 ppm |
| Fragment mass tol. | 0.5 Da | 0.05 Da |
| Dynamic mods. | Oxidation (M)<br>Acetyl (N-term)<br>Amidated (C-term)<br>Gln → pyro-Glu (N-term Q),<br>Glu → pyro-Glu (N-term E) | |
| Fixed mods. | Carbamidomethyl (C) | |
| Deconvolution | No | Yes |

Proteome Discoverer 1.4. Importantly, the data obtained using the dedicated method will need a deconvolution of spectra for a proper identification (*see* **Notes 15** and **16**).

Figure 2 shows the two different workflows used in Proteome Discoverer 1.4. Each workflow consists of different nodes that provide different steps for a successful data processing and identification of the peptides (*see* **Notes 17** and **18**).

### 3.5 Results

Figure 3 shows the fragmentation spectrum of the whole thymosin beta-4 protein identified using a CID-based dedicated method in human tears. This MS/MS event led to the successful identification of the processed form of thymosin beta 4 (except the deleted initial methionine) with an ion score of 147.

The complementarity of different fragmentation methods when applied to the analysis of naturally occurring peptides has been shown elsewhere [10, 11, 15]. In this regard, Fig. 4 shows unpublished data on the analysis of human natural peptides coming from urine samples using both CID and ETD. Ten different human urine

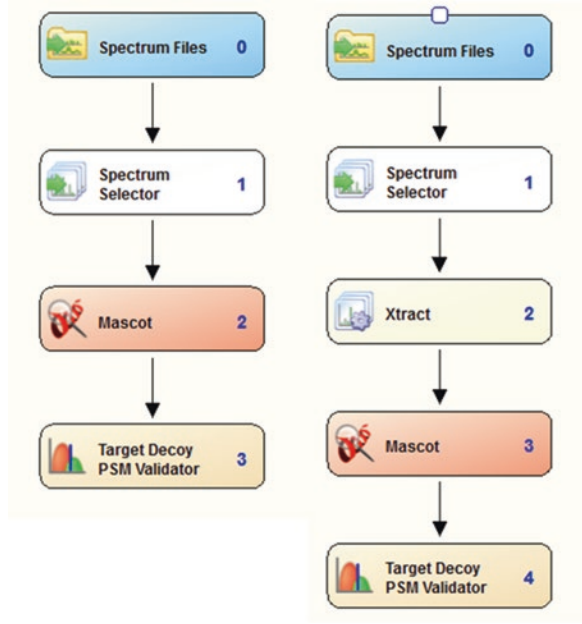



**Fig. 2** Mass spectrometry data processing workflows for the default acquisition (left) and dedicated acquisition (right) within Proteome Discoverer

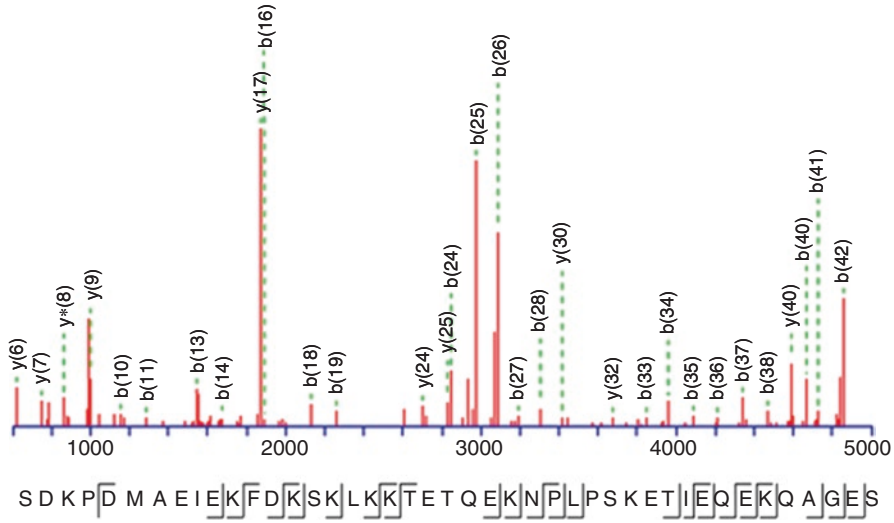



**Fig. 3** Spectrum for the identification of thymosin beta-4 protein in human tears using the dedicated CID-based method from Table 3

samples were prepared using the method described in this chapter. These samples were individually analyzed using a CID-based default method. Then, in order to increase sample coverage, these samples were pooled and analyzed twice using an ETD-based default method. Charge states greater than 2 were specifically selected for

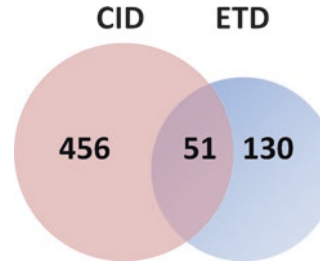



**Fig. 4** Overlap between the identifications provided by CID- and ETD-based methods for natural peptides coming from human urine samples. Ten samples were analyzed independently using a CID-based method. Samples were pooled and analyzed twice using the ETD-based method. Despite the difference in the effort made using each of the methods, ETD contributes substantially to the total number of peptides identified

this acquisition. As the Venn diagram reveals, despite the difference in the total number of runs, the use of ETD contributes substantially to the total number of identifications, with a total of 130 peptides (20% of the total) uniquely identified by this method. This example clearly illustrates the advantage of using alternative fragmentation methods for the analysis of natural peptides.

### *3.6 Bioinformatic Analysis*

The following section pinpoints some of the analyses that can be performed with the protein and the peptide lists obtained after the MS analysis.

The function of the whole protein and the function of the naturally occurring peptides arising from it may not necessarily be the same, but the group of identified proteins can be further characterized through a Gene Ontology enrichment analysis. This analysis will provide a brief description of the processes the identified proteins are involved in. For this purpose, the free online DAVID functional annotation tool (https://david.ncifcrf.gov/) can be a good option [16, 17]. Typically enrichment in Biological Processes (BP), Cellular Component (CC), and Molecular Function (MF) is analyzed, but the program provides of other interesting processes, such as KEGG Pathways, that can be analyzed if desired. Processes enriched with a Fisher's exact test $p$ value <0.05 are considered as significantly enriched, although more restrictive cutoffs, such as <5% FDR, can also be used for a more reliable analysis.

In addition to general analyses, there are a couple of analyses that may provide valuable information about protein processing and characteristics of the peptides themselves. Antimicrobial activity can be determined using the AMP prediction tools in the CAMP Database (http://www.camp.bicnirrh.res.in/) [18, 19]. Three machine learning algorithms for prediction of antimicrobial activity are available: Random Forest (RF), Support Vector Machines (SVM), and Discriminant Analysis (DA). Prediction

results can be displayed as the sum of the probability scores of each classification model, corrected by their accuracy (0.93, 0.92, and 0.88 for RF, SVM, and DA, respectively). Highest scoring peptides can be considered as putative antimicrobial peptides (*see* **Note 19**). This putative activity should be further demonstrated in vitro.

PROSPER webserver (https://prosper.erc.monash.edu.au/webserver.html) provides an in silico prediction of protease substrates and their cleavage sites in the list of protein sequences provided as a FASTA format. Twenty-four different protease types, covering four major protease families—aspartic (A), cysteine (C), metallo (M), and serine (S)—are considered in this analysis, providing information on the possible proteases that are responsible for the formation of the observed natural peptides [20]. However, considering that there are hundreds of proteases and peptidases in mammals, this database is not comprehensive. Most importantly, the website lacks many enzymes that generate peptides in various tissues, such as the proprotein convertases that are responsible for the initial cleavage of the majority of neuropeptide precursors.

## 4 Notes

1. AMBIC solution does not need pH adjustment (should be around pH 8).
2. Do this in glassware instead of plasticware. It is recommended to use dedicated glassware for this purpose so no detergents or any other contaminants interfere with the MS analysis.
3. Reduction and alkylation of the sample are necessary for the proper analysis of peptides containing disulfide bonds. Disruption and inactivation of the disulfide bridges formed by Cys residues help this analysis and have been shown to increase sample coverage when analyzing natural peptides.
4. DTT/IAA ratio may vary depending on the protocol used, but a relation such as the one proposed has been shown to work in our laboratory.
5. This protocol was originally intended to be used for the elimination of detergents in peptide samples. However, given that some biofluids may contain apolar contaminants such as lipids, its use may help obtaining a cleaner sample. A preliminary trial may be useful for a proper determination of its potential advantages and disadvantages, since sample losses may be expected.
6. Accurate measurement of the solution volumes is not required here.
7. Care should be taken when aspirating the sample. Avoid touching the aqueous phase with the pipette tip.

8. Sample recovery may increase with the number of uptakes. However, ten times is the recommended number. Please take care of the zip-tip load limit (2 μg for the product referenced in this chapter).
9. The number of washes can be increased if desired for especially dirty samples. However, washing the samples too much is not recommended since part of the sample might get lost. Optimization of the protocol with a mock sample can help defining the steps to be performed.
10. Sample resuspension volume and amount to be loaded may depend on the starting material.
11. The columns and gradient are illustrative. However, a linear gradient followed by a washing step and an equilibration step are needed as part of the protocol. Length of each phase may depend on sample load, sample complexity, and/or column length, among others. Adaptation to the system and optimization are therefore needed.
12. In addition to these inherent physicochemical particularities, most of the search algorithms have been developed for their use with tryptic peptides, limiting even more the successful analysis of non-tryptic samples.
13. We strongly recommend this preliminary step, since different methods may provide a very different output. A preliminary characterization may help defining the best strategy possible. Use one or a few samples to define if the default or the dedicated methods provide a larger number of identifications. There might be some natural peptide samples where a default method may work better than the dedicated method.
14. Further optimization of the parameters provided might be needed depending on the spectrometer used and the sample under analysis.
15. Proteome Discoverer provides Xtract as deconvolution software, but different options can be used for this purpose.
16. A non-deconvoluted search can be used in order to determine the original charge state of the peptides identified. These searches will provide no identification, but will assign a charge state to a certain scan number (MS/MS event). Subsequently, a correlation between the charge state and the peptide identified in that scan can be easily performed.
17. Default parameters are used except for the ones mentioned before.
18. A decoy validation of the results is recommended, although depending on the sample complexity and number of MS/MS events and their successful identification rate, the estimation might not be fully accurate. Therefore, again, a preliminary

characterization of the sample is recommended for a convenient decision.

19. A possible cutoff for an initial selection of putative antimicrobial peptides can be 0.75 over the weighed probability, but the selection of the highest scoring results is recommended.

## Acknowledgments

This work has been carried out with the financial support of the Basque Government, Bizkaia County, and ETORTEK and ELKARTEK programs.

### References

1. Hallberg M (2015) Neuropeptides: metabolism to bioactive fragments and the pharmacology of their receptors. Med Res Rev 35(3):464–519
2. Schulz-Knappe P, Schrader M, Zucht HD (2005) The peptidomics concept. Comb Chem High Throughput Screen 8:697–704
3. Adermann K, John H, Ständker L et al (2004) Exploiting natural peptide diversity: novel research tools and drug leads. Curr Opin Biotechnol 15(6):599–606
4. Schulte I, Tammen H, Selle H et al (2005) Peptides in body fluids and tissues as markers of disease. Expert Rev Mol Diagn 5:145–157
5. Hathout Y (2007) Approaches to the study of the cell secretome. Expert Rev Proteomics 4:239–248
6. Bauça JM, Martínez-Morillo E, Diamandis EP (2014) Peptidomics of urine and other biofluids for cancer diagnostics. Clin Chem 60:1052–1061
7. Schrader M, Schulz-Knappe P (2001) Peptidomics technologies for human body fluids. Trends Biotechnol 10:S55–S60
8. Ramström M, Bergquist J (2004) Miniaturized proteomics and peptidomics using capillary liquid separation and high resolution mass spectrometry. FEBS Lett 567:92–95
9. Secher A, Kelstrup CD, Conde-Frieboes KW et al (2016) Analytic framework for peptidomics applied to large-scale neuropeptide identification. Nat Commun 7:11436
10. Sasaki K, Osaki T, Minamino N (2013) Large-scale identification of endogenous secretory peptides using electron transfer dissociation mass spectrometry. Mol Cell Proteomics 12:700–709
11. Azkargorta M, Soria J, Ojeda C, Guzmán F et al (2015) Human basal tear peptidome characterization by CID, HCD, and ETD followed by in silico and in vitro analyses for antimicrobial peptide identification. J Proteome Res 14:2649–2458
12. Yeung YG, Nieves E, Angeletti RH et al (2008) Removal of detergents from protein digests for mass spectrometry analysis. Anal Biochem 382:135–137
13. Yeung YG, Stanley ER (2010) Rapid detergent removal from peptide samples with ethyl acetate for mass spectrometry analysis. Curr Protoc Protein Sci Chapter 16:Unit 16.12
14. Tabb DL, Huang Y, Wysocki VH et al (2004) Influence of basic residue content on fragment ion peak intensities in low-energy collision-induced dissociation spectra of peptides. Anal Chem 76:1243–1248
15. Shen Y, Tolić N, Xie F et al (2011) Effectiveness of CID, HCD, and ETD with FT MS/MS for degradomic-peptidomic analysis: comparison of peptide identification methods. J Proteome Res 10:3929–3943
16. Huang d W, Sherman BT, Lempicki RA (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc 4:44–57
17. Huang d W, Sherman BT, Lempicki RA (2009) Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res 37:1–13
18. Thomas S, Karnik S, Barai RS et al (2010) CAMP: a useful resource for research on antimicrobial peptides. Nucleic Acids Res 38:D774–D780
19. Waghu FH, Gopi L, Barai RS et al (2014) CAMP: collection of sequences and structures of antimicrobial peptides. Nucleic Acids Res 42:D1154–D1158
20. Song J, Tan H, Perry AJ et al (2012) PROSPER: an integrated feature-based tool for predicting protease substrate cleavage sites. PLoS One 7:e50300

# Chapter 5

# Bioinformatics for Prohormone and Neuropeptide Discovery

**Bruce R. Southey, Elena V. Romanova, Sandra L. Rodriguez-Zas, and Jonathan V. Sweedler**

## Abstract

Neuropeptides and peptide hormones are signaling molecules produced via complex post-translational modifications of precursor proteins known as prohormones. Neuropeptides activate specific receptors and are associated with the regulation of physiological systems and behaviors. The identification of prohormones—and the neuropeptides created by these prohormones—from genomic assemblies has become essential to support the annotation and use of the rapidly growing number of sequenced genomes. Here we describe a methodology for identifying the prohormone complement from genomic assemblies that employs widely available public toolsets and databases. The uncovered prohormone sequences can then be screened for putative neuropeptides to enable accurate proteomic discovery and validation.

**Key words** Neuropeptide, Prohormone, Homology, Bioinformatics, Cleavage, Gene prediction

## 1 Introduction

The increased speed and decreased cost of genomic sequencing has revolutionized neuropeptide identification from primarily experiment-driven discovery to bioinformatics-driven genomic searches. Currently, the number of sequenced genomes exceeds the number of species that have at least one experimentally confirmed prohormone and neuropeptide. The advances in genomics also enable neuropeptide discovery to extend beyond the specific experimental system being investigated, such as a defined tissue or developmental stage. We have developed a bioinformatics toolset that can be used to uncover neuropeptides in a broad range of organisms, and to predict multiple putative novel neuropeptides that have not been previously characterized, even in well-studied species.

The challenge in using genomic data to discover neuropeptides stems from the fact that one neuropeptide gene typically encodes multiple neuropeptides within a larger precursor protein, which is referred to as a prohormone. Prohormones undergo post-translational enzymatic cleavage and further chemical modifications, resulting in a

Michael Schrader and Lloyd Fricker (eds.), *Peptidomics: Methods and Strategies*, Methods in Molecular Biology, vol. 1719, https://doi.org/10.1007/978-1-4939-7537-2_5, 

set of shorter neuropeptides [1], all products of the same gene. The bioinformatics approach to neuropeptide detection follows this biological process, starting with the identification of the prohormone sequence in the genome. The characteristic features of a complete prohormone protein are (1) a signal peptide at the N-terminus that enables the co-translational translocation of the protein into the secretory pathway [2], and (2) at least one cleavage site that is recognized by endopeptidases that delimits the neuropeptide sequence.

Our efforts to identify prohormones and neuropeptides across taxa and species (e.g., human, mouse, rat, cattle, pig, chicken, song bird, honey bee, flour beetle, California sea slug, fish, camelids [3–10]) have enabled us to identify the bioinformatics steps that are critical for the accurate identification and annotation of prohormones in genomes. These steps provide reliable prohormone gene, protein sequence, and neuropeptide prediction across sequenced genomes, regardless of the degree of experimental validation. Our web application NeuroPred [11, 12] expedites the bioinformatics tasks and has been successfully used to predict the putative neuropeptides nested in these prohormone sequences [7–9, 11, 13]. We have combined the sequence annotations obtained using NeuroPred with other bioinformatics tools into the web portal PepShop [14] to facilitate neuropeptide and prohormone discovery and usage in a wide range of species. Together they provide important resources for the neuropeptide research community.

A major undertaking of the bioinformatics approach for detecting neuropeptides is the identification of candidate prohormones in a target species. The availability of the species genome is not sufficient by itself for accurate neuropeptide identification. Many prohormone genes are predicted using automatized methods to annotate the genome assembly. However, we have demonstrated that some of these prohormone gene predictions encompass inaccurate features due to incorrect exon predictions, and some prohormones have been completely missed. In addition, few neuropeptides from predicted proteins have been manually or empirically confirmed. Accurate prohormone gene prediction also benefits from comparative genomics and recognition of sequence similarities between gene structures of evolutionarily related species because neuropeptides are highly conserved across the metazoans [15].

The prohormone and peptide sequences predicted by our bioinformatics approach (Fig. 1) are readily available to support the identification of novel peptide sequences that have been experimentally detected. The symbiotic integration of bioinformatics predictions with robust mass spectrometry (MS) sequencing enables identification of candidate and novel peptides [16]. Sensitive and specific identification of peptides and their PTMs is essential to the advancement of neuropeptide research.

We demonstrated the benefit of integrating our biologically driven informatics approach with MS analytics to accurately predict prohormones and neuropeptides in the particularly challenging

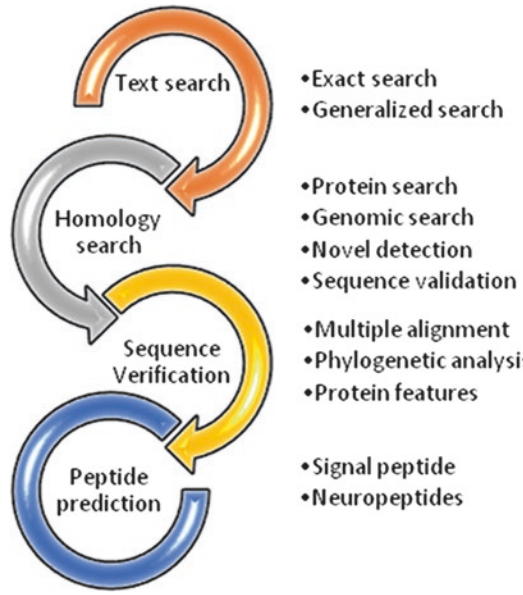



**Fig. 1** Outline of the bioinformatic steps leading to the accurate identification and annotation of a prohormone

genome of *Astatotilapia burtoni*, a teleost fish [4]. The evolutionary ancestor of *A. burtoni* underwent a whole genome duplication event after the divergence of tetrapods and teleosts [17]. Our systematic bioinformatics approach proved to be well suited for sifting through sequence duplications and rearrangements to harvest useful sequence information from divergent species. Throughout this chapter, we provide examples of our approach by describing the individual steps we used in that work [4] to accurately identify and annotate the *A. burtoni* parathyroid hormone family.

## 2 Materials

### *2.1 Genomic Data of Desired Species*

The bioinformatic identification of prohormones integrates genomic, transcriptomic, and proteomic information from the genomic data of the target species. This genomic data include the nucleic assembly, which is typically composed of assembled chromosomes, and scaffolds (genomic sequence interspersed with sections of unknown nucleotides) or contigs (contiguous lengths of known genomic sequence). Transcriptomic data include RNA and expressed sequence tag (EST) sequences, and proteomic data that mainly include sequences predicted by automated tools that are used to annotate the genome assembly. The National Center for Biotechnology Information (NCBI) [18] and ENSEMBL [19] are examples of public repositories that archive genome assemblies and genome, transcriptome, and protein sequences. These public repositories are regularly updated to include the most updated revisions of the genome assemblies and sequence predictions.

### 2.2 Prohormone Protein Sequences of Phylogenetically Close Species

Evolutionary relationships are shaped in part by molecular similarities at both the DNA and protein levels. Therefore, the prohormone protein sequences of the phylogenetically related reference species are expected to be structurally and functionally similar to the target species. Many of these sequences can also be retrieved from the Universal Protein Resource (UniProt) [20] and NCBI RefSeq [21] databases, albeit in a less streamlined manner since these databases are not dedicated to prohormone sequences. These online resources serve for initial prohormone annotation in the target species using cross-species homology. We have built libraries of annotated prohormones for a wide variety of species, including multiple mammalian species, insects, birds, and mollusks, available at the NeuroPred [22] and PepShop [14] websites (*see* **Note 1**). There are other online databases [23–25] that also include prohormone information.

1. UniProt: https://www.uniprot.org/.
2. RefSeq: https://www.ncbi.nlm.nih.gov/refseq/.
3. NeuroPred: http://stagbeetle.animal.uiuc.edu/neuropred/sequences/Sequencedata.html.
4. PepShop: http://stagbeetle.animal.uiuc.edu/pepshop.

### 2.3 Bioinformatic Tools

Essential tools for finding cross-species sequence similarities include routines and software for sequence alignment. Foremost examples of these tools are the Best Linear Alignment Software Tool (BLAST) [26] for pair-wise alignment of a sequence to database sequences, Clustal Omega [27] for multiple sequence alignment, and GeneWise [28] for gene prediction from nucleotide sequences. These web-based tools are highlighted because they are regularly updated, do not require local installation, and outputs can easily be shared, reproduced, and replicated. Additional desirable features of bioinformatic toolsets are enhanced functionality that simplify the steps and allow easy transfer of the output from one program to serve as input for another program. The bioinformatics demonstration described below uses NCBI databases and tools because of their comprehensive scope and ease of integration. These resources will be complemented with Clustal Omega, GeneWise, SignalP, and NeuroPred.

1. Blast: https://blast.ncbi.nlm.nih.gov/Blast.cgi.
2. Clustal Omega: http://www.clustal.org/.
3. GeneWise: http://www.ebi.ac.uk/Tools/psa/genewise/.
4. SignalP: http://www.cbs.dtu.dk/services/SignalP/.
5. NeuroPred: http://neuroproteomics.scs.illinois.edu/neuropred.htm.

### 2.4 Spreadsheet and Text Editor Applications to Record Findings

Gene/protein databases provide important information, notably prohormone gene names, gene locations, and complementary database accession number identifiers that need to be stored for each prohormone gene. The application selected to hold this sequence information must support the addition or update of fundamental information and addendum notes, multifactorial search, and reordering of current information, all within a table format. These features enable the rapid discovery of previously located prohormone genes and entry of new genes and similar genes. Most open-source or commercial spreadsheet applications (e.g., LibreOffice calc and Microsoft Excel) provide these abilities and offer satisfactory functions, including adding new columns and sorting by features such start and end positions of sequence matches or alignments to different genome regions (*see* **Note 2**). A text editor (such as LibreOffice writer or Microsoft Word) primarily enables the accumulation of protein sequences in a table format that can be used for other programs or searched for specific short sequences.

## 3 Methods

The first step in neuropeptide discovery is the identification of candidate prohormone genes and retrieval of the corresponding protein sequences. Next, thorough confirmation of these sequences is required via additional resources, such as other publically available sequences from desired species and trace archives (raw, short sequences generated by a specific genome project) and comparison of sequences across related species. Finally, the signal peptide and putative neuropeptides contained in the predicted prohormone sequence are calculated. A detailed description of these bioinformatic steps follows.

### 3.1 Create a List of Putative Prohormones

Select the phylogenetically closest species that has extensive characterization of the prohormone complement as the reference species. Use this reference species to generate an initial list of candidate prohormone that will be searched in the target species (*see* **Note 3**).

Arrange the prohormones from the reference species in the rows of a spreadsheet application table. The columns of this table will record the sequence accession numbers of the reference and target species, information on the target genome, including location of the reference sequence matches in the target genomic assembly, and miscellaneous details, e.g., the completeness of the reference sequence in the target genome assembly.

Each candidate in the list of prohormones from the reference species is evaluated in the target species following the bioinformatics steps shown in Fig. 1. Keep in mind that there is no single ideal

approach to use when evaluating the candidate list; therefore, the candidate prohormones are processed individually. Searches can benefit from concurrent evaluation of all the sequences within a prohormone family because of the high sequence similarity among prohormone family members. This sequential processing of the candidate list should be amended to accommodate new putative prohormones that are identified in subsequent bioinformatics steps. Novel prohormones identified in the target species and incorporated in the candidate list should be searched in the reference sequence in an iterative manner.

Putative matches to the candidate sequences identified in the target species must be recursively evaluated before moving to the next candidate prohormone in the list. Information on each putative match, such as gene(s) identifier, genome location(s), significance threshold of detection (e.g., e-value, percentage identity), and comments, is added to the table (*see* **Note 4**). The evaluation of the best match (i.e., lower e-value) between the candidate and target sequences must be followed by the evaluation of weaker matches (within 25% of the best match). This extended investigation permits identification of prohormones in the target species that originated from duplication events. Additional matches identified in the target species must be included in the candidate list, and the table entry should be completed with any new information. A candidate prohormone sequence is considered complete once all of the bioinformatics steps in Fig. 1 have been successfully completed.

Some candidate prohormones from the reference species will not have matches in the target species. An example of this is the mammalian relaxin 1 gene that has been lost in ruminants [6]. In other cases the lack of a candidate on the target genome may be associated with reference-specific duplication events after the branching of the reference and target species. An example of this is the insulin-like 4 gene that resulted from a primate-only gene duplication event [29]. Any candidates without a match should be denoted as not being present before moving on to the next candidate in the list.

Our bioinformatics approach was used to identify the *A. burtoni* members of the parathyroid hormone (PTH) family. Consistent with the *Homo sapiens* (human) PTH family [30], the fish PTH family contains at least three members: parathyroid hormone 1 and 2 (PTH1 and PTH2), and parathyroid hormone-related protein (PTHLH) [31, 32]. Therefore, the initial candidate list for *A. burtoni* had three entries: PTH1, PTH2, and PTHLH. This list does not account for the teleost whole-genome duplication because it is uncertain if any of the duplicated prohormone genes resulting from whole-genome duplication have been retained in *A. burtoni*.

### 3.2 Identification of Putative Prohormones

#### 3.2.1 Text Search for Prohormone and Neuropeptide Names

Searches for candidate prohormones include text searches for names and gene symbols of the prohormone genes that have already been annotated in the target species. The existing annotation could be the output of the target genome project or other research efforts (e.g., evolutionary studies of specific prohormone genes). A text search in a gene-centric database, such as NCBI Gene [33], is recommended as the first step rather than a search in a protein database because gene databases are typically more complete and maintain more information at both the nucleotide and amino acid levels than protein databases.

Effective text searches use the common prohormone name or neuropeptide, followed by one or more unique words that are part of the prohormone or neuropeptide name (*see* **Note 5**). Text matches are then recorded in the candidate prohormone table. These details may expedite the entire process since some of this information is identified in the subsequent steps and can assist in the discrimination between duplicated genes and homologous genes from the same prohormone family in the target species.

In our example, a text search was conducted within the NCBI Gene database using the phrase: "parathyroid hormone"[title] AND "Astatotilapia burtoni"[orgn]. Here the "[title]" and "[orgn]" are used to limit the search to specific words in the descriptive text, such as gene name and species, respectively. Our phrase identifies genes that have the compound word "parathyroid hormone" in the descriptive text and are found in the target species, "Astatotilapia burtoni." In addition to finding PTH1, this text search also identified the other genes, PTH2 and PTHL, because the phrase "parathyroid hormone" is part of the gene names, as well as part of the name of the prohormone family.

The outcome of this search included three prohormone genes and three prohormone receptors (Fig. 2), and is summarized in Table 1. An equivalent search within the NCBI protein database [34] only identified two proteins. A more general search for the phrase "parathyroid"[title] AND ("Astatotilapia burtoni"[orgn]) identified 6 proteins corresponding to 3 receptor proteins, PTH1, and the 2 PTHLH protein isoforms. PTH2 was not identified in the previous text search because the database entry referred to a former name, "tuberoinfundibular peptide of 39 residues" (TIP39).

#### 3.2.2 Homology Search Against Protein Databases and Genome Assembly Databases

The initial search of a protein sequence in a protein database uses BLASTP, a version of BLAST that searches protein sequence databases using a protein sequence as a query. The candidate prohormone protein sequence from the reference species is searched against the protein database of the target species. Consideration of BLASTP matches that have an e-value $\leq 10$ using the default BLOSSUM 62 substitution matrix enables the identification of matches that have conserved, albeit small,

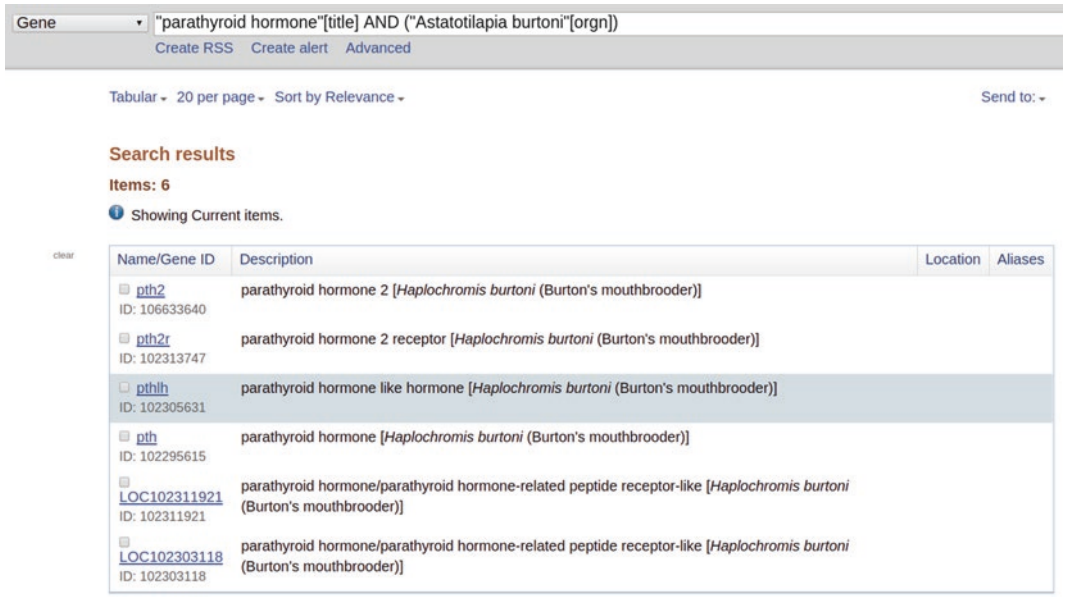



**Fig. 2** Result of a text search for parathyroid hormone family genes in *Astatotilapia burtoni* within the NCBI Gene database where *Haplochromis burtoni* is the former name of *A. burtoni*

**Table 1**
**Summary of a text-based search for parathyroid hormone family genes in *Astatotilapia burtoni* within the NCBI Gene database**

| | | Genomic information | | |
|---|---|---|---|---|
| **Prohormone** | **Gene ID** | **NCBI accession number of Contig** | **Position with contig** | **NCBI protein accession number** |
| PTH1 | 102295615 | NW_005179605.1 | 11468..12171, complement | XP_005936534.1 |
| PTH2 | 106633640 | NW_005179731.1 | 310219..312857 | XP_014195500.1 |
| PTHLH | 102305631 | NW_005179673.1 | 98846..102040 | XP_005939766.1; XP_014194690.1[a] |

[a]PTHLH has two predicted protein isoforms due to different initiation codons. Subsequent evaluation indicated that the isoform (XP_005939766.1) has the same sequence length as the mammalian protein homologs and thus will be entered in the prohormone table of candidate prohormones

regions across species. Effective BLASTP searches disable filtering and masking of low-complexity regions to favor the visualization of long sequence alignments. BLASTP searches using a less significant (higher) e-value threshold, as well as substitution matrices that support more distance sequence matches (e.g., BLOSSUM 50), should be investigated when the original search fails to identify matches.

All BLASTP matches that meet the e-value cutoff and have at least one conserved region should be recorded in the table of candidate prohormones, including locations and e-values. Each BLASTP match should be referenced to the candidate list, as all of the entries identified by the text search should also be identified from this sequence similarity search. These existing entries should be updated with the BLASTP match details.

Matches to the candidate prohormone protein sequence could correspond to predicted splice variants resulting from alternative splicing, duplicated genes, other members from the same prohormone family, and other homologs (*see* **Note 6**). Information from the text and any prior searches enables the initial differentiation of these redundant matches and elimination of matches already evaluated, or matches to another member from the same prohormone family. The best BLASTP match (lowest e-value) should be evaluated for completeness, even if the protein sequence was identified by the previous text search. This enables the identification of possible splice variants or correction of inaccurate protein sequences, such as those generated by automated methods in annotation of the genome assembly.

Each candidate prohormone protein sequence match must be aligned against the prohormone sequence in the reference species, and potentially other species, using tools such as Clustal Omega (*see* **Note 7**). True matches are characterized by alignments that have similar lengths and regions of high amino acid identity or similarity. The information gained from the alignment of the best prohormone sequence match to sequences in multiple species must be added to the table of candidates. These details can include the additional species that contain matches, accession numbers, matching sequences, and the most complete and longest alignments between the candidate sequence and the sequences in other species. All candidate prohormone sequences that have at least one complete protein match that surpasses the e-value cutoff in another species must be marked as completed in the candidate prohormone table. Protein sequence matches supersede text matches, and thus a text match must be replaced with a protein match when available.

The remaining BLASTP matches that surpass the e-value cutoff (other than the best match) include those that contain a high degree of homology (likely duplicated genes), homologous genes from the same gene family, and new prohormone genes. All these matches should be collected into a single table of protein matches and then each entry evaluated one at a time. Entries that match existing complete entries in the candidate prohormone table, or are clearly not prohormone proteins, are removed from the table of protein matches. Entries that match any candidate prohormone or new prohormones should be evaluated in a similar manner to the best match. The resulting complete entries are used to update the table of candidate prohormones, including information on sequence and accession number, before the entry is removed from the table of protein matches.

Some of the remaining entries may be resolved with alternative searches that can result in the identification of a new prohormone that may be unique to the target species. Conducting a protein search in the protein sequences from another species can provide matches that can determine if the current entry is a match to a prohormone, or that can be used instead of the candidate protein sequence. Similarly, searches using the EST database may provide the gene identity or provide a larger region that aids in the resolution of the initial match. Any remaining entries in the table of protein matches must be resolved with genome searches.

In our example, a BLASTP search was initially conducted with the *H. sapiens* PTH1, PTH2, and PTHLH sequences because this species is not encumbered by the teleost whole genome duplication event. Our searches (results not shown) did not uncover additional information to add to our initial text search (Table 1). Searches with *Danio rerio* (zebrafish) also did not uncover additional protein sequences. However, the *D. rerio* search results indicated that expected gene duplications resulting from the teleost whole genome duplication event were also not present in the *A. burtoni* protein sequences (results not shown).

Genome searches need to be conducted for any candidate prohormones that were not identified in the previous BLASTP searches and any remaining entries in the table of protein matches. A genome search should also be conducted if additional gene duplications are expected, e.g., from whole genome duplications or known tandem duplications. The TBLASTN version of BLAST that searches a protein sequence against translated nucleotide sequences of the genome should be used with the same settings as the previous BLASTP search. The expected matching regions should comprise the exons of the candidate gene as well as exons from genes homologous to the candidate protein sequence. All of these matches should be collected into a single table of genome matches and any duplicate matches should be removed. Any entries that correspond to exons of identified prohormones should be removed, after ensuring that the genomic region only contains the identified prohormone gene.

The remaining entries in the genome table that are likely to be from the same gene should be grouped together. These groups should contain the genomic regions of at least one of the prohormone gene exons, and then evaluated one at a time. Each group must be evaluated for assembly artifacts that must be resolved before continuing (*see* **Note 8**). Subsequently, the genome region of the group entry is expanded by ~ 2000 bp beyond the 5′ and 3′ ends of the group entry and the sequence from this extended region corresponding to the correct reading frame of the TBLASTN match is then extracted. This strategy maximizes the likelihood that the complete prohormone gene will be identified. A gene prediction program (e.g., GeneWise) is used to predict a protein sequence from this extracted genome region (*see* **Note 9**). If the

predicted protein sequence is a new prohormone, then the table of candidate prohormones is updated accordingly. All entries in the genome table corresponding to genome group are then removed.

We performed genome searches for PTH1, PTH2, and PTHLH in the *A. burtoni* genome using TBLASTN (results not shown). All matches from the protein sequences of the three genes were entered as a single list of matches since the same match can occur with different members of the same prohormone family. Any duplicate entries were removed so that each match was only processed once. The first entry processed was a single match to PTH2, corresponding to the previously identified PTH2 gene (Table 1). This result indicated that there is only one copy in the genome, so this entry was removed from the list of matches.

The next entries processed were 2 matches to PTH1 (Fig. 3). The first of these entries matched the region of the NW_005179605.1 contig, which contains the previously identified PTH1 gene (XP_005936534.1; Table 1). This entry did not provide additional information and was removed from the list of matches. The second PTH1 entry corresponded to another contig, NW_005179496.1, and exhibited high similarity to the start of the second exon of PTH1. This discovery likely corresponds to the duplicated PTH1 gene resulting from the teleost whole genome duplication event.

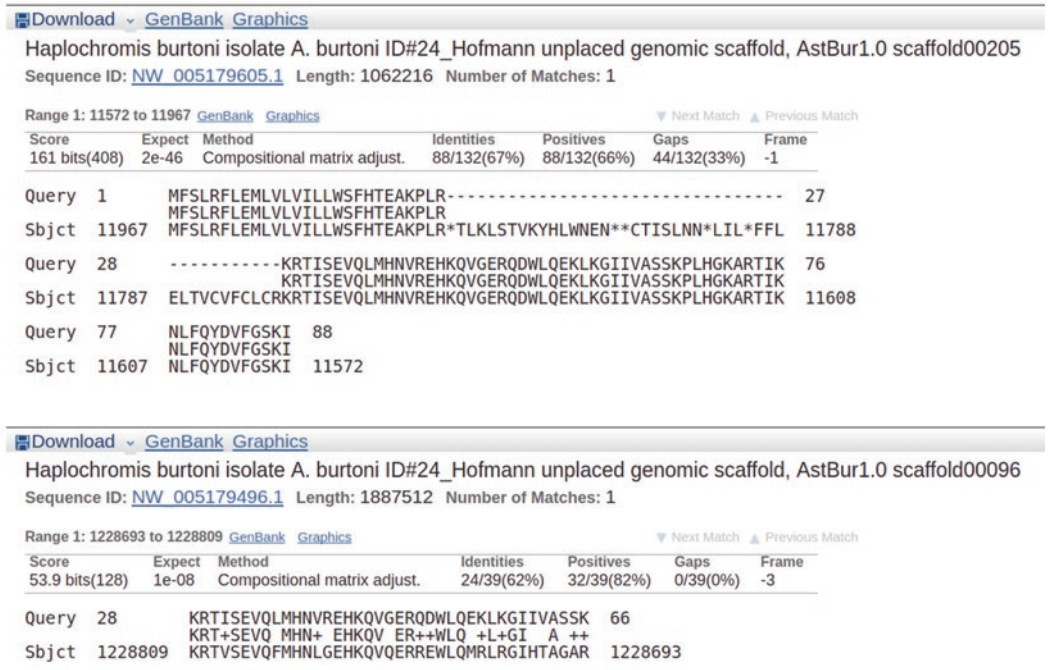



**Fig. 3** Output of the NCBI TBLASTN search of the parathyroid hormone 1 (PTH1) gene in the *Astatotilapia burtoni* genome


**Fig. 4** Output of the sequence of a TBLASTN match of PTH1 in the *Astatotilapia burtoni* genome

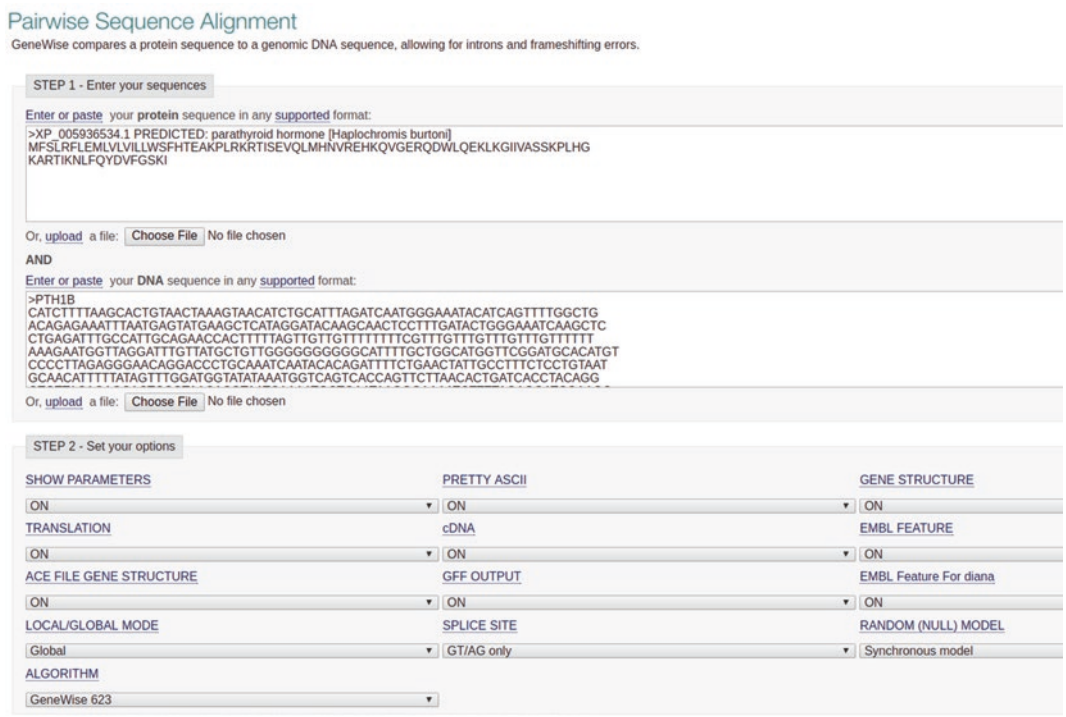



**Fig. 5** Input to the GeneWise program using the *Astatotilapia burtoni* parathyroid hormone 1 protein sequence and genome region containing the suspected duplicated parathyroid hormone 1 gene

To confirm that this match was a duplication of PTH1, we used the "GenBank" link in NCBI output to retrieve the genome record and the matched sequence. The TBLASTN search had located a match in the complement strand, and a customized view was developed by selecting the reverse complement option and "update view" in NCBI (Fig. 4). This region was expanded ~2000 bp from the 5′ and 3′ ends and extracted. The GeneWise program was then used to predict a protein sequence from this region. The input (Fig. 5) included the *A. burtoni* PTH1 protein sequence (denoted as PTH1A) and the extracted *A. burtoni*

genomic region. The output confirms our hypothesis of a duplicated PTH1 gene, providing the complete predicted gene sequence (Fig. 6), including an intron that is consistent with the known PTH1 gene. The previous PTH1 gene was renamed as PTH1A, and this new prediction denoted as PTH1B; the letters after the initial gene symbol (PTH1) are used to differentiate the different copies of the parathyroid hormone 1 gene. The protein sequence and information pertaining to PTH1B were added to candidate list and the entry was removed from the list of matches.

Our TBLASTN search of PTHLH in the target *A. burtoni* genome uncovered five matches with an e-value <10 (Fig. 7). The best match (lower e-value) was to the region of the NW_005179673.1 contig containing the previously identified PTHLH gene (XP_005939766.1; Table 1). This entry was removed from the list of matches because no additional information was obtained. The two matches with the highest e-values matched non-prohormone genes and these entries were discarded from the list of matches. The second-best match exhibited high similarity to the signal peptide region of PTHLH and the third-best match mapped to the start of the PTHLH sequence (Fig. 7). The locations of the second and third matches excluded the other known members of the parathyroid hormone family. This implied that these could be 3 *A. burtoni* PTHLH genes. Following the same approach previously used to obtain the PTH1B gene, augmented genome regions corresponding to these matches were extracted and the two different protein sequences were predicted. The initial PTHLH was renamed as PTHLH1 and the two predictions were denoted as PTHLH2 and PTHLH3 to differentiate the different PTHLH genes. After updating the candidate list and entering the proteins for these three PTHLH genes, these final entries were removed from the list of matches, marking the successful completion of the homology search step.

#### 3.2.3 Novel Detection Based on Neuropeptide Motifs

We recommend homology searches centered on conserved regions, such as neuropeptides, and using relatively lax criteria as helpful strategies to identifying novel prohormones. A more general approach is to search for neuropeptide motifs. This approach has successfully led to the discovery of new prohormones [35, 36]. However, searches for more generic motifs [37] or conserved regions inferred in silico using machine learning algorithms, such as hidden Markov models [38–40], have a high false-positive rate [41].

#### 3.2.4 Validation of Predicted Prohormone Protein Sequences

The predicted protein sequences must be validated for accuracy. This validation includes additional support from data not included in genome repository of the target species and information from other species. Homology searches for the prohormone protein sequences should be conducted in EST, RNA, and protein sequence databases that have not been incorporated

```
genewise $Name: wise2-4-1 $ (unreleased release)
This program is freely distributed under a GPL. See source directory
Copyright (c) GRL limited: portions of the code are from separate copyright

Query protein:       XP_005936534.1
Comp Matrix:         BLOSUM62.bla
Gap open:            12
Gap extension:       2
Start/End            global
Target Sequence      PTH1B
Strand:              forward
Start/End (protein)  global
Gene Parameter file: gene.stat
Splice site model:   GT/AG only
GT/AG bits penalty   -9.96
Codon Table:         codon.table
Subs error:          1e-06
Indel error:         1e-06
Null model           syn
Algorithm            623

genewise output
Score 28.31 bits over entire alignment
Scores as bits over a synchronous coding model

Warning: The bits scores is not probablistically correct for single seqs
See WWW help for more info

XP_005936534.1     1 MFSLRFLEMLV-LVILLWSFHTEAKPL
                     M    +  +L+ L +L +S H + +PL
                     MGKTDYKILLISLCLLHFSVHCQGRPL
PTH1B           1826 agaagtaaccatctctcttgctcgccc
                     tgacaaattttctgttatctagaggct
                     gaatctgtattacctaccctctaaaaa


XP_005936534.1    27                                KRTISEVQLMHNVREHKQV
                                                    KRT+SEVQ MHN+ EHKQV
                                 R:S[agt]           KRTVSEVQFMHNLGEHKQV
PTH1B           1907 AGGTAATCC  Intron 1       CAGTaaagaggctacacggcacg
                       <2-----[1909   :   1999]-2> agctgatattaatgaaaat
                                                    agagtgcgcgcccagcggg


XP_005936534.1    47 GERQDWLQEKLKGIIVASSKPLHGKARTIKNLFQYDVFGSKI
                      ER++WLQ +L+GI  A ++    +  T +   ++ +   ++
                     QERREWLQMRLRGIHTAGARN-SSRETTGRRRRRWPLRLEEM
PTH1B           2058 cgccgtccaaccgacagggca aacgaagcaaaatccccgga
                     aaggagtatgtggtaccgcga gggaccgggggggctgtaat
                     ggcgggggacgtccgaacgc ctggggcgggaggggtgggg


//
Gene 1
Gene 1826 2180
  Exon 1826 1908 phase 0
  Exon 2000 2180 phase 2
//
FT          CDS    join(1826..1908,2000..2180)
FT                   /note="Match to XP_005936534.1"
//
FT   misc_feature    join(1826..1908,2000..2180)
FT                   /note="Match to XP_005936534.1 Score 28.31"
//
>PTH1B.[1826:2180].sp.tr
MGKTDYKILLISLCLLHFSVHCQGRPLSKRTVSEVQFMHNLGEHKQVQERREWLQMRLRG
IHTAGARNSSRETTGRRRRRWPLRLEEM
//
```

**Fig. 6** Output from GeneWise depicting the alignment between the amino acid sequence from the region extracted and the *Astatotilapia burtoni* PTH1 prohormone sequence and the predicted protein sequence of duplicated parathyroid hormone gene (denoted as PTH1B)


Sequences producing significant alignments:

Select: All None Selected:0

Alignments Download GenBank Graphics

| Description | Max score | Total score | Query cover | E value | Ident | Accession |
|---|---|---|---|---|---|---|
| Haplochromis burtoni isolate A. burtoni ID#24_Hofmann unplaced genomic scaffold, AstBur1.0 scaffold00273 | 269 | 348 | 100% | 9e-83 | 84% | NW_005179673.1 |
| Haplochromis burtoni isolate A. burtoni ID#24_Hofmann unplaced genomic scaffold, AstBur1.0 scaffold00732 | 51.2 | 51.2 | 18% | 1e-06 | 77% | NW_005180132.1 |
| Haplochromis burtoni isolate A. burtoni ID#24_Hofmann unplaced genomic scaffold, AstBur1.0 scaffold00021 | 43.1 | 43.1 | 24% | 0.001 | 48% | NW_005179421.1 |
| Haplochromis burtoni isolate A. burtoni ID#24_Hofmann unplaced genomic scaffold, AstBur1.0 scaffold00408 | 32.7 | 32.7 | 71% | 2.7 | 29% | NW_005179808.1 |
| Haplochromis burtoni isolate A. burtoni ID#24_Hofmann unplaced genomic scaffold, AstBur1.0 scaffold00276 | 31.2 | 31.2 | 17% | 7.1 | 41% | NW_005179676.1 |

**Fig. 7** Output of the *Astatotilapia burtoni* parathyroid hormone-like prohormone TBLASTN search

into the genomic repository of the target species. These databases can also provide valuable complementary information, such as single nucleotide polymorphisms (SNPs), insertions, and deletions. Moreover, this complementary information can be helpful when the coverage of the target genome assembly varies across regions or is incomplete.

Multiple matches between the protein sequence from the target species and the EST sequences, including the full sequence or overlapping regions, offer the strongest empirical evidence for the gene. Gaps in the alignment could be attributed to alternative splicing, and thus all EST forms should be extracted and tested for sequence accuracy and completeness. When protein isoforms from the same prohormone are available, then all isoforms should be predicted from the same genomic region.

Confirmation using sequence information from nongenomic databases of other phylogenetically close species can be used to further validate the prohormone predictions. Matches to EST or other transcriptomic databases provide evidence of the presence of the predicted prohormone. However, the actual protein sequence cannot be completely verified since differences between sequences can be actual sequence differences between species, as well as a result of sequencing errors.

Our predicted *A. burtoni* PTH1B gene has no supporting confirmatory experimental evidence in any of the current *A. burtoni* resources. A suitable match in *Neolamprologus brichardi*, a species very closely related to *A. burtoni*, was found to the NCBI-predicted protein, "XP_006804843.1 PREDICTED: parathyroid hormone-like." This cross-species analysis offers further confirmation of our *A. burtoni* PTH1B prediction.

The new *A. burtoni* PTHLH2 and PTHLH3 prohormone protein sequences were validated using the NCBI *A. burtoni* transcriptomic-based databases. The PTHLH2 protein sequence had multiple matches in the NCBI *A. burtoni* "Non-RefSeq RNA" database (this database contains RNA sequences that are present in GenBank but have not been included in the RefSeq database). One of these additional matches was an *A. burtoni*

RNA sequence in the Transcriptome Shotgun Assembly (TSA) database [42] (titled "GBDH01011309.1 TSA: Haplochromis burtoni comp20762_c0_seq1 transcribed RNA sequence"). The PTHLH3 protein sequence matched the *A. burtoni* EST, DY630955.1. Extracting and translating these sequences from each source confirmed the previously predicted PTHLH2 and PTHLH3 protein sequences. Recently PTHLH3 was identified in zebrafish (*D. rerio*) as parathyroid hormone 4 (PTH4; XP_005168342.1) and was shown to interact with the known zebrafish PTH receptors [43].

### 3.3 Sequence Verification of Predicted Prohormone Proteins

The comprehensive validation of the predicted prohormone in the target species enhances confidence in the detection. Further biological support can be gained using motif- or block-centered multiple sequence alignments. The alignment of the predicted prohormone sequences to databases of protein families using tools such as Pfam [44] further ensures that the target sequence is a member of the expected protein family. This additional validation step is particularly helpful when the candidate sequence or the validation information corresponds to a phylogenetically distant related species.

We recommend an additional bioinformatics step for accurate annotation of prohormone paralogs and orthologs in the target species, particularly in protein families that have a high level of homology. This step encompasses multiple sequence alignments of the candidate sequences and known prohormone sequences from multiple species, including phylogenetically distant species, using tools such as Clustal Omega. In addition to discrimination between prohormone family members and identification of new duplicates in the target species, multiple sequence alignments support the assessment of expected structural features, such as conserved regions that are often characteristic for bioactive neuropeptides. For example, in our *A. burtoni* annotation [4], we confirmed the loss of melanocyte-stimulating hormone (MSH) peptide, γ-MSH [45].

To validate the accuracy of our parathyroid prohormone predictions, the 7 *A. burtoni* sequences were searched against the NCBI-predicted genes from 4 related cichlid species with sequenced genomes: *Oreochromis niloticus* (Nile tilapia), *N. brichardi, Pundamilia nyererei*, and *Maylandia zebra*. Matches to all of our parathyroid prohormone predictions were identified in multiple species (Table 2). Two PTH2 protein isoforms with corresponding mRNA evidence were detected in *O. niloticus*. Based on this evidence, both PTH2 protein isoforms were subsequently predicted from the same *A. burtoni* genomic region using GeneWise. Our bioinformatics approach uncovered 2 *A. burtoni* PTH2 isoforms that are the result from alternative splicing events.

**Table 2**
**Matches of the seven predicted parathyroid prohormone family protein sequences to the protein databases for *Astatotilapia burtoni*, *Oreochromis niloticus*, *Neolamprologus brichardi*, *Pundamilia nyererei*, and *Maylandia zebra***

| | NCBI protein accession number | | | | |
|---|---|---|---|---|---|
| **Gene symbol** | ***A. burtoni*** | ***O. niloticus*** | ***N. brichardi*** | ***P. nyererei*** | ***M. zebra*** |
| PTH1A | XP_005936534.1 | | XP_006780577.1 | XP_005724427.1 | XP_004543577.1 |
| PTH1B | | | XP_006804843.1 | | |
| PTH2_v1 | | NP_001266421.1 | | | XP_012779771.1 |
| PTH2_v2 | XP_014195500.1 | XP_013120746.1 | | XP_013771025.1 | |
| PTHLH1 | XP_005939766.1 | XP_003443941.1 | XP_006782617.1 | XP_005728183.1 | XP_004563786.1 |
| PTHLH2 | | XP_003448833.1 | XP_006781115.1 | | XP_004564921.1 |
| PTHLH3 | | | XP_006783019.1 | XP_005729696.1 | |

Phylogenetic trees depicting the relationship between the sequences from the same prohormone or prohormone family across species can also be constructed using tools such as ETE3 [46]. These trees facilitate the identification of potential sources of prohormone sequence alignment errors, such as unusual or unexpected prediction, gaps, large mismatched regions, or misnamed predictions. Also, these phylogenetic trees provide insights into the evolutionary relationship between the sequences and can be used to discover novel taxa-specific prohormone genes.

All of the known sequences from the *A. burtoni*, *D. rerio*, and *H. sapiens* parathyroid prohormone family were aligned together using Clustal Omega. A phylogenetic tree (Fig. 8) depicting the relationship between these parathyroid prohormone family sequences was constructed using the GenomeNet ETE 3 (v3.0.0b32) implementation [47]. This tree confirms that the sequences appear to be correctly named and illustrates the expected relationships between members of the parathyroid prohormone family. The branching in the PTHLH tree suggests 2 gene duplication events. The first is the teleost-specific whole genome duplication that resulted in PTHLH1 and PTHLH2; the second event occurred prior to the teleost-specific whole genome duplication and resulted in PTHLH3 (PTH4) and PTHLH. There is no *H. sapiens* PTHLH3 because PTHLH3 (PTH4) was lost in eutherian mammals after the eutherian-metatherian split [43]. These discoveries highlight the potential of our bioinformatic approach to annotate prohormones in target species, discover novel taxa-specific prohormone genes, and identify losses of prohormone genes.

### 3.4 Peptide Prediction from Prohormone Protein Sequences

Tools such as SignalP [48] can be used to predict the signal peptide and associated cleavage site in the N-terminal region of the predicted prohormone sequence in the target species (*see* **Note 10**). Rigorous application of the previous prohormone prediction steps guarantees the identification of a complete or nearly complete prohormone sequence. Incomplete protocol implementation or limited sequence information could produce partial, incorrect, or chimeric sequences. The predicted prohormone protein sequence must be revised when the signal peptide cleavage site cannot be identified. The sequence must also be revised when the location of the signal peptide cleavage site is predicted >40 amino acids from the prohormone initiation methionine because this location is generally atypical.

#### 3.4.1 Signal Peptide Prediction

#### 3.4.2 Prediction of Putative Peptides

The predicted prohormone sequence must be subsequently analyzed using NeuroPred. This public web service supports the prediction of putative peptide cleavage sites in a protein sequence, assessment of the likelihood of cleavage, and evaluation of potential PTMs. This resource accepts one or more protein sequences, and signal peptides are identified using

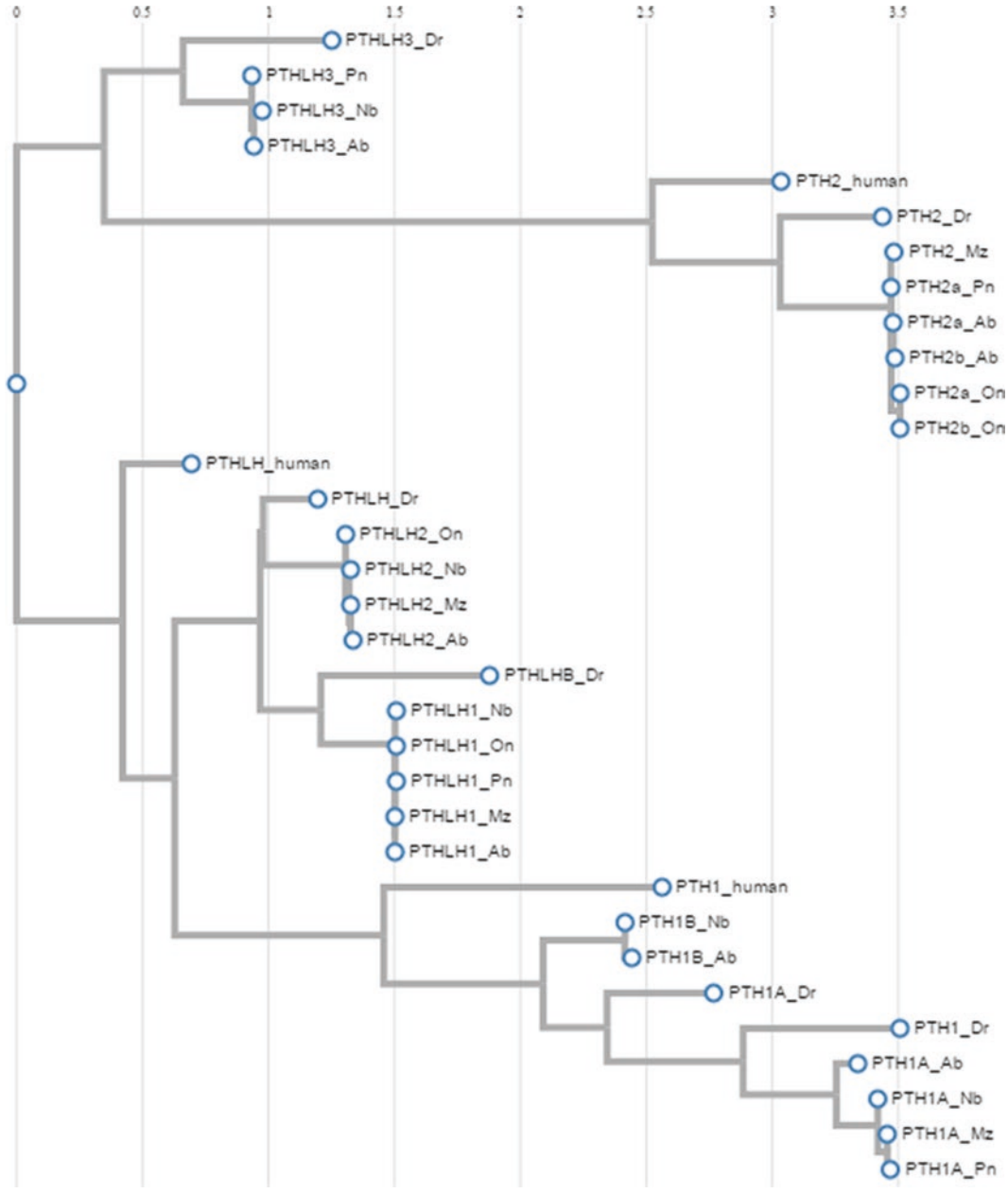



**Fig. 8** Phylogenetic tree of parathyroid prohormone genes obtained using the GenomeNet ETE3. Leaves represent the gene symbol followed by species suffix: *Astatotilapia burtoni* (Ab); Danio rerio (Dr); *Maylandia zebra* (Mz); *Oreochromis niloticus* (On); *Neolamprologus brichardi* (Nb); *Pundamilia nyererei* (Pn); and *Homo sapiens* (human)

default or user-defined specifications (*see* **Note 11**). NeuroPred encompasses multiple peptide cleavage prediction methods that have been optimized for multiple species (*see* **Note 12**). After removal of the signal peptide, NeuroPred predicts the cleavage sites and, depending on the selected options, identifies the

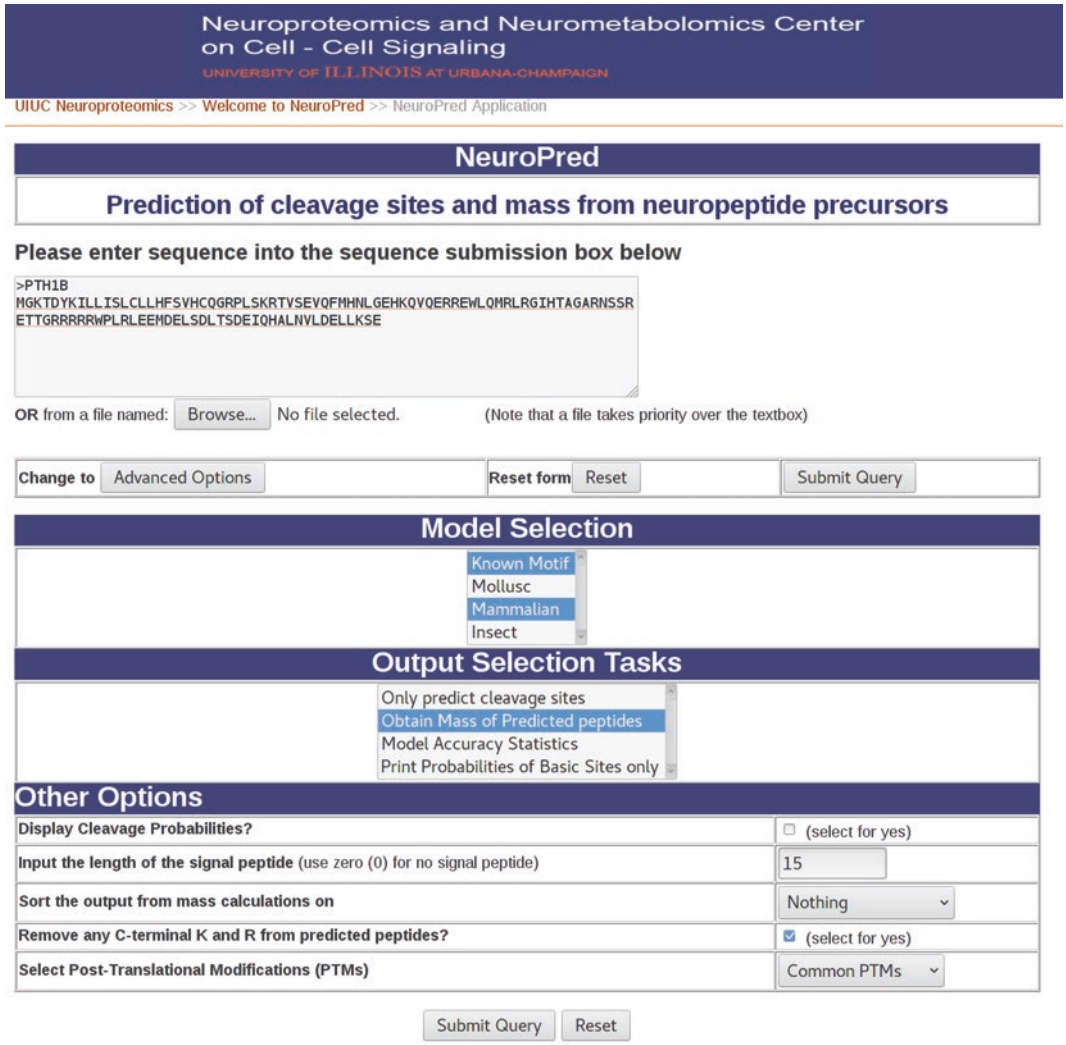
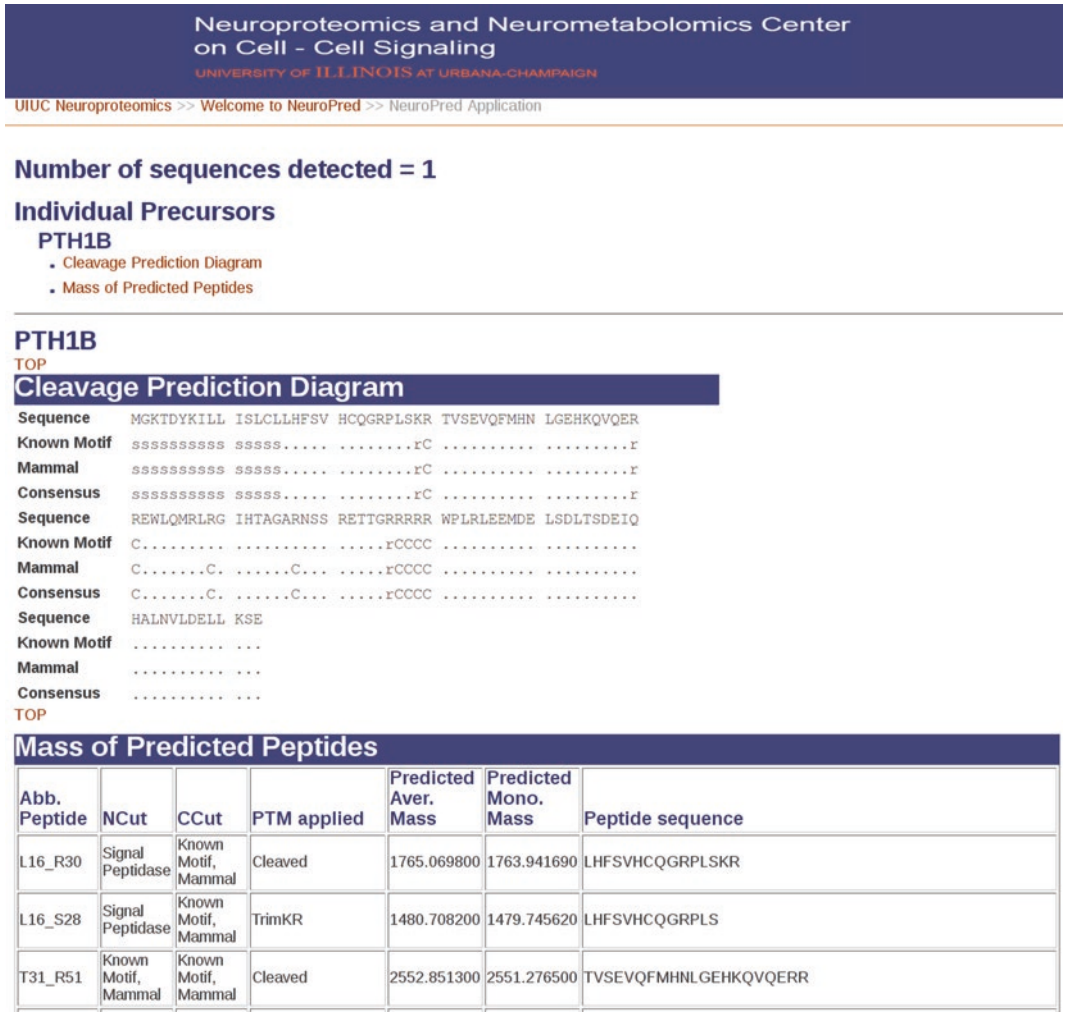



**Fig. 9** Input to the NeuroPred tool to predict cleavage sites for *Astatotilapia burtoni* parathyroid hormone 1B (PTH1B) prohormone protein sequence with the "Known Motif" and "Mammalian" cleavage models selected

resulting peptides. The peptide sequences, PTMs, and masses predicted by NeuroPred have been successfully used in support of neuropeptide identification in MS analyses.

We used NeuroPred to predict the peptides in the *A. burtoni* PTH1B protein sequence (Figs. 9 and 10). The NeuroPred input options depicted in Fig. 9 include the model used to predict the cleavage sites (the "Known Motif" and "Mammalian" cleavage models were selected for the *A. burtoni* example). The NeuroPred output depicted in Fig. 10 includes a cleavage diagram that localizes the signal peptide used and all predicted cleavage sites. The output also includes information on the peptides resulting from cleavage at the predicted sites.


Neuroproteomics and Neurometabolomics Center on Cell - Cell Signaling
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

UIUC Neuroproteomics >> Welcome to NeuroPred >> NeuroPred Application

**Number of sequences detected = 1**

**Individual Precursors**

**PTH1B**

- Cleavage Prediction Diagram
- Mass of Predicted Peptides

**PTH1B**

TOP

**Cleavage Prediction Diagram**

| | |
|---|---|
| Sequence | MGKTDYKILL ISLCLLHFSV HCQGRPLSKR TVSEVQFMHN LGEHKQVQER |
| Known Motif | sssssssss sssss..... ........rC .......... .........r |
| Mammal | sssssssss sssss..... ........rC .......... .........r |
| Consensus | sssssssss sssss..... ........rC .......... .........r |
| Sequence | REWLQMRLRG IHTAGARNSS RETTGRRRRR WPLRLEEMDE LSDLTSDEIQ |
| Known Motif | C......... .......... .....rCCCC .......... .......... |
| Mammal | C.......C. ......C... .....rCCCC .......... .......... |
| Consensus | C.......C. ......C... .....rCCCC .......... .......... |
| Sequence | HALNVLDELL KSE |
| Known Motif | .......... ... |
| Mammal | .......... ... |
| Consensus | .......... ... |

TOP

**Mass of Predicted Peptides**

| Abb. Peptide | NCut | CCut | PTM applied | Predicted Aver. Mass | Predicted Mono. Mass | Peptide sequence |
|---|---|---|---|---|---|---|
| L16_R30 | Signal Peptidase | Known Motif, Mammal | Cleaved | 1765.069800 | 1763.941690 | LHFSVHCQGRPLSKR |
| L16_S28 | Signal Peptidase | Known Motif, Mammal | TrimKR | 1480.708200 | 1479.745620 | LHFSVHCQGRPLS |
| T31_R51 | Known Motif, Mammal | Known Motif, Mammal | Cleaved | 2552.851300 | 2551.276500 | TVSEVQFMHNLGEHKQVQERR |

**Fig. 10** Output of the NeuroPred tool showing the cleavage diagram, where the predicted cleavage sites are denoted by the letter "C" for "Known Motif" and "Mammalian" cleavage models, and sequences and masses of first putative peptides

## 4 Notes

1. The accuracy of the prohormone gene prediction increases exponentially with increasing sequence similarity between species studied. Thus, the species phylogenetically closest to the target species should be used in the first bioinformatics steps. Sequences from more phylogenetically distant species should be used to resolve contradictory findings among closest species and whenever substantial uncertainty arises from poor matches or unreliable predictions during the final bioinformatic steps.

2. The ability to identify the start and end locations of exons from multiple genome matches is critical to the discrimination between duplicated genes located on the same chromosome or contigs in the target species. Furthermore, accurate exon delimitation enables the reduction of multiple matches to the same region by different genes, typically from the same gene family.
3. The nomenclature used to name a prohormone on a target species should follow community annotation guidelines. When these guidelines are not in place, the recommendation is to follow the guidelines of a major genome annotation project within the taxa. Accepted gene symbols should be used for prohormones to facilitate search and validation, especially across species, and to avoid confusion between proteins and peptides and between duplicated genes.
4. Characterization of prohormones resulting from tandem duplication requires recursive iteration of the bioinformatic steps described. An example of this scenario is mammalian calcitonin, which has varying copy numbers among mammalian species. Knowledge of potential tandem duplication aids in the accurate identification of the individual prohormone genes.
5. Known genes may not be found using text searches because the names of genes, proteins, and neuropeptides can vary between species and across time. In the absence of identifications using text searches, subsequent searches should explore gene symbols and synonyms. In some cases, text searches in other species may uncover alternative annotation, nomenclature, and search terms. The candidate prohormone table should include all synonyms of gene names and symbols.
6. Duplicated genes or homologs may match to the same region as the candidate protein sequence in addition to other regions in the genome of the target species. Expected regions based on comparative mapping information should be evaluated first to identify the primary ortholog gene, and nearby matches are likely to be the result of tandem duplication. Also, incomplete coverage or assembly may cause one gene sequence to be mapped to substantially different regions. In these cases, the matches tend to be short and lack conserved regions. Other types of sequences such as ESTs can be very helpful to address gaps or low quality sequences.
7. Multiple sequence alignment is the favored method for identifying inaccuracies in prohormone protein prediction. An incorrect initiation codon will result in predicted protein isoforms that are contained within a subset of a larger protein isoform. Incorrect termination codons will result in the last exon being incorrectly predicted or completely missed. Gaps and mismatches indicate incomplete coverage or species dif-

ferences that can only be resolved by other sources. Multiple sequence alignment should be used to resolve inaccuracies in the prediction, starting from the conserved regions and extending to both ends.

8. There are various assembly artifacts that have varying impacts on prohormone gene identification. Often many of these artifacts, such as matches in different strands or where exons of the gene are not sequentially located on the same contig, can be resolved by manually placing the translated sequence from exons in the correct order based on the candidate prohormone protein sequence. A different resolution is required for erroneous insertions that result in part of all of an exon located multiple times in nearby regions. Typically only one of these regions will provide a complete gene prediction. If remaining incorrect regions does not provide a complete prohormone gene prediction, then the matches to those regions can be safely discarded.
9. Alternative specifications of the gene prediction tool should be investigated when the extracted genome sequence does not support the prediction of a complete or nearly complete protein, or when the predicted protein is substantially shorter or longer than expected. Effective strategies include varying the sequence region being analyzed and removing excessive gaps or strings of ambiguous nucleotides. The global model option in GeneWise should be used when a high degree of homology is expected between the protein sequence and the genome of the target species. Using the gene prediction tool to predict the protein from the genome sequence of other species can offer insights into the expected prediction. Searching for the prohormone in other nucleotide databases of the target species could improve the prohormone prediction. Protein sequences should not be predicted using TBLASTN, because this tool may provide a low prediction accuracy at the intron-exon splice boundaries and may fail to identify all prohormone exons.
10. Signal peptide cleavage sites are typically located between 15 and 40 amino acids from the N-terminal start of the protein sequence. Signal peptides are highly conserved, and thus the predicted signal peptide cleavage is identical or very similar across related species, and often across members within a prohormone family. Multiple sequence alignment of prohormone sequences from multiple species that have known or predicted signal peptide cleavage sites will enhance the accuracy of the prediction when the predicted prohormone sequence in the target species is partial or encompasses highly uncertain positions. If a site is not predicted, alternative

specifications of the signal peptide prediction tool or different tools should be investigated.

11. The default specifications to predict peptides cleaved from protein sequences in NeuroPred were designed based on information for a large number of prohormones across multiple species and accommodate the sequences of many prohormones and neuropeptides. Users can overwrite these default specifications, e.g., by specifying the position of the signal peptide cleavage site when the input sequence is incomplete. The NeuroPred specifications for false-positive and false-negative cleavage predictions can also be adjusted according to the goals of the study.
12. Multiple species-specific models to predict cleavage sites and resulting peptides are available in NeuroPred. Cleavage and resulting peptide prediction models from the species phylogenetically closest to the target species should be used first. Models from more phylogenetically distant species should be examined for confirmation. NeuroPred supports the estimation of cleavage probability for each predicted cleavage site, which further empowers users to prioritize among the predicted peptides.

## Acknowledgments

This work was supported by the National Institutes of Health, Award No. P30 DA018310 from the National Institute on Drug Abuse (NIDA), the US Department of Agriculture (USDA) National Institute of Food and Agriculture (NIFA) project No. ILLU-538-909, and the National Science Foundation, Award No. CHE-16-06791. The content is solely the responsibility of the authors and does not necessarily represent the official views of the funding agencies.

### References

1. Burger E (1988) Peptide hormones and neuropeptides. Proteolytic processing of the precursor regulatory peptides. Arzneimittelforschung 38(5):754–761
2. von Heijne G (1990) The signal peptide. J Membr Biol 115(3):195–201. https://doi.org/10.1007/bf01868635
3. Amare A, Hummon AB, Southey BR, Zimmerman TA, Rodriguez-Zas SL, Sweedler JV (2006) Bridging neuropeptidomics and genomics with bioinformatics: prediction of mammalian neuropeptide prohormone processing. J Proteome Res 5(5):1162–1167. https://doi.org/10.1021/pr0504541
4. CK H, Southey BR, Romanova EV, Maruska KP, Sweedler JV, Fernald RD (2016) Identification of prohormones and pituitary neuropeptides in the African cichlid, Astatotilapia Burtoni. BMC Genomics 17(1):660. https://doi.org/10.1186/s12864-016-2914-9
5. Porter KI, Southey BR, Sweedler JV, Rodriguez-Zas SL (2012) First survey and functional annotation of prohormone and convertase genes in the pig. BMC Genomics 13:582. https://doi.org/10.1186/1471-2164-13-582
6. Southey BR, Rodriguez-Zas SL, Sweedler JV (2009) Characterization of the prohormone complement in cattle using genomic libraries and cleavage prediction approaches. BMC Genomics 10:228. https://doi.org/10.1186/1471-2164-10-228
7. Southey BR, Sweedler JV, Rodriguez-Zas SL (2008) A python analytical pipeline to identify prohormone precursors and predict prohormone

cleavage sites. Front Neuroinform 2:7. https://doi.org/10.3389/neuro.11.007.2008

8. Southey BR, Sweedler JV, Rodriguez-Zas SL (2008) Prediction of neuropeptide cleavage sites in insects. Bioinformatics 24(6):815–825. https://doi.org/10.1093/bioinformatics/btn044
9. Tegge AN, Southey BR, Sweedler JV, Rodriguez-Zas SL (2008) Comparative analysis of neuropeptide cleavage sites in human, mouse, rat, and cattle. Mamm Genome 19(2):106–120. https://doi.org/10.1007/s00335-007-9090-9
10. Murphy D, Alim FZD, Hindmarch C, Greenwood M, Rogers M, Gan CK, Yealing T, Romanova EV, Southey BR, Sweedler JV (2016) Seasonal adaptations of the hypothalamo-neurohypophyseal system of the Arabian one-humped camel. Paper presented at the Plant and Animal Genome, San Diego, CA, USA. https://pag.confex.com/pag/xxiv/webprogram/Paper18655.html
11. Southey BR, Amare A, Zimmerman TA, Rodriguez-Zas SL, Sweedler JV (2006) NeuroPred: a tool to predict cleavage sites in neuropeptide precursors and provide the masses of the resulting peptides. Nucleic Acids Res 34 (Web Server issue):W267–272. doi:https://doi.org/10.1093/nar/gkl161
12. Southey BR, Amare A, Zimmerman TA, Rodriguez-Zas SL, Sweedler JV (2017) NeuroPred application. http://neuroproteomics.scs.illinois.edu/neuropred.htm. Accessed 21 Feb 2017
13. Southey BR, Rodriguez-Zas SL, Sweedler JV (2006) Prediction of neuropeptide prohormone cleavages with application to RFamides. Peptides 27(5):1087–1098. https://doi.org/10.1016/j.peptides.2005.07.026
14. Southey BR, Rodriguez Zas SL (2017) PepShop application. http://stagbeetle.animal.uiuc.edu/pepshop. Accessed 21 Feb 2017
15. Grimmelikhuijzen CJ, Hauser F (2012) Mini-review: the evolution of neuropeptide signaling. Regul Pept 177(Suppl):S6–S9. https://doi.org/10.1016/j.regpep.2012.05.001
16. Romanova EV, Sweedler JV (2015) Peptidomics for the discovery and characterization of neuropeptides and hormones. Trends Pharmacol Sci 36(9):579–586. https://doi.org/10.1016/j.tips.2015.05.009
17. Glasauer SM, Neuhauss SC (2014) Whole-genome duplication in teleost fishes and its evolutionary consequences. Mol Gen Genomics 289(6):1045–1060. https://doi.org/10.1007/s00438-014-0889-2
18. Coordinators NR (2017) Database resources of the National Center for Biotechnology Information. Nucleic Acids Res 45(D1):D12–D17. https://doi.org/10.1093/nar/gkw1071
19. Yates A, Akanni W, Amode MR, Barrell D, Billis K, Carvalho-Silva D, Cummins C, Clapham P, Fitzgerald S, Gil L, Giron CG, Gordon L, Hourlier T, Hunt SE, Janacek SH, Johnson N, Juettemann T, Keenan S, Lavidas I, Martin FJ, Maurel T, McLaren W, Murphy DN, Nag R, Nuhn M, Parker A, Patricio M, Pignatelli M, Rahtz M, Riat HS, Sheppard D, Taylor K, Thormann A, Vullo A, Wilder SP, Zadissa A, Birney E, Harrow J, Muffato M, Perry E, Ruffier M, Spudich G, Trevanion SJ, Cunningham F, Aken BL, Zerbino DR, Flicek P (2016) Ensembl 2016. Nucleic Acids Res 44(D1):D710–D716. https://doi.org/10.1093/nar/gkv1157
20. UniProt C (2015) UniProt: a hub for protein information. Nucleic Acids Res 43(Database issue):D204–D212. https://doi.org/10.1093/nar/gku989
21. Pruitt KD, Brown GR, Hiatt SM, Thibaud-Nissen F, Astashyn A, Ermolaeva O, Farrell CM, Hart J, Landrum MJ, McGarvey KM, Murphy MR, O'Leary NA, Pujar S, Rajput B, Rangwala SH, Riddick LD, Shkeda A, Sun H, Tamez P, Tully RE, Wallin C, Webb D, Weber J, Wu W, DiCuccio M, Kitts P, Maglott DR, Murphy TD, Ostell JM (2014) RefSeq: an update on mammalian reference sequences. Nucleic Acids Res 42(Database issue):D756–D763. https://doi.org/10.1093/nar/gkt1114
22. Southey BR, Amare A, Zimmerman TA, Rodriguez-Zas SL, Sweedler JV (2017) NeuroPred sequence data. http://stagbeetle.animal.uiuc.edu/neuropred/sequences/Sequencedata.html. Accessed 21 Feb 2017
23. Liu F, Baggerman G, Schoofs L, Wets G (2008) The construction of a bioactive peptide database in Metazoa. J Proteome Res 7(9):4119–4131. https://doi.org/10.1021/pr800037n
24. Burbach JP (2010) Neuropeptides from concept to online database www.neuropeptides.nl. Eur J Pharmacol 626(1):27–48. https://doi.org/10.1016/j.ejphar.2009.10.015
25. Falth M, Skold K, Norrman M, Svensson M, Fenyo D, Andren PE (2006) SwePep, a database designed for endogenous peptides and mass spectrometry. Mol Cell Proteomics 5(6):998–1005. https://doi.org/10.1074/mcp.M500401-MCP200
26. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25(17):3389–3402. https://doi.org/10.1093/nar/25.17.3389

27. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Soding J, Thompson JD, Higgins DG (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal omega. Mol Syst Biol 7:539. https://doi.org/10.1038/msb.2011.75
28. Birney E, Clamp M, Durbin R (2004) GeneWise and Genomewise. Genome Res 14(5):988–995. https://doi.org/10.1101/gr.1865504
29. Wilkinson TN, Speed TP, Tregear GW, Bathgate RA (2005) Evolution of the relaxin-like peptide family. BMC Evol Biol 5:14. https://doi.org/10.1186/1471-2148-5-14
30. Wysolmerski JJ (2012) Parathyroid hormone-related protein: an update. J Clin Endocrinol Metab 97(9):2947–2956. https://doi.org/10.1210/jc.2012-2142
31. Bhattacharya P, Yan YL, Postlethwait J, Rubin DA (2011) Evolution of the vertebrate pth2 (tip39) gene family and the regulation of PTH type 2 receptor (pth2r) and its endogenous ligand pth2 by hedgehog signaling in zebrafish development. J Endocrinol 211(2):187–200. https://doi.org/10.1530/JOE-10-0439
32. Guerreiro PM, Renfro JL, Power DM, Canario AV (2007) The parathyroid hormone family of peptides: structure, tissue distribution, regulation, and potential functional roles in calcium and phosphate balance in fish. Am J Physiol Regul Integr Comp Physiol 292(2):R679–R696. https://doi.org/10.1152/ajpregu.00480.2006
33. NCBI (2017) Gene database. https://www.ncbi.nlm.nih.gov/gene/. Accessed 21 Feb 2017
34. NCBI (2017) Protein database. https://www.ncbi.nlm.nih.gov/protein/. Accessed 21 Feb 2017
35. Nathoo AN, Moeller RA, Westlund BA, Hart AC (2001) Identification of neuropeptide-like protein gene families in *Caenorhabditis elegans* and other species. Proc Natl Acad Sci U S A 98(24):14000–14005. https://doi.org/10.1073/pnas.241231298
36. Hummon AB, Richmond TA, Verleyen P, Baggerman G, Huybrechts J, Ewing MA, Vierstraete E, Rodriguez-Zas SL, Schoofs L, Robinson GE, Sweedler JV (2006) From the genome to the proteome: uncovering peptides in the Apis brain. Science 314(5799):647–649. https://doi.org/10.1126/science.1124128
37. Gustincich S, Batalov S, Beisel KW, Bono H, Carninci P, Fletcher CF, Grimmond S, Hirokawa N, Jarvis ED, Jegla T, Kawasawa Y, Lemieux J, Miki H, Raviola E, Teasdale RD, Tominaga N, Yagi K, Zimmer A, Hayashizaki Y, Okazaki Y, RIKEN GER Group; GSL Members (2003) Analysis of the mouse transcriptome for genes involved in the function of the nervous system. Genome Res 13(6B):1395–1401. https://doi.org/10.1101/gr.1135303
38. Shi L, Ko ML, Abbott LC, Ko GY (2012) Identification of Peptide Lv, a novel putative neuropeptide that regulates the expression of L-type voltage-gated calcium channels in photoreceptors. PLoS One 7(8):e43091. https://doi.org/10.1371/journal.pone.0043091
39. Mirabeau O, Perlas E, Severini C, Audero E, Gascuel O, Possenti R, Birney E, Rosenthal N, Gross C (2007) Identification of novel peptide hormones in the human proteome by hidden Markov model screening. Genome Res 17(3):320–327. https://doi.org/10.1101/gr.5755407
40. Sonmez K, Zaveri NT, Kerman IA, Burke S, Neal CR, Xie X, Watson SJ, Toll L (2009) Evolutionary sequence modeling for discovery of peptide hormones. PLoS Comput Biol 5(1):e1000258. https://doi.org/10.1371/journal.pcbi.1000258
41. Ozawa A, Lindberg I, Roth B, Kroeze WK (2010) Deorphanization of novel peptides and their receptors. AAPS J 12(3):378–384. https://doi.org/10.1208/s12248-010-9198-9
42. NCBI (2017) Transcriptome Shotgun Assembly database. https://www.ncbi.nlm.nih.gov/genbank/tsa/. Accessed 21 Feb 2017
43. Suarez-Bregua P, Torres-Nunez E, Saxena A, Guerreiro P, Braasch I, Prober DA, Moran P, Cerda-Reverter JM, SJ D, Adrio F, Power DM, Canario AV, Postlethwait JH, Bronner ME, Canestro C, Rotllant J (2017) Pth4, an ancient parathyroid hormone lost in eutherian mammals, reveals a new brain-to-bone signaling pathway. FASEB J 31(2):569–583. https://doi.org/10.1096/fj.201600815R
44. Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, Potter SC, Punta M, Qureshi M, Sangrador-Vegas A, Salazar GA, Tate J, Bateman A (2016) The Pfam protein families database: towards a more sustainable future. Nucleic Acids Res 44(D1):D279–D285. https://doi.org/10.1093/nar/gkv1344
45. Dores RM, Baron AJ (2011) Evolution of POMC: origin, phylogeny, posttranslational processing, and the melanocortins. Ann N Y Acad Sci 1220:34–48. https://doi.org/10.1111/j.1749-6632.2010.05928.x
46. Huerta-Cepas J, Serra F, Bork P (2016) ETE 3: reconstruction, analysis, and visualization of phylogenomic data. Mol Biol Evol 33(6):1635–1638. https://doi.org/10.1093/molbev/msw046
47. ETE G (2017) GenomeNet ETE3 application. http://www.genome.jp/tools/ete/. Accessed 21 Feb 2017
48. Petersen TN, Brunak S, von Heijne G, Nielsen H (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Methods 8(10):785–786. https://doi.org/10.1038/nmeth.1701

# Chapter 6

# Substrate Capture Assay Using Inactive Oligopeptidases to Identify Novel Peptides

**Vanessa Rioli and Emer S. Ferro**

## Abstract

Researchers are always searching for novel biologically active molecules including peptides. With the improvement of equipment for electrospray mass spectrometry, it is now possible to identify hundreds of novel peptides in a single run. However, after identifying the peptide sequences it is expensive to synthesize all the peptides to perform biological activity assays. Here, we describe a substrate capture assay that uses inactive oligopeptidases to identify putative biologically active peptides in complexes peptide mixtures. This methodology can use any crude extracts of biological tissues or cells, with the advantage to introduce a filter (i.e., binding to an inactive oligopeptidase) as a prior step in screening to bioactive peptides.

**Key words** Catalytically inactive oligopeptidase, Bioactive peptide, Gel filtration "semi-dry", Electrospray mass spectrometry

## 1 Introduction

Peptides are produced by cells from proteins specifically synthesized for this purpose, or as protein metabolism products [1, 2]. Many of the products formed in the first cases are known modulators of cellular communication known as neuropeptides [3]. In the second case, in compartments other than those specializing in protein degradation, specific cellular mechanisms induce proteins to limited digestion, generating intermediate peptides, many of which are completely unknown [1, 2]. The enzymes involved in the synthesis and metabolism of neuropeptides, as well as where this process takes place within cells is well known [3]. On the other hand, not much is known about the products of limited protein degradation that are formed in cells.

Several metalloendopeptidases have been isolated and characterized as putative peptide-metabolizing enzymes [4]. However, for most of these enzymes no natural substrates have been characterized. Here, we describe a feasible method that allows the inactivation of metalloproteases by site-direct mutagenesis and their further use for the isolation of substrates from crude tissue extracts [5].

Michael Schrader and Lloyd Fricker (eds.), *Peptidomics: Methods and Strategies*, Methods in Molecular Biology, vol. 1719, https://doi.org/10.1007/978-1-4939-7537-2_6, 

## 2 Materials

Prepare all solutions using ultrapure water (prepared by purifying deionized water, to attain a sensitivity of 18 MΩ•cm at 25 °C) and analytical grade reagents (*see* **Note 1**).

Use only low-binding tubes and low-binding filtration membranes.

### 2.1 Inactive Protease

Homogeneously purified recombinant inactive form of oligopeptidase (1–5 nmol) or site-directed inactive protease of choice. Expression and purification of the inactive form "E474A" of thimet oligopeptidase (EC 3.4.24.15; EP24.15) was conducted as previously described [4, 5] (*see* **Note 2**).

### 2.2 Preparing Crude Peptide Extracts

The procedures described in this section have been previously published [6–8].

1. Microwave oven (*see* **Note 3**).
2. Homogenizers, such as a mechanical tissue disruptor (e.g., Polytron) and sonicator with microtip probe.
3. 80 °C water bath.
4. Refrigerated centrifuge.
5. Refrigerated ultracentrifuge.
6. Sonicator.
7. 5 M HCl.
8. 1 M Tris–HCl (pH 7.4).
9. HCl 50 mM at 70 °C.
10. Phosphate-buffered saline (PBS).
11. Universal polymeric reversed-phase sorbent (we use Oasis HLB 1 cc Vac Cartridge, 30 mg Sorbent per Cartridge, 30 μm Particle Size from Waters).
12. Centrifugal Filter Unit "low binding" of 10 kDa MW cutoff (we use units from Millipore).
13. Gloves, latex, which should be worn at all times to prevent contamination of the samples.

### 2.3 Enzyme–Peptide Binding Assay

1. Inactive protease (1–5 nmol).
2. Crude peptide extracts (50–100 μg) from biological material of interest (*see* Subheading 3.1.1).
3. TBS buffer (25 mM Tris–HCl, pH 7.5, containing 125 mM NaCl) containing 0.1% of bovine serum albumin (TBS/BSA).

### 2.4 "Semi-Dry" Gel Filtration

1. Size exclusion gel (we use Sephadex G-25 fine from Sigma) previously washed in TBS buffer, and maintaining humidity ("semi-dry").
2. Centrifuge with swinging rotor (to be used at 1000 × *g* at room temperature).

### 2.5 High Performance Liquid Chromatography (HPLC)

1. Reverse phase C18 column.
2. Acetonitrile, grade HPLC (≥99, 9% purity).
3. Trifluoroacetic acid (TFA).

### 2.6 Mass Spectrometry (ESI-MS/MS) Coupled to a Capillary Liquid Chromatography (LC) System

1. Electrospray mass spectrometer coupled to a capillary LC system (e.g., Synapt G2 coupled to a nanoAcquity LC system, or equivalent).
2. Reverse phase trapping column with 5 μm particles, 180 μm inner diameter, 20 mm length (we use Symmetry C18, Waters).
3. Reverse phase C18 column of 75 μm inner diameter, 100 mm length and around 1.7 μm particles; we use BEH 130, Waters).
4. Eluents: 7–65% of phase B (0.1% formic acid in acetonitrile).

## 3 Methods

### 3.1 Crude Peptide Extracts

Peptides sourced from animal tissues or cell culture, as previously described [5–8].

#### 3.1.1 Tissues

1. Dissect the organ and immediately heat it for 10 s in the microwave oven; the temperature in the interior of the organ must reach 80 °C during this time without cooking the tissue (*see* **Note 3**).
2. Homogenize the tissue in 10 volumes (vol:vol) of hot water (80 °C) using a mechanical tissue disrupter device (e.g., Polytron). It is also desirable to further homogenize your samples using a sonicator, usually 20 pulses of 1 s each pulse, at a power output of 10 W.
3. Cool the homogenate in ice bath to 4 °C (*see* **Note 4**).
4. Add cold HCl solution to a final concentration of 10 mM and briefly homogenize using both the mechanical homogenizer (e.g., Polytron) and sonicator (*see* **Note 5**).
5. Centrifuge at 35,000 × *g* for 10 min at 4 °C. Collect the supernatant and discard the pellet (*see* **Note 6**).
6. Centrifuge at 100,000 × *g* for 1 h at 4 °C. Collect the supernatant and discard the pellet (*see* **Note 7**).

#### 3.1.2 Cell Culture

1. At least $1 \times 10^8$ cells (*see* **Note 8**) are needed.
2. Wash cells three times with warm phosphate-buffered saline (PBS).
3. Gently remove cells using cell scraper and rinse using five rounds of centrifugation at 800 × *g* for 5 min/resuspension (*see* **Note 9**).

4. Resuspend the final rinsed cell pellet in 10 mL (i.e., 10 volumes of the corresponding cell pellet; 10 vol:vol) of hot (80 °C) water and heat it for 20 min in a water bath at 80 °C, in order to inactivate proteases.
5. It is also desirable to further homogenize your samples using a sonicator, usually 20 pulses of 1 s each pulse, at a power output of 10 W.
6. Cool the homogenate in ice bath to 4 °C (*see* **Note 4**).
7. Add cold HCl solution to a final concentration of 10 mM and briefly homogenize using the polytron and sonicator (*see* **Note 5**).
8. Centrifuge at 35,000 × *g* for 10 min at 4 °C. Collect the supernatant and discard the pellet (*see* **Note 6**).
9. Centrifuge at 100,000 × *g* for 1 h at 4 °C. Collect the supernatant and discard the pellet (*see* **Note 7**).

#### *3.1.3 For Tissues and Cell Culture*

The supernatants collected should be filtered into size-exclusion filter device of 10 kDa at 1500 × *g* at 4 °C (*see* **Note 10**). The eluate containing the peptide crude fraction (MW < 10 kDa) must be reserved, and further subjected to concentration and chromatography onto a universal polymeric reversed-phase sorbent column according to manufacturer instructions (*see* **Note 11**). The volume of the crude peptide fraction eluted from these columns can be further concentrated to approximately 20–50 μL in speed vacuum centrifuge at 30 °C (*see* **Note 11**). Peptide concentrations in the crude extracts can be determined using fluorescamine, using a peptide of known concentration as standard [6, 9–11] (*see* **Note 12**).

### *3.2 Enzyme–Peptide Binding Assay (Fig. 1)*

1. Incubate for 30 min, at 21 °C (room temperature), the recombinant catalytically inactive oligopeptidase (1–5 nmol) with 50 μg of crude peptide diluted in TBS/BSA buffer to a final volume of 200 μL. For the control reaction, use the crude peptide mixture without the inactive oligopeptidase (*see* **Note 13**).
2. Add the above reaction mixture containing the inactive enzyme and the crude peptide mixture to a 1 mL bed of semi-dry Sephadex G-25 column (the plastic column device has 5 mm of diameter). Proceed with the control reaction similarly. Collect the flow-through (~200 μL) from experimental and control reactions, and analyze the peptide content by high performance liquid chromatography (HPLC) as described below. The flow through can be collected by gravity into a 1.5 mL low-binding plastic tube (*see* **Note 14**).
3. Analyze the peptide content by high performance liquid chromatography (HPLC) using a standard reverse phase C18

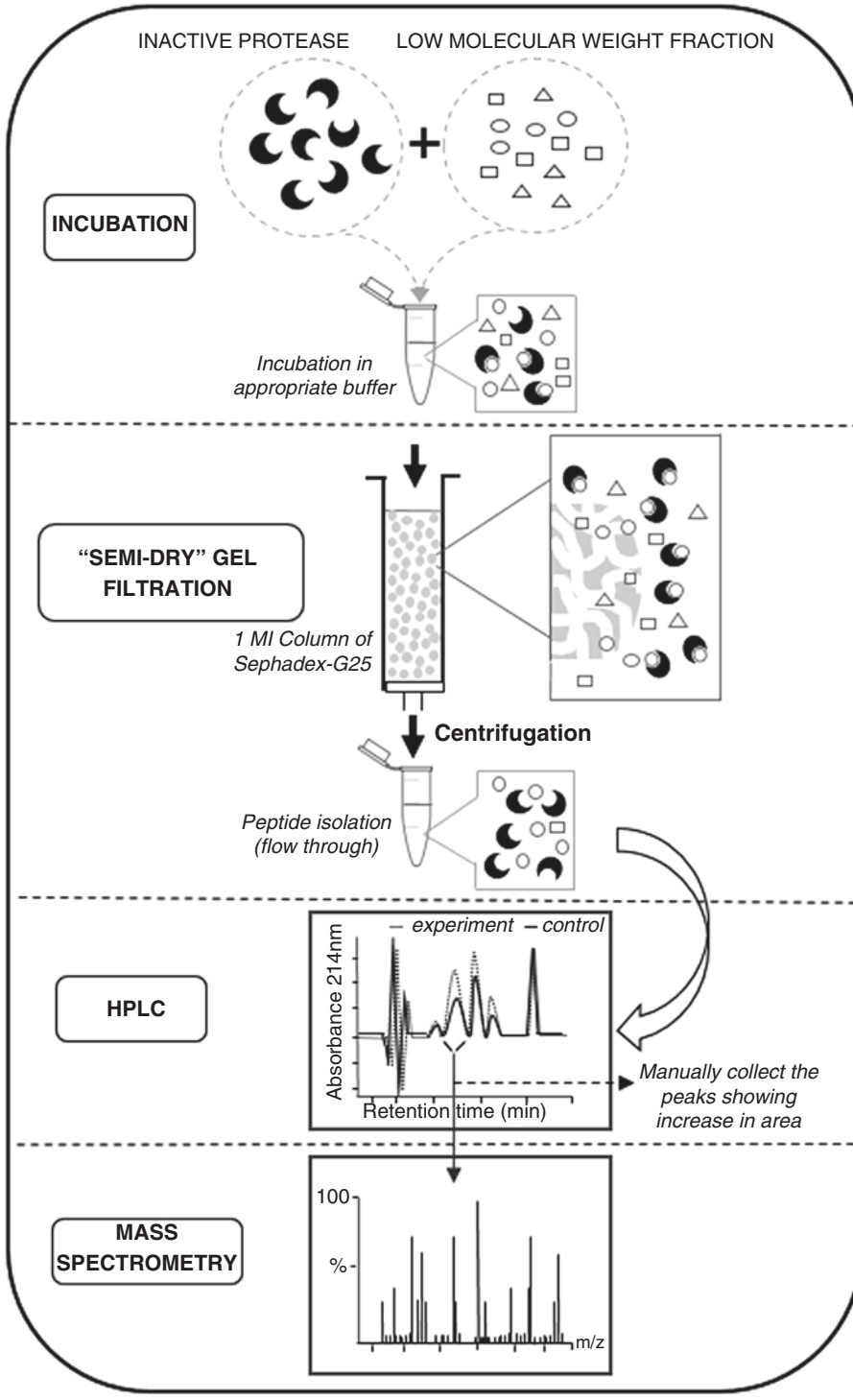



**Fig. 1** Schematic representation of the steps needed for the isolation of novel bioactive peptides from complex biological peptide samples using mutated an inactive oligopeptidase

column, in a linear gradient of 5–35% acetonitrile in 0.1% trifluoroacetic acid, for 20 min at a flow rate of 1 mL/min (*see* **Note 15**). First, run the control reactions and later the experimental samples containing the inactive enzyme. Manually collect specific chromatographic peaks that have shown increase in signal intensity when compared to the same peak from the control reaction (according to the retention time from the HPLC chromatogram, Fig. 1) and identify every tube with a crescent number according to the retention time it elutes.

4. Chromatographic peaks collected manually should be individually injected and containing peptides sequenced by electrospray mass spectrometry (*see* **Note 16**). MS spectra are analyzed using the manufacturer's software (we use MassLynx 4.0 software, Waters). To identify peptides, MS/MS data were analyzed using the Mascot search engine (Matrix Science Ltd., UK). No cleavage site should be specified. Modifications can include N-terminal protein acetylation and methionine oxidation. Results should be manually checked and interpreted to eliminate false positives. To identify the putative protein precursors of the peptides sequenced by ESI-MS/MS, a protein data base (available on the World Wide Web) will be searched for short, nearly exact matches (rodent origin), as previously described [12]. The prediction of posttranslational modification sites in sequenced peptides will be able to be made with the ExPASy Proteomics Server on the World Wide Web.

## 4 Notes

1. Having the best water quality is the key if you are working with peptides and mass spectrometry. Make sure your purification system is all set up and filters have been replaced appropriately.
2. Before using your enzyme preparation make sure you run it on a SDS-PAGE and check its homogeneity. It is not uncommon that after freezing some proteases may aggregate and/or precipitate and start to show more bands than originally in the SDS-PAGE. The enzyme used in this step must be at least 95–98% pure on the gel after coomassie staining.
3. The strength of microwave ovens varies among different models. Before starting you must standardize these conditions, for example, using a digital temperature probe inserted into the animal organ to check the temperature after microwaving. You want to reach a temperature of 80–90 °C within the organ from which the peptides will be extracted. Temperatures in this range are ideal because they inactivate proteases that can

destroy the peptides you are trying to isolate. For small animal organs (e.g., mice brain) we usually add 2–3 beakers containing 50 mL of water into the microwave to help absorb microwave radiation and provide for more consistent heating of the tissue between replicates.

4. When homogenizing your tissue you want the temperature to be 80 °C or higher to inactivate proteases. However, before acidification with HCl (to reach pH 2), you must cool down your homogenate to prevent undesired hydrolysis of Pro-Asp peptide bonds, which can occur under hot acidic conditions.
5. This acidification step seems important to release peptides from the proteins, as it is believed that many peptides can escape further degradation binding to proteins inside the cells.
6. If you have a large volume of homogenate that contains considerable debris, you can centrifuge at 1000 × *g* for 10 min and use the supernatant for the next step. Usually this first centrifugation will clear your supernatant and make the next step easier.
7. You can skip this step if you don't have an ultracentrifuge. There will be not much difference in the peptide recovery at the end, but you may experience some delay during filtration in the next steps.
8. The amount of cells depends on its peptide content. For HEK293 cells, for example, two 150 mm cell culture plates containing confluent cells are fine. If you can start with more cells, it may be better. If you recover more peptides than required for subsequent steps, you can freeze them and use at later times.
9. We have found that five times of rinsing cells with PBS buffer greatly reduces contamination from fetal bovine serum. However, you can rinse the cells additional times, especially if the amount of serum is higher than the regular 10%, or if the amount of cells is larger than $1 \times 10^8$ cells. In these cases, you may also want to increase the PBS volume during the washing.
10. Before using the size-exclusion filter device you must wash it five times with ultrapure water to remove any traces of contaminants that may be present. For example, some size-exclusion filters contain polyethylene glycol (PEG) which greatly interferes in the mass spectrometry analysis. PEG also interferes with the nano columns.
11. We prefer using the Oasis HLB cartridge columns that cannot dry during the process without losing the chromatography capacity. Other types of reverse phase cartridge columns can be used but be aware to follow the instruction from the manufacturer. You can reduce the final volume of the eluent even more, but don't let the pellet dry—this will make the pellet hard to solubilize.

12. This step can be avoided if you don't trust fluorescamine to quantify peptides (some researchers suggest that fluorescamine can quantify other compounds such as free amino acids, and is not useful to determine the peptide concentrations of your samples), and/or you simply don't need to estimate the amount of peptide you have at this step. However, we strongly encourage the use of fluorescamine to quantify peptides at this point as it gives you at least a good estimative of how much peptide you recovered in your samples. In our laboratories we use a standard peptide of amino acid sequence LTLRTKL, because it contains one internal lysine plus the N-terminal amine. This is an empirical rough estimate of the average natural short peptide, which typically has two reactive amines to react with fluorescamine.
13. The incubation period and volume can be adapted to your experiments. In our experience, this time is sufficient to equilibrate the enzyme–substrate interactions. If you change the reaction volume make sure to use larger columns, keeping the proportion of 1/5 of the columns bed volume (here of 1 mL).
14. It is possible to use a quick spin into a micro centrifuge to speed the process. However, it is very important to use low-binding tubes. Many peptides bind to plastic tube surfaces, creating undesired bias to your experiments, as it will analyze only peptides that don't interact with the plastic surface. Low retention tubes reduce this bias.
15. Here, the recommendation is to use a pre-column as the BSA from the TBS/BSA buffer reduces the half-life of your C18 column.
16. Typical mass spectrometry conditions are: LC-MS/MS experiments can be performed on a Synapt G2 mass spectrometer coupled to a nanoAcquity capillary liquid chromatography (LC) system (Waters, Milford, MA, USA). The peptide mixture is desalted online for 3 min at a flow rate of 5 μL/min of phase A (0.1% formic acid) using a Symmetry C18 trapping column (5-μm particles, 180-μm inner diameters, 20-mm length; Waters). The mixture of trapped peptides is subsequently separated by elution with a gradient of 7–65% of phase B (0.1% formic acid in acetonitrile) through a BEH 130 C18 column (1.7-μm particles, 75-μm inner diameter, 100-mm length; Waters) in 42 min. The data is acquired in the data-dependent mode and the MS spectra of multiple-charged protonated peptides generated by electrospray ionization were acquired for 0.2 s from $m/z$ 300–1600. The three most intense ions exceeding base peak intensity threshold of 2500 counts are automatically mass selected and dissociated in MS/MS by 15- to 60-eV collisions with argon for 0.2 s. The typical LC and electrospray ionization conditions consist of a flow rate of 250 nL/min, a capillary voltage of 3.0 kV, a block temperature of 70 °C, and a cone voltage of 50 V. The dynamic peak exclusion window was set to 90 s.

## Acknowledgments

We are thankful to Rosangela Aparecida dos Santos Eichler for her critical reading and suggestions on this manuscript. This work was supported by Brazilian National Research Council (CNPq) grants 400944/2014-6, 445363/2014-2, and FAPESP (CeTICS 2013/07467-1 and Temático 2016/04000-3). ESF is a CNPq fellowship recipient.

### References

1. Ferro ES, Hyslop S, Camargo ACM (2004) Intracellullar peptides as putative natural regulators of protein interactions. J Neurochem 91:769–777
2. Fricker LD (2010) Analysis of mouse brain peptides using mass spectrometry-based peptidomics: implications for novel functions ranging from non-classical neuropeptides to microproteins. Mol Biosyst 6:1355–1365
3. Fricker LD (2005) Neuropeptide-processing enzymes: applications for drug discovery. AAPS J 7:E449–E455
4. Rioli V, Kato A, Portaro FC, Cury GK, te Kaat K, Vincent B et al (1998) Neuropeptide specificity and inhibition of recombinant isoforms of the endopeptidase 3.4.24.16 family: comparison with the related recombinant endopeptidase 3.4.24.15. Biochem Biophys Res Commun 250:5–11
5. Rioli V, Gozzo FC, Heimann AS, Linardi A, Krieger JE, Shida CS, Almeida PC, Hyslop S, Eberlin MN, Ferro ES (2003) Novel natural peptide substrates for endopeptidases 24.15, neurolysin, and angiotensin-converting enzyme. J Biol Chem 278:8547–8555
6. Cunha FM, Berti DA, Ferreira ZS, Klitzke CF, Markus RP, Ferro ES (2008) Intracellular peptides as natural regulators of cell signaling. J Biol Chem 283:24448–24459
7. Berti DA, Morano C, Russo LC, Castro LM, Cunha FM, Zhang X, Sironi J, Klitzke CF, Ferro ES, Fricker LD (2009) Analysis of intracellular substrates and products of thimet oligopeptidase in human embryonic kidney 293 cells. J Biol Chem 284:14105–14116
8. Che FY, Yan L, LI H, Mzhavia N, Lakshmi AD, Fricker LD (2001) Identification of peptides from brain and pituitary of Cpe(fat)/Cpe(fat) mice. Proc Natl Acad Sci U S A 98(17):9971–9976
9. Undenfriend S, Stein S, Bohlen P, Dairman W, Leimgruber W, Weigele M (1972) Fluorescamine: a reagent for assay of amino acids, peptides, proteins, and primary amines in the picomole range. Science 178:871–872
10. Saric T, Graef CI, Goldberg AL (2004) Pathway for degradation of peptides generated by proteasomes: a key role for thimet oligopeptidase and other metallopeptidases. J Biol Chem 279:46,723–46,732
11. Rioli V, Prezoto BC, Konno K, Melo RL, Klitzke CF, Ferro ES, Ferreira-Lopes M, Camargo AC, Portaro FC (2008) A novel bradykinin potentiating peptide isolated from Bothrops jararacussu venom using catallytically inactive oligopeptidase EP24.15. FEBS J 275:2442–2454
12. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389–3402

# Chapter 7

# Non-targeted Identification of D-Amino Acid-Containing Peptides Through Enzymatic Screening, Chiral Amino Acid Analysis, and LC-MS

**Hua-Chia Tai, James W. Checco, and Jonathan V. Sweedler**

## Abstract

D-Amino acid-containing peptides (DAACPs) in animals are a class of bioactive molecules formed via the posttranslational modification of peptides consisting of all-L-amino acid residues. Amino acid residue isomerization greatly impacts the function of the resulting DAACP. However, because isomerization does not change the peptide's mass, this modification is difficult to detect by most mass spectrometry-based peptidomic approaches. Here we describe a method for the identification of DAACPs that can be used to systematically survey peptides extracted from a tissue sample in a non-targeted manner.

**Key words** D-Amino acid-containing peptides, Posttranslational modifications, Bioactive peptides, Peptide isomerization, Chirality

## 1 Introduction

While peptides derived from ribosomal synthesis are translated exclusively using L-amino acids, a bioactive group of D-amino acid-containing peptides (DAACPs) exists in diverse animal species [1, 2]. Formation of D-amino acid residues in animal peptides often appears to result from a posttranslational modification (PTM) in which an L-amino acid residue is enzymatically converted into a D-amino acid residue in the peptide chain. This modification has a profound impact on the structure and functional properties of a peptide, and often leads to enhanced biological activity and increased protease stability for the DAACP relative to its all-L-residue counterpart [3–8].

Despite the functional importance of peptide isomerization, this PTM is seldom explored in peptidomic studies. This is because L- to D-residue isomerization does not change the mass or chemical composition of a peptide, making it difficult to detect in mass spectrometry (MS)-based peptide characterizations. Although a number of analytical approaches have been developed for differentiating peptide diastereomers using

Michael Schrader and Lloyd Fricker (eds.), *Peptidomics: Methods and Strategies*, Methods in Molecular Biology, vol. 1719, https://doi.org/10.1007/978-1-4939-7537-2_7, 

tandem MS (MS/MS), capillary electrophoresis, ion mobility MS, radical-directed dissociation MS, and other techniques [9–14], specific optimization of experimental conditions for a particular peptide of interest is often required. Therefore, these methods are useful for confirming the chirality of a peptide already suspected to exist as a DAACP, but are less efficient at identifying new DAACPs without prior knowledge. In this chapter we describe a non-targeted method to identify DAACPs extracted from complex biological samples [15]. This approach does not require prior knowledge about suspected DAACPs and can be applied to peptides extracted from various tissues or regions of interest across different animal models.

The first stage in this method involves a screening process based on the observation that DAACPs tend to have increased resistance to degradation by peptidases relative to their all-L-residue peptide analogs [16]. Since nearly all known DAACP neuropeptides and neurohormones have the D-residue near the N-terminus [2], an aminopeptidase should degrade these DAACPs at a slower rate than it degrades most peptides containing only L-residues. Aminopeptidase M (APM) is used in this protocol to assay complex mixtures of peptides extracted from animal tissues. By analyzing the APM reaction mixture over time using liquid chromatography coupled to mass spectrometry (LC-MS), peptides that are digested relatively slowly are marked as DAACP candidates and isolated for further study. In the second stage, purified DAACP candidates are hydrolyzed into their component amino acids, derivatized with Marfey's reagent to enhance chiral separation [17], and then analyzed with an LC-MS/MS system suitable for multiple reaction monitoring (MRM) to determine the chirality of the amino acids. If a peptide is found to contain a D-amino acid residue, the final stage is to chemically synthesize the proposed DAACP and compare its LC-MS retention time to that of the endogenous peptide. If the synthetic DAACP matches the retention time of the endogenous peptide obtained by LC-MS (and differs from that of the all-L-residue analog), then one can conclude that the endogenous peptide is indeed a DAACP. The method described is derived from our prior work characterizing DAACPs from the GFFD prohormone in *Aplysia californica* [8, 15, 18].

## 2 Materials

Prepare all solutions using LC-MS grade solvents, reagents and ultrapure water. Use low protein binding microcentrifuge tubes for peptide samples.

### 2.1 Aminopeptidase M Digestion

1. Aminopeptidase M, E.C. number 3.4.11.2, activity ≥50 units/mL.
2. Reaction buffer: 25 mM Tris–HCl, 0.5 M NaCl, pH 7.5.
3. pH paper.
4. 37 °C water bath or incubator.
5. Reversed-phase C18 solid-phase extraction columns for sample cleanup (e.g., ZipTip$_{C18}$ pipette tips or C18 spin columns). Wetting solution: 100% acetonitrile (ACN). Equilibration solution: 0.1% trifluoroacetic acid (TFA) in water. Wash solution: 5% methanol in water, 0.1% TFA. Elution solution: 50% ACN in water, 0.1% formic acid.
6. LC-MS/MS system for peptide characterization and for structure confirmation; e.g., a nanoLC system (we use Dionex Ultimate, Thermo Fisher Scientific) coupled to a QTOF mass spectrometer (we use Impact, Bruker).
7. Data analysis software for peptide identification; e.g., PEAKS Studio (Bioinformatics Solutions Inc.).

### 2.2 Chiral Amino Acid Analysis

1. Microwave reactor, e.g., Discover (CEM Corporation), with fiber optic temperature probe for acid hydrolysis, valve panel connected to a nitrogen source and a vacuum source, with vacuum gauge and pressure sensor, and Teflon PFA vessel with microvial insert tray and tubing.
2. 300 µL glass sample vials.
3. 200 µL extra-long pipet tips.
4. 6 N DCl with 1% phenol: add 2.5 mL of concentrated DCl to 2.5 mL of 2% (w/w) phenol dissolved in $D_2O$ (*see* **Note 1**).
5. Centrifugal evaporator; e.g., Savant SpeedVac (Thermo Scientific).
6. 0.5 M $NaHCO_3$ solution in water.
7. Marfey's reagent (1-fluoro-2,4-dinitrophenyl-5-L-alanine amide), 1 mg/mL dissolved in ACN.
8. Standard amino acids mixture: a 200 µM (total) mixture of glycine plus the L- and D-forms of the other 19 proteinogenic amino acids.
9. Mixing block with temperature control.
10. LC-MS/MS-MRM system; e.g., an UHPLC system (we use Advance, Bruker) coupled to a triple quadrupole mass spectrometer (we use EVOQ, Bruker).
11. LC column suitable for separating aromatic hydrocarbons; e.g., a phenyl-hexyl column (we use Kinetex, Phenomenex).
12. LC solvents: Solvent A: 25 mM ammonium formate in water; Solvent B: methanol.

## 3 Methods

### 3.1 Aminopeptidase M Digestion

1. Mix 15 μL reaction buffer with 5 μL peptide sample (5–10 μg in aqueous solution) in a low protein binding microcentrifuge tube. Check the pH of the solution with pH paper and adjust to ~pH 7.5 using 0.1 M HCl or 0.1 M NaOH solutions. Add 1 μL of APM (at 60 U/mL) and pipette up and down to mix. Save half of the mixture as the "0 h" sample and incubate the other half of the mixture at 37 °C for 24 h. For the 0 h sample, proceed immediately to the next step; for the 24 h sample, perform sample cleanup after incubation (*see* **Note 2**).
2. Sample cleanup by solid-phase extraction using reversed-phase C18 pipette tips: adjust sample to 0.1% TFA, pH < 4. Attach pipette tip to a pipettor set at 10 μL maximum volume. Wash the tip twice by slowly aspirating the wetting solution, then dispense to waste. Repeat twice with the equilibration solution. Aspirate and dispense the sample for 10 cycles to bind peptides to the pipette tip. Wash the tip twice with the wash solution to remove salts, contaminants, and unbound molecules. Elute the purified peptide sample in 5 μL of elution solution. Samples can be stored at −20 °C (*see* **Note 3**).
3. Analyze samples with the LC-MS/MS system for peptide characterization. Use the data analysis software to identify peptide sequences through database searching and/or *de novo* sequencing.
4. Compare peptide content between the 0 and 24 h APM digestion samples to identify peptides that reproducibly show resistance to digestion by APM as possible DAACP candidates (Fig. 1) (*see* **Note 4**).
5. Isolate DAACP candidates using successive rounds of HPLC purifications. Detailed strategies vary with each specific peptide. Different LC gradients, solvents, and column binding properties are commonly employed to achieve peptide isolation [19, 20]. At least 5 nmol of each peptide should be isolated (purity >80%) for the next stage of analysis (*see* **Note 5**).

### 3.2 Chiral Amino Acid Analysis

1. Place each glass vial inside a 1.5 mL microcentrifuge tube for easy handling. Transfer peptide sample (1–5 nmol in solution) into a glass vial, then evaporate the solvent using a centrifugal evaporator (*see* **Note 6**).
2. Add 5 mL of 6 N DCl with 1% phenol to the bottom of the Teflon PFA vessel. Place glass vials containing the dried peptide samples into the microvial insert tray, and place the tray inside the vessel body. Assemble the vessel and seal tightly, then connect the vessel to the valve panel and place it into the microwave

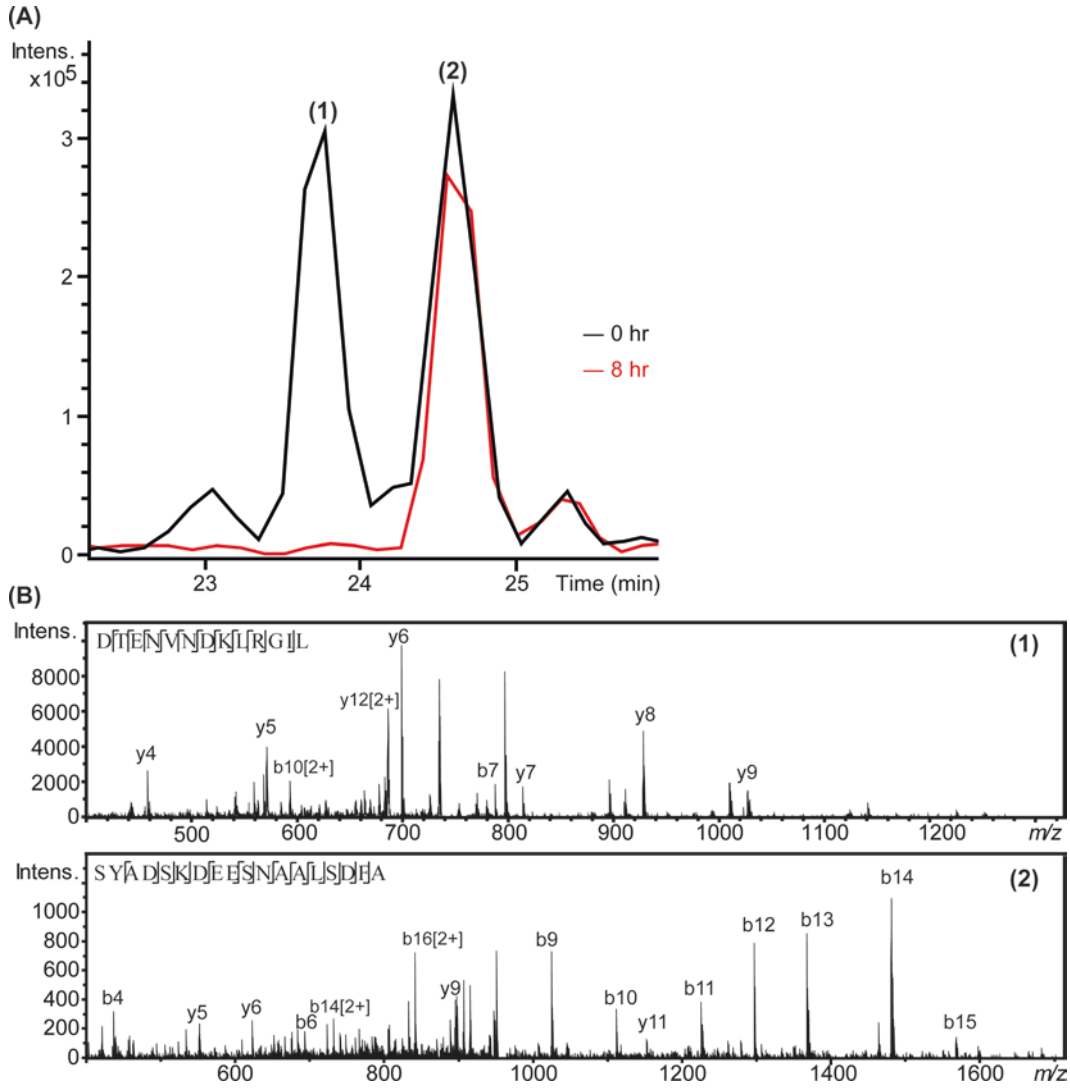



**Fig. 1** Identify peptides that are resistant to APM by comparing peptide signals before (black chromatogram) and after (red chromatogram) digestion. In this example, an extracted peptide mixture from *Aplysia* ganglia was subjected to APM digestion for 8 h at 37 °C. (**a**) LC-MS results showed that peptide (1) was degraded while peptide (2) remained at similar levels throughout the APM digestion. Therefore, peptide (2) becomes a putative DAACP candidate and was isolated for further analysis. (**b**) The identities of peptides (1) and (2) were determined based on MS/MS data in a database search

reactor cavity. The valve panel should be connected to a nitrogen source (set at 15 psi) and a vacuum source. Flush the vessel with nitrogen for 10 s, then evacuate down to −20 mmHg. Repeat for three cycles. Adjust the final pressure to 5–10 psi of nitrogen in the vessel. Insert the fiber optic temperature probe (*see* **Note 7**).

3. Run the microwave reactor using the following conditions: power = 200 W, pressure limit = 120 psi, temperature = 165 °C,

and hold time = 15 min (power and pressure settings may need to be adjusted based on the specific instrument used). At the end of the hydrolysis run, allow the vessel to cool before disassembling.

4. Remove glass vials from the vessel using forceps and place them back into 1.5 mL microcentrifuge tubes. Dry sample vials in the centrifugal evaporator to remove any solvent introduced to the sample during the hydrolysis.
5. Using extra-long pipet tips, add 25 μL of 0.5 M $NaHCO_3$ solution to each glass vial to redissolve the amino acid hydrolysate. Transfer samples to 0.5 mL microcentrifuge tubes and add in 20 μL of 1 mg/mL Marfey's reagent in ACN. In a separate vial, also prepare a sample consisting of 25 μL of the standard amino acids mixture and 20 μL of Marfey's reagent. Place tubes in a heated mixing block at 60 °C for 3 h under gentle mixing.
6. Separately prepare amino acid standards derivatized with Marfey's reagent and use them to establish MRM channels for each derivatized amino acid on the LC-MS-MRM system. Ensure that the LC method is capable of separating the L- and D-forms of every derivatized amino acid. Our setup uses a phenyl-hexyl column (2.6 μm particle size, 100 Å pore size, 100 × 2.1 mm, Phenomenex Kinetex) with a gradient elution using 25 mM ammonium formate as Solvent A and methanol as Solvent B, at a flow rate of 300 μL/min. The gradient is as follows: 5% B for 2 min, 5–15% B over 5 min, 15–60% B over 5 min, 60% B for 3 min, 60–100% B over 3 min, 100% B for 3 min, 100–5% B over 1 min, and 5% B for 2 min (*see* **Note 8**).
7. Analyze the derivatized amino acid hydrolysate samples on the LC-MS/MS-MRM system. Also include a run of the standard amino acid mixture in the same batch. Determine the chirality of each amino acid residue from a peptide sample by matching the peak retention time from the experimental sample to the retention times of the L- or D-amino acid standards (Fig. 2). For peptides that are suspected to contain a D-amino acid residue based on the chiral amino acid analysis, proceed to the next stage for final structure confirmation.

### 3.3 LC-MS for Structure Confirmation

1. Synthesize both the all-L-residue peptide and the proposed DAACP through commercial custom peptide synthesis or in-house synthesis (*see* **Note 9**).
2. Develop an LC method that separates the two peptide isomers on the LC-MS/MS system for structure confirmation.
3. Analyze each synthetic peptide separately on this LC-MS/MS system to establish retention times, MS, and MS/MS data for these standards. Analyze a sample of the endogenous peptide on the LC-MS/MS system. The existence of an endogenous

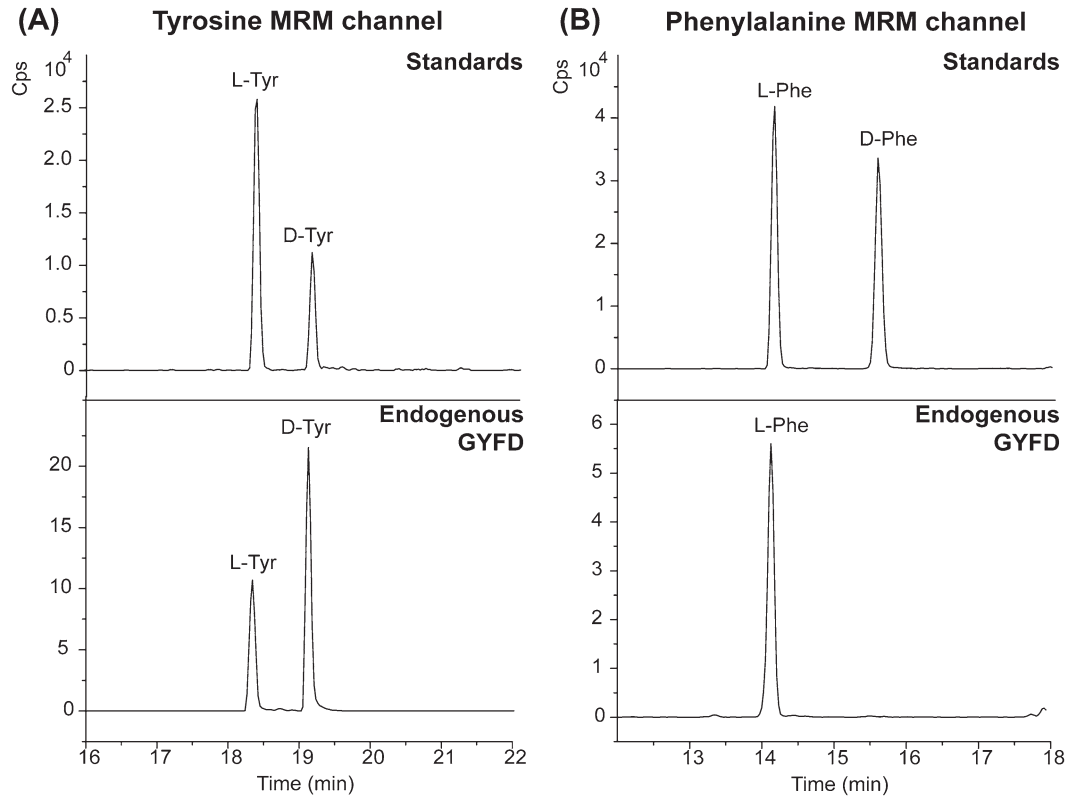



**Fig. 2** Chiral amino acid analysis determines the chirality of amino acid residues in a candidate peptide. In this example of a chiral analysis on the extracted *Aplysia* peptide GYFD, D-tyrosine was detected in the tyrosine MRM channel (**a**) while only L-phenylalanine was detected in the phenylalanine MRM channel (**b**). This suggests that the endogenous peptide exists as GdYFD

DAACP can be confirmed if the retention time, mass, and MS/MS fragmentation pattern of the endogenous peptide matches that of the synthetic DAACP (Fig. 3) [15].

## 4 Notes

1. Prepare this solution immediately before use. "D" refers to deuterium (i.e., $^{2}H$). DCl and $D_2O$ are used in place of HCl and $H_2O$ so that D-amino acids resulting from spontaneous racemization during acid hydrolysis would have a mass shift of +1 Dalton and thus be distinguished from any D-amino acid residue originally present in the peptide [21]. Phenol is added to prevent degradation of tryptophan and other amino acids [22].
2. The peptide sample should be prepared from peptides extracted from a target tissue using a suitable peptide extraction procedure, dried, and then reconstituted in aqueous solution. Peptide extraction protocols differ based on the tissue

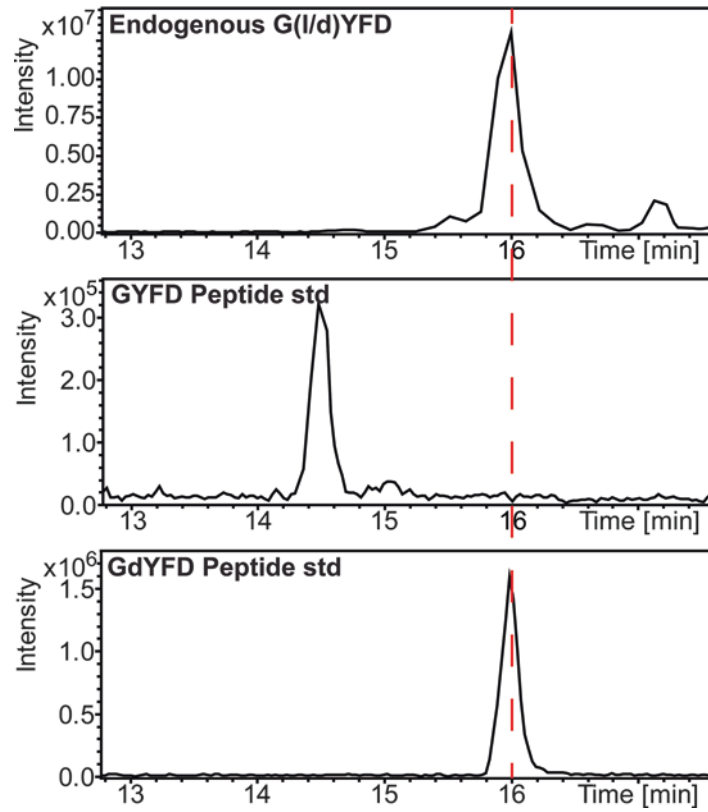



**Fig. 3** Structure confirmation using synthetic peptide standards. LC-MS/MS analysis showed that the retention time (base peak chromatograms shown above) of the endogenous peptide, after 48 h of APM digestion, matches that of the synthetic DAACP standard, thereby confirming the structure of the endogenous peptide as GdYFD. (Adapted with permission from ref. [15], Copyright 2016 American Chemical Society)

type and are not described here. Total peptide concentration may be measured with a commercial kit (e.g., Pierce Quantitative Colorimetric Peptide Assay). One or more all-L-residue peptide standards (avoid peptides with pyroglutamate or proline near the N-terminus since these peptides may have higher resistance to APM) and DAACP standards (e.g., deltorphin) may be spiked into the sample as controls for enzyme activity. Choose peptides that are not endogenous to the biological sample for these controls. Each all-L-residue peptide standard should be present in the analysis of the 0 h time point but absent after 24 h. Failure to degrade all-L-residue standards indicates insufficient APM activity. Reaction conditions (APM concentration, digestion time) should then be adjusted accordingly. Each DAACP standard should be present at the 24 h time point at concentrations similar to

those seen at the 0 h time point. A dramatic loss of DAACP standards in the 24 h time point indicates excess APM or digestion time is being used. Test the activity of each batch of APM and adjust the enzyme concentration and reaction time to allow for full hydrolysis of all-L-residue standards while leading to minimal degradation of DAACP standards.

3. A 2% solution of aqueous TFA can be used to adjust the sample to a final concentration of 0.1% TFA. The low pH promotes peptide binding to the C18 stationary phase of the pipette tip, and also inactivates APM to stop the digestion reaction. Sample cleanup by solid-phase extraction is intended to desalt, purify, and concentrate peptide samples prior to LC-MS analysis. The capacity of a $ZipTip_{C18}$ pipette tip is typically 5 μg. If higher binding capacity or higher loading volume (>15 μL) is needed, a good alternative is the Pierce C18 spin column, which can process samples of 10–150 μL and has a peptide binding capacity of up to 30 μg.
4. Note that not all APM-resistant peptides are expected to be DAACPs. Some N-terminal modifications (e.g., pyroglutamylation, acetylation) and certain sequences (e.g., proline residues near the N-terminus) can also increase a peptide's resistance to APM digestion [23]. It is also possible that some DAACPs may be degraded by APM, depending on the position of the D-residue in the peptide sequence. DAACPs with a D-residue far from the N-terminus may be degraded in this screening procedure (although they may give rise to truncated peptide fragments as APM stalls its digestion near the D-residue).
5. Peptide concentration can be estimated by UV absorbance at 280 nm (for peptides containing tryptophan or tyrosine), 214 nm, or 205 nm [24–26], or with a commercial peptide assay kit (e.g., Pierce Quantitative Colorimetric Peptide Assay). In addition, MALDI-TOF MS may be used to assess the integrity and purity of the isolated peptide. The peptide purity needed for chiral analysis is flexible and can be determined on a case by case basis. If other peptides in the sample do not contain the amino acid residue suspected to be a D-amino acid, then these impurities may not interfere with interpretation of the chiral analysis data and a lower purity is acceptable. On the other hand, if the residue in question is present in multiple peptides in the sample, a higher purity might be desired at this stage to avoid testing a large number of peptide conformations in the confirmation stage.
6. Transfer peptide sample in a solution of 20–100 μL so that the peptide sample is evenly distributed around the bottom of the vial rather than in a small clump; this allows for a more complete hydrolysis of the peptides.

7. To check if the hydrolysis vessel is sealed properly after connecting the tubing, fill the vessel with nitrogen (15 psi), then turn the valve to the "run" position. Make sure the pressure reading is stable (for 30 s) before moving on to the vacuum/nitrogen cycle. If the pressure drops while the valve is in the "run" position, check and tighten the ferrules, cap, and connections.
8. MRM parameters will vary with the MS instrument. Table 1, adapted from prior work [15], shows the MRM transition, collision energy, and ionization mode of each derivatized amino acid analyzed on a Bruker EVOQ triple quadrupole mass spectrometer using a heated electrospray ionization source with the following settings: spray voltage at 3500 V, cone temperature at 250 °C, cone gas flow at 20 units, probe

**Table 1**
**MRM parameters for amino acids derivatized with Marfey's reagent**

| Amino acid | Parent ion (*m/z*) | Collision energy (eV) | Fragment ion (*m/z*) | Mode |
|---|---|---|---|---|
| Alanine | 340.0 | 14 | 278.1 | – |
| Arginine | 426.9 | 10 | 70.1 | + |
| Asparagine | 382.9 | 33 | 175.9 | – |
| Aspartic acid | 384.0 | 24 | 267.9 | – |
| Cysteine | 372.1 | 21 | 284.9 | – |
| Glutamic acid | 398.0 | 24 | 202.0 | – |
| Glutamine | 396.9 | 20 | 353.0 | – |
| Glycine | 326.1 | 32 | 162.0 | – |
| Histidine | 658.2 | 30 | 549.1 | – |
| Isoleucine/Leucine | 382.0 | 18 | 319.8 | – |
| Lysine | 649.1 | 36 | 479.1 | – |
| Methionine | 400.0 | 17 | 337.9 | – |
| Phenylalanine | 416.0 | 20 | 337.1 | – |
| Proline | 365.9 | 15 | 321.9 | – |
| Serine | 355.9 | 16 | 263.9 | – |
| Threonine | 370.0 | 17 | 263.9 | – |
| Tryptophan | 457.1 | 10 | 188.0 | + |
| Tyrosine(+2 Da)[a] | 686.1 | 28 | 353.9 | – |
| Valine | 367.9 | 15 | 306.0 | – |

[a]Derivatized tyrosine was observed with a mass shift of +2 Da following DCl hydrolysis (but not HCl hydrolysis), perhaps due to the incorporation of deuterons at positions *ortho* to the phenolic hydroxyl group [27]. (Table adapted with permission from ref. [15], Copyright 2016 American Chemical Society)

temperature at 400 °C, and probe gas flow at 45 units. These parameters may be used as a starting point for new method development.

9. Since chiral amino acid analysis does not provide information about the position of the D-residue in the peptide, multiple conformations may need to be synthesized. For example, if a D-phenylalanine signal was detected during chiral analysis, but the peptide sequence has more than one phenylalanine near the N-terminus, then peptide structures with D-phenylalanine at each position should be considered.

## Acknowledgments

This work was supported by the National Institutes of Health, Award No. P30 DA018310 from the National Institute on Drug Abuse (NIDA) and Award No. 2 R01NS031609 from the National Institute of Neurological Disorders and Stroke (NINDS). J.W.C. was supported in part by a Beckman Institute Postdoctoral Fellowship, funded by a Beckman Foundation gift to the Beckman Institute for Advanced Science and Technology at the University of Illinois at Urbana-Champaign. The content is solely the responsibility of the authors and does not necessarily represent the official views of the funding agencies.

### References

1. Bai L, Sheeley S, Sweedler JV (2009) Analysis of endogenous D-amino acid-containing peptides in Metazoa. Bioanal Rev 1(1):7–24. https://doi.org/10.1007/s12566-009-0001-2
2. Ollivaux C, Soyez D, Toullec JY (2014) Biogenesis of D-amino acid containing peptides/proteins: where, when and how? J Pept Sci 20(8):595–612. https://doi.org/10.1002/psc.2637
3. Richter K, Egger R, Kreil G (1987) D-alanine in the frog skin peptide dermorphin is derived from L-alanine in the precursor. Science 238(4824):200–202
4. Kamatani Y, Minakata H, Kenny PT, Iwashita T, Watanabe K, Funase K, Sun XP, Yongsiri A, Kim KH, Novales-Li P et al (1989) Achatin-I, an endogenous neuroexcitatory tetrapeptide from Achatina fulica Ferussac containing a D-amino acid residue. Biochem Biophys Res Commun 160(3):1015–1020
5. Ohta N, Kubota I, Takao T, Shimonishi Y, Yasuda-Kamatani Y, Minakata H, Nomoto K, Muneoka Y, Kobayashi M (1991) Fulicin, a novel neuropeptide containing a D-amino acid residue isolated from the ganglia of Achatina fulica. Biochem Biophys Res Commun 178(2):486–493. https://doi.org/10.1016/0006-291X(91)90133-R
6. Buczek O, Yoshikami D, Bulaj G, Jimenez EC, Olivera BM (2005) Post-translational amino acid isomerization: a functionally important D-amino acid in an excitatory peptide. J Biol Chem 280(6):4247–4253. https://doi.org/10.1074/jbc.M405835200
7. Ollivaux C, Gallois D, Amiche M, Boscameric M, Soyez D (2009) Molecular and cellular specificity of post-translational aminoacyl isomerization in the crustacean hyperglycaemic hormone family. FEBS J 276(17):4790–4802. https://doi.org/10.1111/j.1742-4658.2009.07180.x
8. Bai L, Livnat I, Romanova EV, Alexeeva V, Yau PM, Vilim FS, Weiss KR, Jing J, Sweedler JV (2013) Characterization of GdFFD, a D-amino acid-containing neuropeptide that functions as an extrinsic modulator of the Aplysia feeding circuit. J Biol Chem 288(46):32837–32851. https://doi.org/10.1074/jbc.M113.486670

9. Bai L, Romanova EV, Sweedler JV (2011) Distinguishing endogenous D-amino acid-containing neuropeptides in individual neurons using tandem mass spectrometry. Anal Chem 83(7):2794–2800. https://doi.org/10.1021/ac200142m
10. Sachon E, Clodic G, Galanth C, Amiche M, Ollivaux C, Soyez D, Bolbach G (2009) D-amino acid detection in peptides by MALDI-TOF-TOF. Anal Chem 81(11):4389–4396. https://doi.org/10.1021/ac9002886
11. Koehbach J, Gruber CW, Becker C, Kreil DP, Jilek A (2016) MALDI TOF/TOF-based approach for the identification of D-amino acids in biologically active peptides and proteins. J Proteome Res 15(5):1487–1496. https://doi.org/10.1021/acs.jproteome.5b01067
12. Sheeley SA, Miao H, Ewing MA, Rubakhin SS, Sweedler JV (2005) Measuring D-amino acid-containing neuropeptides with capillary electrophoresis. Analyst 130(8):1198–1203. https://doi.org/10.1039/b504717j
13. Jia C, Lietz CB, Yu Q, Li L (2014) Site-specific characterization of D-amino acid containing peptide epimers by ion mobility spectrometry. Anal Chem 86(6):2972–2981. https://doi.org/10.1021/ac4033824
14. Tao Y, Quebbemann NR, Julian RR (2012) Discriminating D-amino acid-containing peptide epimers by radical-directed dissociation mass spectrometry. Anal Chem 84(15):6814–6820. https://doi.org/10.1021/ac3013434
15. Livnat I, Tai HC, Jansson ET, Bai L, Romanova EV, Chen TT, Yu K, Chen SA, Zhang Y, Wang ZY, Liu DD, Weiss KR, Jing J, Sweedler JV (2016) A D-amino acid-containing neuropeptide discovery funnel. Anal Chem 88(23):11868–11876. https://doi.org/10.1021/acs.analchem.6b03658
16. Ewing MA, Wang J, Sheeley SA, Sweedler JV (2008) Detecting D-amino acid-containing neuropeptides using selective enzymatic digestion. Anal Chem 80(8):2874–2880. https://doi.org/10.1021/ac7025173
17. Bhushan R, Bruckner H (2004) Marfey's reagent for chiral amino acid analysis: a review. Amino Acids 27(3–4):231–247. https://doi.org/10.1007/s00726-004-0118-0
18. Yang CY, Yu K, Wang Y, Chen SA, Liu DD, Wang ZY, YN Su, Yang SZ, Chen TT, Livnat I, Vilim FS, Cropper EC, Weiss KR, Sweedler JV, Jing J (2016) Aplysia locomotion: network and behavioral actions of GdFFD, a D-amino acid-containing neuropeptide. PLoS One 11(1):e0147335. https://doi.org/10.1371/journal.pone.0147335
19. Finoulst I, Pinkse M, Van Dongen W, Verhaert P (2011) Sample preparation techniques for the untargeted LC-MS-based discovery of peptides in complex biological matrices. J Biomed Biotechnol 2011:245291. https://doi.org/10.1155/2011/245291
20. Conlon JM (2007) Purification of naturally occurring peptides by reversed-phase HPLC. Nat Protoc 2(1):191–197. https://doi.org/10.1038/nprot.2006.437
21. Liardon R, Ledermann S, Ott U (1981) Determination of D-amino acids by deuterium labelling and selected ion monitoring. J Chromatogr A 203:385–395. https://doi.org/10.1016/S0021-9673(00)80309-X
22. Fountoulakis M, Lahm HW (1998) Hydrolysis and amino acid composition of proteins. J Chromatogr A 826(2):109–134
23. Turner AJ (2013) Chapter 79—Aminopeptidase N In: Rawling ND, Salvesen G (eds) Handbook of Proteolytic Enzymes. Academic Press, London, pp 397–403. https://doi.org/10.1016/B978-0-12-382219-2.00079-X
24. Gill SC, von Hippel PH (1989) Calculation of protein extinction coefficients from amino acid sequence data. Anal Biochem 182(2):319–326. https://doi.org/10.1016/0003-2697(89)90602-7
25. Conibear AC, Daly NL, Craik DJ (2012) Quantification of small cyclic disulfide-rich peptides. Biopolymers 98(6):518–524. https://doi.org/10.1002/bip.22121
26. Anthis NJ, Clore GM (2013) Sequence-specific determination of protein and peptide concentrations by absorbance at 205 nm. Protein Sci 22(6):851–858. https://doi.org/10.1002/pro.2253
27. Murai Y, Wang L, Masuda K, Sakihama Y, Hashidoko Y, Hatanaka Y, Hashimoto M (2013) Rapid and controllable hydrogen/deuterium exchange on aromatic rings of α-amino acids and peptides. Eur J Org Chem 2013(23):5111–5116. https://doi.org/10.1002/ejoc.201300405

# Part II

## Quantitative Methods in Peptidomics

# Chapter 8

# Quantitative Peptidomics: General Considerations

Lloyd Fricker

## Abstract

Peptidomics is the detection and identification of the peptides present in a sample, while quantitative peptidomics provides additional information about the amounts of these peptides. Comparison of peptide levels among two or more samples is termed relative quantitation. It is also possible to perform absolute quantitation of peptide levels in which the biological sample is compared to synthetic standards, which requires a separate standard for each peptide. In contrast, relative quantitation can compare levels of all peptides that are detectable in a sample, which can exceed 1000 peptides in a complex sample. In this chapter, various techniques used for quantitative peptidomics are described along with discussion of the advantages and disadvantages of each approach. A guide to selecting the optimal quantitative approach is provided, based on the goals of the experiment and the resources that are available.

**Key words** TMAB, Formaldehyde, Isotopic label, Isobaric tags, iTRAQ, DiLeu, Label-free quantitation

## 1 Introduction

Quantitative peptidomics provides information about the relative levels of peptides between two or more samples, and has found many uses for biological studies. Unlike radioimmunoassays, which measure levels of a single peptide in a complex sample, peptidomics approaches can measure the relative levels of hundreds or thousands of peptides present in the sample. Radioimmunoassays require a specific antiserum for each peptide and can detect low abundance peptides that are not readily detected using peptidomics approaches [1]. While radioimmunoassays are useful to measure absolute levels of a specific peptide, they are not able to measure levels of the vast majority of peptides present in biological samples, and often cannot distinguish between highly similar peptides that differ by a single amino acid residue or posttranslational modifications (e.g., phosphorylation, amidation). In contrast, quantitative peptidomics can easily discriminate peptides with posttranslational modifications that alter the mass of the peptide [2–4]. While quantitative peptidomics are powerful techniques, the results need to be interpreted with caution, as would the results from any study that measures peptides (*see* **Note 1**).

Michael Schrader and Lloyd Fricker (eds.), *Peptidomics: Methods and Strategies*, Methods in Molecular Biology, vol. 1719, https://doi.org/10.1007/978-1-4939-7537-2_8, 

Two fundamentally different approaches have been used for quantitative peptidomics. One approach takes the samples to be compared and labels them with different stable isotopes (*see* **Note 2**). After labeling, the samples are pooled, fractionated by reversed-phase liquid chromatography (LC), and analyzed by mass spectrometry (MS), usually electrospray ionization (ESI)-MS [3]. The heavy and light forms of the peptide can be distinguished by MS and the relative peak intensity of each form can be determined, providing a measure of the relative levels of peptide present in the different biological samples (Fig. 1). The isotopic labels can be incorporated into the peptides during their biosynthesis in cell culture, which is termed Stable Isotope Labeling with Amino Acids in Cell Culture (SILAC) [5, 6]. It is also possible to label whole organisms with heavy isotopes and compare with organisms exposed to the natural abundance of these isotopes (commonly referred to as "light"); this is described in detail in another chapter in this volume (Kunz et al.). While useful for small organisms such as *Drosophila* (Kunz et al.), the cost of the stable isotopes is very high for studies on mice and larger organisms, and this approach cannot be used for studies on human samples. More commonly, the "heavy" and "light" isotopes are incorporated into peptides using synthetic chemicals that covalently react with the peptides. Several different isotopic tags have been used for quantitative peptidomics and the most commonly used reagents are described in this chapter. Some of these reagents are covered in more detail in other chapters in this volume (Boonen et al., Dasgupta et al., DeLaney et al.).

It is possible to perform quantitative peptidomics without using isotopic labels—this is referred to as label-free peptidomics. Because the label-free approach avoids the labeling step, it may seem that the label-free approach is simpler than quantitative approaches using isotopic labels. However, label-free approaches are usually more complicated because they require many more LC runs and subsequent MS analysis (either by LC/MS with ESI, or MALDI analysis of the LC fractions). Furthermore, label-free approaches previously used for peptidomic studies have not reliably quantified small changes of 50% or less [7], although there are some mass spectrometers that can provide highly accurate label-free quantitation (*see* Subheading 1.1). In contrast, some studies with isotopic labels have detected statistically significant differences as small as 15%, providing that there is low variability among the replicates [8–13]. One reason for the greater accuracy of quantitation using isotopic labels is that the replicates being compared are analyzed on the same LC-MS run, whereas for label-free quantification each replicate is run separately, introducing run-to-run variation. This run-to-run variation can be reduced by performing technical replicates of each biological replicate and averaging the technical replicates, but this adds to the total number of runs. For example, a study to compare five biological replicates of a treated sample with five biological replicates of a control sample using label-free quantitation with three technical

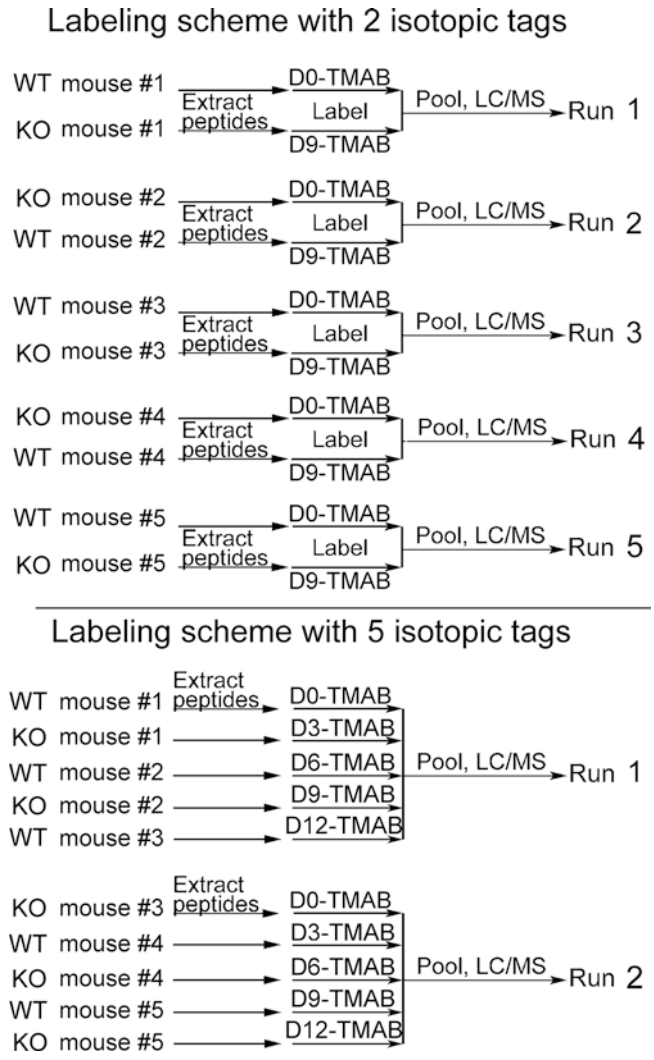



**Fig. 1** General scheme for quantitative peptidomics using two isotopic tags (top) or five isotopic tags (bottom). In both schemes, a total of five wild-type (WT) and five mutant mice (knock-out, or KO) are shown, but the actual number of replicates needs to take into account the experimental objectives, animal-to-animal variation, and the expected degree of the changes. To accurately measure small changes, a large number of biological replicates are needed, especially for peptides that show large animal-to-animal variation. Note that in both schemes, the two conditions (WT and KO) are switched between runs to control for possible variation in the chemical reactivity of the different isotopic reagents, and it is important that replicates "reverse" the labels. For example, in the top panel with two labels, runs 1, 3, and 5 involve WT samples labeled with D0, while in runs 2 and 4 the WT samples are labeled with D9. In the bottom panel with five labels, the WT samples are labeled with D0, D6, and D12 (Run 1) or with D3 and D9 (Run 2). As shown, the experiment with two isotopic tags requires five LC/MS runs, but this does not include separate runs that compare WT and WT animals. Thus, five additional LC/MS runs are needed in order to establish the baseline of variation of each peptide among biological replicates, so that statistical calculations can compare the observed KO/WT ratio to the WT/WT ratio. However, when using five isotopic tags, it is possible to compare WT/WT variations in the same experiment used to measure the KO/WT levels, and only two LC/MS runs are necessary for the entire experiment. To quantify using label-free approaches would require 30 LC/MS runs if three technical replicates were included for each of the five WT and five KO mice. Figure modified from [65]

replicates for each sample would require 30 LC/MS runs. In contrast, these samples could be analyzed on a single LC/MS run if using a 10-plex set of isotopic tags, or in two LC/MS runs if using a 5-plex set of isotopic tags (Fig. 1).

This chapter provides a brief overview of the major approaches that have been used for quantitative peptidomics. In addition, a general list of materials required for quantitative peptidomics is provided, and an overview of the technical methods. Guidelines are provided to help select the optimal method based on the experimental objectives and the resources that are available.

### 1.1 Label-Free Approaches

Many proteomic studies, and some peptidomic studies, have quantified protein/peptide levels by label-free approaches [7, 14–17]. In general, MS is not quantitative when comparing two or more distinct peptides because the ionization efficiency of peptides is dependent on their amino acid composition and posttranslational modifications. One solution is to synthesize standard peptides and analyze them on the same LC/MS system that will be used for the biological samples. Some instruments, such as triple quadrupole MS, are capable of high accuracy quantitation of peptides and other analytes. Large libraries containing tens of thousands of peptides have been synthesized and used to quantify levels of proteins in biological samples [18, 19]. Although not yet applied to peptidomic applications, the basic approach should be possible to adapt from proteomics to peptidomics applications for accurate label-free quantitation of known peptides in a biological sample.

For proteomic applications, most of the approaches include data from multiple peptides that are derived from the protein by enzymatic digestion, usually with trypsin [14]. By including distinct peptides from a particular protein, the variations in ionization efficiency between peptides are averaged out and the results can be reasonably accurate for proteomic studies [14]. For example, a typical protein is cleaved into 40 or more peptides of which many can be detected on LC/MS analysis, thus providing a large number of data points for consideration. In contrast, peptidomics is usually aimed at measuring individual peptides, and peptides from the same precursor may not be co-regulated. For example, distinct neuropeptides are often produced from the same prohormone through differential proteolytic cleavages, and their regulation can be distinct [20]. Thus, while label-free approaches are common for proteomic applications, these techniques are difficult to adapt to peptidomic studies with high accuracy unless a large number of replicates are performed or synthetic standards of each peptide are produced and analyzed.

There are three basic methods for label-free quantitation: peak intensity, spectral counting, and selected reaction monitoring (SRM). Of these, SRM is the most accurate and has the largest dynamic range. This general approach has been used for many years by the pharmaceutical industry to determine plasma or tissue levels of drugs and other

small molecules. SRM is usually performed on a triple quadrupole mass spectrometer, and requires a synthetic peptide standard for each peptide to be quantified. Thus, this technique is limited to known peptides for which standards are available, and while used for large-scale proteomic analysis [18, 19] it has not been used for peptidomic studies.

The other two label-free approaches have been used for peptidomic studies [7]. As the name implies, peak intensity involves the measurement of the ion current of the peptide as it elutes from LC/MS, usually integrating the signal over the entire elution period, but on occasion taking the peak value. Spectral counting is simpler, and counts only the number of times a peptide is observed. For proteomics this can be reasonably accurate because an abundant protein is likely to result in a larger number of detected fragments than a less abundant protein. However, this approach is quite limited for peptidomics because of the small number of data points. For example, if a peptide is detected in 2 out of 3 replicates of a treated sample, but only 1 out of 3 replicates of a control sample, are levels twice as high in the treated sample? It is possible, but there is no confidence in this interpretation due to the low number of data points. For this reason, it is important to include a large number of replicates when performing label-free peptidomics. Furthermore, there is a difference between a peptide being present in a sample and the identification of that peptide from its fragmentation pattern during MS/MS analysis and subsequent database searches. Only a fraction of the peptides present in a sample can usually be identified by MS/MS analysis and database searches. For complex mixtures of peptides analyzed by LC/MS, the MS instrument only has time to select a subset of peptides for fragmentation. If not selected for fragmentation, there is no MS/MS data and the peptide cannot be identified in that run. Furthermore, of the peptides selected for fragmentation, many do not result in a sufficient number of fragment ions that can allow for the sequence to be determined (this is especially a problem for short peptides, but can be a problem for peptides of any length). Thus, if the spectral counting is based only on peptides identified in each run by database searches of MS/MS data, then this approach will incorrectly miss many peptides and this can lead to major errors in the analysis. Once again, this is less of a problem with proteomics because multiple peptides are detected for each protein and the results averaged, but for peptidomics this represents a major problem.

Label-free quantitation using peak intensity takes into account the relative levels of the peptide in each LC/MS run. It is difficult to handle missing data—if a peptide is found in 2 out of 3 runs, is the run without the peptide considered "zero" and averaged with the other 2 runs, or is the missing run not included in the average? Neither of these solutions to missing data is ideal, especially if the peptide isn't really "missing" from the MS data (i.e., if it wasn't detected from database searches due to missing or low quality MS/MS data, but was present in the MS spectra).

Success with either of these label-free approaches depends on several factors. First, a key consideration for label-free (as well as methods using isotopic labels) is the quality of the samples and reproducibility of the peptide extraction. Postmortem changes and loss of peptide can introduce variability, as can the presence of contaminants. The number of replicates is important, both biological (i.e., separate samples from different animals or plates of cells) and technical (i.e., repeat LC/MS runs of the same biological sample). Some investigators leave out the technical replicates and perform a larger number of biological replicates in cases where the sample preparation is relatively simple. If a large number of biological replicates are difficult to obtain, technical replicates can be included to increase the total number of replicates. The instruments available for MS analysis can also influence the results, as some are more quantitative than others (such as triple quadrupole MS instruments). This is also a consideration if the results are dependent on peptide identifications in each run from MS/MS data—if so, then instruments capable of rapid MS/MS sampling times are ideal (*see* **Note 3**). Another important consideration is the software available for analysis, much of which has been developed for proteomic (and not peptidomic) applications.

A final point to consider is the magnitude of the changes that are expected between the sample groups. If the changes are orders of magnitude, it will be difficult to determine the precise value (e.g., 10-fold increase versus 20-fold increase) but relatively easy to determine that there was a major change (as shown in representative data in Fig. 2). This latter point is also true for methods that use isotopic (Fig. 3) and isobaric labels. As mentioned above, label-free approaches with SRM using a triple quadrupole mass spectrometer can detect a large dynamic range and accurately quantify both large and small changes of peptides, providing that synthetic standard peptides are available. If these resources are not available and small changes in peptide levels are anticipated (20–50%), then isotopic or isobaric tags are preferable to label-free approaches.

### 1.2 Isotopic Tags

Over the past 15 years, a number of different isotopic labels have been developed for proteomics and some have been tested for peptidomics [21–27]. None of these labels are truly ideal, although some are much better than others. One key property of the ideal isotopic label is that it can be incorporated into every peptide present in the sample. For this reason, one of the earliest reagents developed for proteomics, named ICAT (isotope-coded affinity tag), is not useful for general peptidomic studies because the ICAT reagent binds to Cys residues—these are rarely found in peptides [28]. Instead, it is optimal to use isotopic labels that will be incorporated into most, if not all peptides in the sample. Studies involving cell culture can grow the cells in amino acids labeled with stable heavy isotopes ($^{2}H$, $^{15}N$, etc.)—these become incorporated into proteins and peptides if cells are treated for a sufficient length of time [5]. However, the cost of labeled amino acids is high for cell culture studies and prohibitively

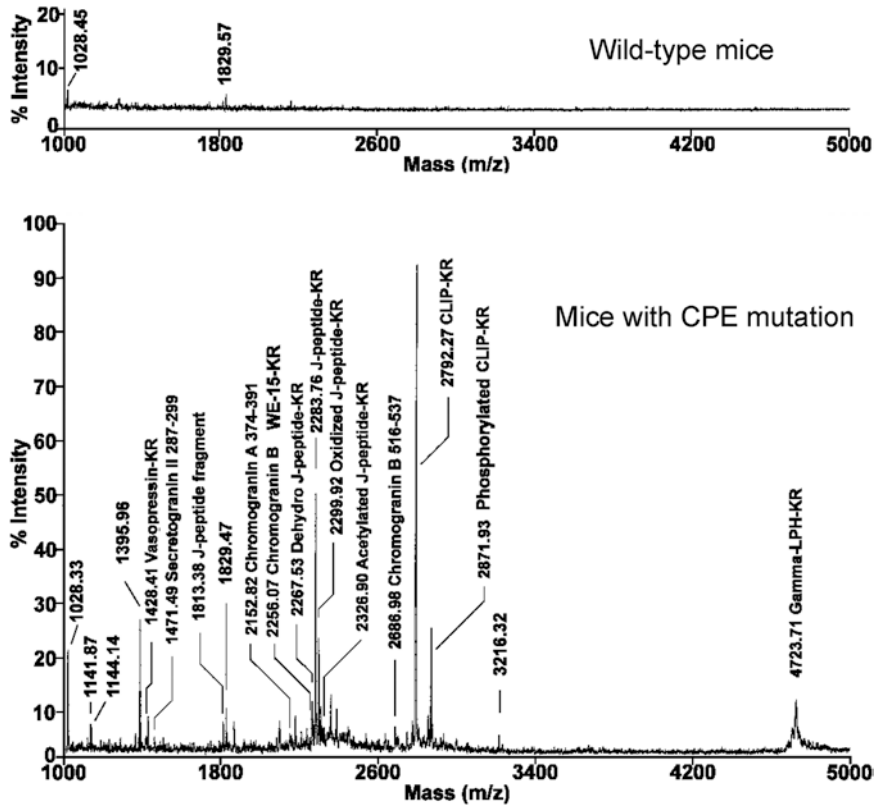



**Fig. 2** MALDI-TOF analysis of peptides extracted from 20 mouse pituitaries and purified on an anhydrotrypsin column. (Upper) Extract from wild-type mouse pituitary. (Lower) Extract from pituitaries of mice with mutation of carboxypeptidase E (CPE). The observed [MH]+ mass of selected ions is indicated along with peptides that were subsequently identified by using MS/MS analysis. Abbreviations: CLIP, corticotropin-like immunoreactive peptide; LPH, lipotropin. Figure modified from [66]

expensive for most animal studies, except for small organisms that can be grown in a laboratory such as *C. elegans* and *Drosophila* (*see* Kunz et al. chapter in this volume). Thus, most quantitative peptidomic studies perform post-extraction labeling of peptides with reagents that target free amines. Greater than 90% of peptides in previous studies of mouse brain have a free primary amine, either on the N-terminus or the side chain of Lys residues [29]. While some peptides will be missed by this approach, including some with important biological activities, reagents that target amines are preferable to those that target specific residues such as Cys. As a side point, reagents that target carboxylic acid groups could potentially be useful to complement the amine-based tags, but the chemistry is not as simple and no multiplex tags have been reported.

Another important property of the ideal isotopic tag is that peptides labeled with the light and heavy forms co-elute from the chromatography step, usually reversed-phase LC. If these forms do not precisely co-elute, the quantification is less accurate [22]. Some

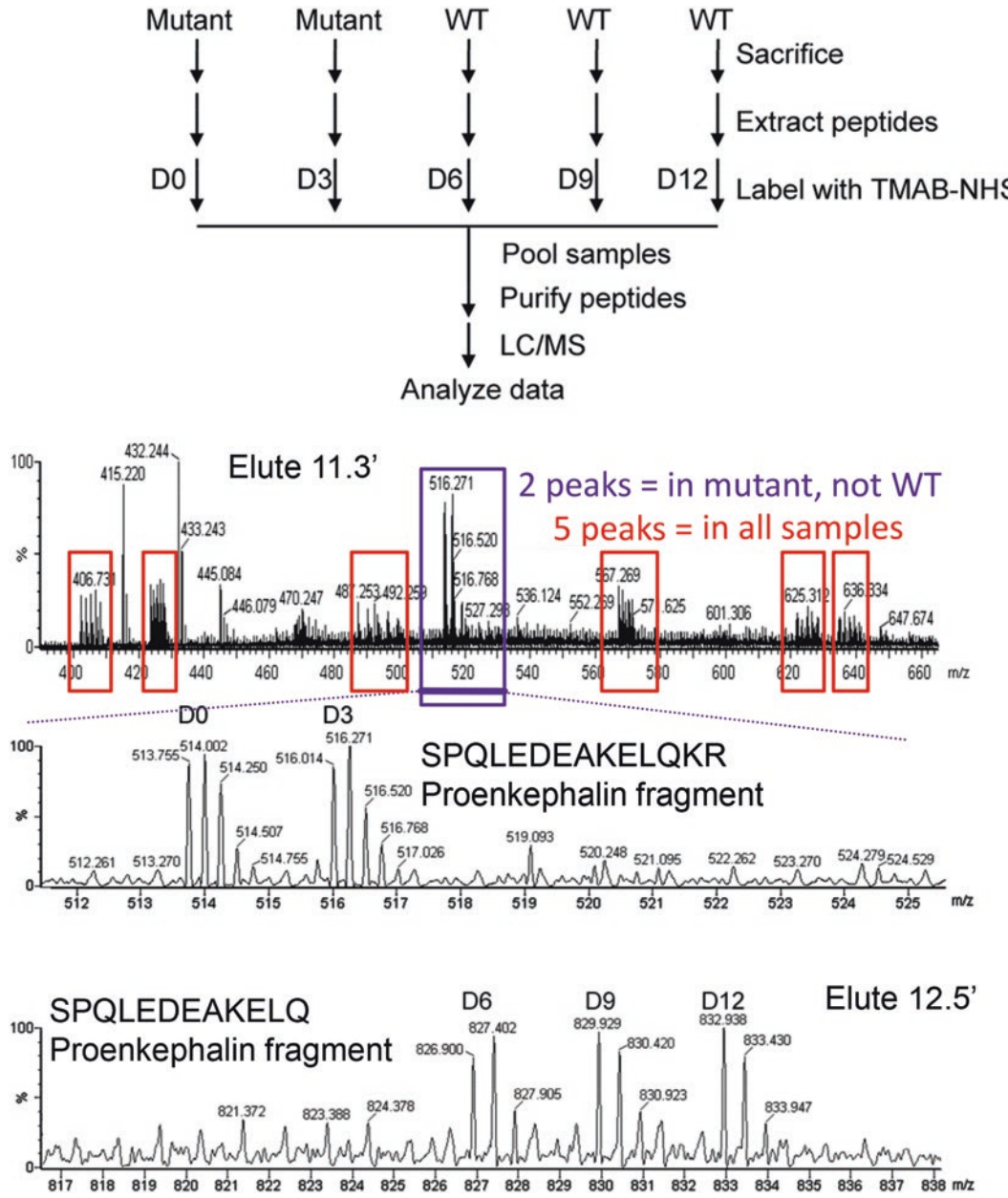



**Fig. 3** Representative data from an experiment with isotopic tags. As with Fig. 2, this experiment compared wild-type (WT) mice with *Cpe*$^{fat/fat}$ mice (mutant), in this case analyzing brain extracts without affinity chromatography and using TMAB isotopic tags to compare the five mice, as shown in the schematic diagram at the top of the figure. A portion of the MS spectrum of the peptides eluting at 11.3 min is shown (the full range analyzed spans from *m*/*z* 300 to 1800). This spectrum shows groups of five peaks with generally comparable peak heights (indicated by red boxes), groups of two peaks (indicated by purple boxes), and single peaks with *m*/*z* 415, 432, and 445. The single peaks presumably represent peptides that do not contain a free amine, but may also represent peptides with amines that did not get labeled during the reaction. The region from *m*/*z* 512 to 525 that eluted at 11.3 min is shown in the middle panel. From this analysis, there is no detectable signal for the D6, D9, or D12-labeled peptide, indicating that levels are >10-fold higher in mutant than in WT mouse brain. The peptide was subsequently identified by MS/MS analysis as a fragment of proenkephalin containing

of the isotopic tags that have been tested for peptidomics show a large isotope effect in the elution from reversed-phase LC columns. For example, peptides labeled with light and heavy forms of acetic anhydride (which acetylates primary and secondary amines) elute at different times, with the deuterated form eluting earlier than the hydrogenated form [22]. This is because hydrogen bonds contribute to the retention of peptides on the reversed-phase column, and deuterium bonds are weaker than hydrogen bonds. Also, acetylation converts primary and secondary amines to acetylated amides and these are not charged under the acidic conditions used for LC/MS. Thus, unless the peptide contains an Arg or His residue, the acetylated peptide will not be positively charged and cannot be detected on MS using positive ion mode (which is more sensitive and more commonly performed than negative ion mode). Another relatively inexpensive reagent, succinic anhydride, largely avoids the problem of different elution time from LC columns—peptides labeled with the deuterated and light forms of succinic anhydride generally co-elute from reversed-phase LC columns [22]. However, succinic anhydride converts amines into succinylated amides and the resulting peptides cannot be detected in positive ion mode unless they contain an Arg or His residue.

The isotopic tags most commonly used for quantitative proteomics and peptidomics avoid the problems with acetic and succinic anhydride, and also come in multiple isotopic forms which allows for inclusion of several replicates in the same LC/MS run. All of these tags react with amines and maintain the charge state of the peptide. The main isotopic tags used for quantitative peptidomics are described in this section, and the related isobaric tags are described in Subheading 1.3.

*1.2.1 Trimethylaminobutyrate (TMAB) Tags*

Together with colleagues, we have performed many peptidomic studies with TMAB tags [8–13, 22, 30–47]. The TMAB tags were originally developed in two isotopic forms by Regnier and colleagues [48] and extended to five isotopic forms by my colleagues [21]. The reagent used to label the peptides contains a carboxyl group activated with *N*-hydroxysuccinimide (NHS) [21, 48]. This reagent, referred to as TMAB-NHS, covalently labels primary and secondary amines. TMAB-NHS can be synthesized in five isotopic forms that differ by 3 Da, and for most peptides this difference is sufficient for accurate determination of peak intensity without peak overlap from other isotopic forms [21]. The difference in mass is readily detectable on the MS spectra (unlike the isobaric tags). Because the deuterium in the heavy forms is adjacent

**Fig. 3** (continued) C-terminal Lys-Arg, which is expected to be elevated in mice lacking carboxypeptidase E activity (*see* other Fricker chapter in this volume). The proenkephalin fragment lacking the C-terminal basic residues was identified from MS/MS analysis of a peptide that eluted at 12.5 min, which was only observed in the D6, D9, and D12-labeled forms (bottom panel). This indicates that the peptide lacking C-terminal basic residues is >10-fold higher in the WT mice than in the mutant mice lacking carboxypeptidase E activity

to a quaternary amine, this part of the molecule does not interact with the reversed-phase columns and as a result, peptides labeled with different isotopic forms of TMAB co-elute from LC columns [22].

While the TMAB reagents are excellent labels for many applications (described by Boonen et al. in another chapter in this book), there are problems with these labels. The reagents are not commercially available and need to be custom synthesized. Although the reagents are inexpensive and the reaction can be performed in a standard biochemistry laboratory with some specialized chemistry equipment (such as a rotary evaporator), it is difficult to achieve consistent purity of the labels [21]. Impurities can contribute to side reactions, such as iodination of His and Tyr residues in peptides [49]. If these impurities are present in any of the five reagents, it interferes with quantification of peptides that contain either His or Tyr residues. The labeled peptides can be detected using quadrupole time-of-flight mass spectrometers (Q-TOFs) but are unstable on some mass spectrometers, such as MALDI and ion traps (including the popular Orbitrap instruments). Another problem is that quantitation of peptides labeled with the five different TMAB reagents must currently be done manually. A computer program has been developed and is publicly available for quantification of peptides labeled with just two of the TMAB reagents: D0- and D9-TMAB. This is described in more detail in another chapter in this book (*see* Boonen et al.). However, the use of only two isotopic forms greatly increases the number of LC/MS runs needed for a typical experiment, and also requires separate analyses to establish a baseline variation of the peptides within the samples (Fig. 1).

### *1.2.2 Reductive Methylation with Formaldehyde*

Another approach for isotopic labeling of peptides is the reductive methylation of amines (described in more detail in Dasgupta et al. chapter in this volume). Both primary and secondary amines (e.g., N-terminal Pro residues) are labeled—the primary amines incorporate two methyl groups while the secondary amines incorporate a single methyl group. As a side point, this reaction is sometimes referred to as reductive dimethylation because nearly all peptides have free primary amines and therefore incorporate two methyl groups, but reductive methylation is technically the correct term. The labeling step is simple and uses reagents which are commercially available and inexpensive [50]. Two reagents are needed: formaldehyde and sodium cyanoborohydride, both of which are commercially available in heavy and light forms. With various combinations of these reagents, one can generate five isotopic forms which differ by 2 Da per primary amine that is labeled (*see* Dasgupta et al.). Peptides with multiple primary amines (i.e., a free N-terminus and one or more Lys residues) incorporate four or more methyl groups, resulting in mass differences between each isotopic form that are sufficient for accurate quantitation without problems due to peak overlap (*see* Dasgupta et al.). But for peptides with a single primary amine, there is considerable overlap in the spectra between

the isotopic forms, which complicates the analysis and requires calculation of the contribution from $^{13}C$–containing peptides and subtraction of this signal. This labeling approach and the calculations for subtracting overlapping signals are described in another chapter in this book (*see* Dasgupta et al.).

In our experience, peptides labeled with different isotopic forms of formaldehyde and borohydride co-elute from reversed-phase HPLC columns. This is presumably because the H and/or D atoms are present on methyl groups attached to a tertiary amine which is protonated at the acidic pH used for HPLC, and therefore these methyl groups are positively charged and do not interact with the reversed-phase matrix. However, it has been claimed that the deuterated peptides elute slightly ahead of the light ones when using a combination of basic pH reversed-phase chromatography followed by microcapillary LC-MS/MS on reversed-phase columns at acidic pH, although it was not specified whether the differential elution was observed at basic pH or acidic pH [51]. To minimize potential complications from differential elution of light and heavy forms, it is important to quantify the ratio of the forms over the entire peak and not use a single spectrum for quantification.

### 1.3 Isobaric Tags

Whereas the isotopic tags described above are identical chemical reagents except for the difference of heavy/light isotopic atoms, isobaric tags are chemical reagents with distinct structures for each form. Isobaric tags include a balancing group on each of the different tags so that each tag is the same mass until it is fragmented in the collision cell into products of different masses. Thus, the different isotopic tags are isobaric (i.e., same mass) prior to fragmentation. Because the balancer group contains the opposite numbers of heavy/light isotopic atoms as the reporter group, peptides labeled with different tags generally co-elute from reversed-phase LC columns. The differently labeled peptides show a single peak in the MS mode (unlike isotopically labeled peptides), simplifying the spectra with isobarically labeled peptides. However, quantification can only be achieved for those peptides which are selected for MS/MS analysis, and this is usually a subset of the total peptides that are detected in complex mixtures.

The first isobaric tags developed for proteomic studies, which also work for peptidomic analysis, were named iTRAQ (isobaric tag for relative and absolute quantitation) [52]. Initially these were available in four forms (4-plex), and more recently in eight forms (8-plex). Another set of isobaric tags, named TMT (tandem mass tags) was originally developed for 6-plex, and is currently available for 10-plex quantitation [53]. All of these commercial isobaric tag sets are expensive, compared to the isotopic tags mentioned in Subheading 1.2. Several recent peptidomic studies have used the TMT reagents for quantitation [27, 54].

Another set of isobaric tags for quantitative peptidomics is *N*,*N*-dimethyl leucine, abbreviated DiLeu that were developed by Li and colleagues [26]. These reagents can be easily synthesized in a standard chemistry laboratory, at a cost estimated to be 100-fold lower than the commercial isobaric tags (*see* DeLaney et al. chapter in this book). Li and colleagues have expanded the multiplexing capacity to 8-plex [55] and 12-plex DiLeu [56] enabling simultaneous quantification of 12 samples via the use of high resolution Orbitrap platforms and mass defect, which offers the highest throughput for isobaric tagging quantitation. In addition to isobaric quantitation enabling relative quantitation, DiLeu-based reagents can also be used for absolute quantitation via the synthesis of 5-plex isotopic DiLeu reagents (iDiLeu) with differing masses by incorporation of different deuterium atoms for each tag [57]. Both DiLeu and iDiLeu tags were also found to increase the fragmentation efficiency of neuropeptides, which offers unique advantages for de novo sequencing of neuropeptides and improved confidence for peptide identifications [58]. Quantification of peptides with these tags can be further improved using an ion mobility technique to reduce co-isolation and co-fragmentation of labeled peptides [59]. These tags have been used for the relative quantification of peptides and proteins in animal and human studies [58–60]. These tags have also been used for the absolute quantification of peptides, by including a standard curve of synthetic peptide labeled with some of the tags and pooling together with a biological sample labeled with another of the tags [24].

### 1.4 Summary

The various approaches for quantitative peptidomics all have strengths as well as weaknesses. The choice of technique to use depends largely on the goals of the experiment. Another consideration is the MS instrumentation available as well as the expertise of the collaborators who will assist with the analysis. To help in the decision process, **step 1** of Subheading 3 describes the important considerations for selection of the ideal method for different types of experiments.

## 2 Materials

Only general materials are indicated here because this is not a detailed protocol for a specific method, and is instead an overview of the different types of quantitative peptidomics approaches and materials needed for each. The focus of this chapter is to provide a guide for the selection of the technique, described in **step 1** of Subheading 3, based on what materials are available. An assumption is that the reader of this chapter is interested in peptidomics and will be working with a proteomics facility that will perform the MS and assist with the data analysis. But because of the differences between proteomics and peptidomics (described in Subheading 1), it is

important to use methods that are optimal for peptidomics; these depend on the goal of the studies (Subheading 3) as well as the materials that are available (briefly described below in this section).

### 2.1 Materials Required for Label-Free and Isotopic/Isobaric Label Approaches

1. Ultrapure water, such as distilled water that has been further purified through resins to remove ions and organic contaminants, the latter being the major problem for MS analyses.
2. Low retention microfuge tubes. Peptides tend to be sticky molecules that bind to plastic surfaces. The use of low retention tubes improves the recovery of peptides.
3. Biological samples containing peptides; minimum of three biological replicates of each group that is to be compared and ideally five or more.
4. Microfiltration units with 10 kDa molecular weight cutoff. These need to be washed with water prior to use in order to remove contaminants that would otherwise interfere with MS analyses.
5. C18 reversed-phase material to remove salts from peptides. Examples include spin columns, cartridges, or pipette tips (e.g., ZipTip) filled with C18 resin.
6. Vacuum centrifuge/concentrator (e.g., SpeedVac or equivalent).
7. HPLC system for fractionation of peptides, ideally coupled to ESI for direct interface with a mass spectrometer. Alternatively, some investigators use a fraction collector for the HPLC system that saves the eluate on plates that can be directly analyzed with a MALDI MS instrument (after addition of matrix).
8. Mass spectrometer capable of fragmentation of the peptide and analysis of the fragments (i.e., MS/MS, such as QTOF, ion trap, Orbitrap, TOF/TOF or comparable instrument).
9. Software for analyzing the spectra and converting raw data into searchable files. The choice of software will depend on the MS instruments used, and most proteomic facilities provide access to computers with the software installed.
10. Software for database searches to identify peptides from MS/MS data (*see* chapters in this volume by Azkargorta et al. and Southey et al.). Software is often provided by proteomic facilities, and many databases of commonly studied species are available on the internet.

### 2.2 Additional Materials for Label-Free Quantitation

1. Sufficient quantities of biological samples to allow for multiple technical replicates (ideally three) of each group, or a larger number of biological replicates if no technical replicates are performed.
2. Software capable of comparing multiple LC/MS runs, identifying peptides common to multiple runs, and quantifying levels

in each run either by spectral counting or by peak intensity. For review of software developed for proteomic studies, *see* [61].

### 2.3 Additional Materials for Quantitation with Isotopic Labels

1. Isotopic tags, with several options available, depending on goals of study.
   (a) If the study uses cell culture or small organisms that can be raised in a laboratory, the cells/organisms can be grown with regular media/food (often termed "light," *see* **Note 2**) or with media/food enriched in a heavy isotope. For cell culture media, amino acids containing leucine with ten atoms of deuterium were used for a SILAC study on neuropeptides in cultured cells [5]. The labeling of *Drosophila* with $^{15}N$ is described in another chapter in this volume (Kunz et al.).
   (b) For all other studies, chemical reagents are needed that covalently label peptides with light or heavy isotopes. Some are commercially available such as TMT, iTRAQ, and the reagents for methylation of amines using formaldehyde and borohydride/borodeuteride. Others require custom synthesis such as TMAB and DiLeu/iDiLeu (*see* chapters in this volume Boonen et al. and DeLaney et al.).
2. Reagents to quench the reaction prior to pooling (unless using SILAC or comparable approaches).
3. Software that can be used to quantify the data, either automatically (for some isotopic labels) or manually by inspection of the MS spectra for each peptide. Many proteomic facilities provide computers with the appropriate software installed.

## 3 Methods

For people new to the peptidomics field, the choice of which quantitative approach to use is difficult. Most proteomic facilities use a particular approach and will guide the new user toward those techniques, but they may not be optimal for peptidomics analyses. Instead, some of the better peptidomics methods are relatively easy to adapt to the equipment and techniques used in standard proteomic facilities. In this section, a series of questions is intended to guide the selection of the optimal quantitative peptidomics technique for different types of research projects.

1. What is the overall goal of the study?
   If the goal is to compare two or more experimental groups and identify the major differences in abundant peptides, then any of the approaches can be used. If the goal is to identify all differences in peptide levels between two or more experimental groups, then isotopic labels with manual analysis of the data

is optimal; isobaric tags and label-free approaches will miss many peptides that can be detected by manual analysis of MS data performed with isotopic labels.

2. What is the anticipated difference between levels of peptides in the two samples?

   If peptides are expected to show major changes, then any method can be used. For example, if the experiment is comparing extracts from a wild-type animal with extracts from an animal lacking a peptide-processing enzyme, then it is likely that many peptides will be greatly affected (*see* representative data in Figs. 2 and 3). If the peptides are expected to show small changes of only 20%, then it is better to use isotopic or isobaric labels to achieve accurate quantitation, and not the label-free approaches (unless using a triple quadrupole with SRM, which is highly accurate).

3. What is your knowledge of chemistry and access to standard organic chemistry laboratory equipment?

   If you can perform simple chemistry procedures in your laboratory or have access to equipment such as a rotary evaporator and a chemical fume hood, then you can consider the isotopic and isobaric tags that require custom synthesis (e.g., TMAB, DiLeu, and iDiLeu). These reagents are easy to synthesize and very inexpensive (relative to many commercial tags) but not all biology laboratories have the equipment necessary for their synthesis. An alternative is the formaldehyde/borohydride reagents for methylation of amines—the chemistry is performed during the labeling procedure and is as easy as adding commercial reagents to tubes containing biological samples. A chemical fume hood is required, but this is standard equipment in most biology laboratories. The commercial isotopic and isobaric tags require no knowledge of chemistry—these procedures just involve adding reagents to tubes containing the biological samples. The label-free approaches require no chemistry.

4. What is your budget, and do you have to pay for each LC/MS run?

   If you have unlimited funding, then the commercial isotopic or isobaric tags are a good option. If you do not have to pay for the LC/MS runs, then label-free is an option. But if you have limited funding, and especially if you have to pay a sizable amount for each LC/MS run, then consider the custom-synthesized tags or the formaldehyde/borohydride reagents. The label-free approaches save money by not requiring isotopic/isobaric tags, but the down-side is the need for many more LC/MS runs.

5. How many peptidomics studies do you plan to do?

   If a small number, then it is probably not worthwhile to custom synthesize the TMAB, DiLeu, or iDiLeu reagents,

which are typically produced in gram quantities that can be used for dozens of experiments, each involving multiple samples and several replicates. The cost of the reagents for the commercial tags is high but approximately the same as the cost of reagents to synthesize a batch of the 5-plex TMAB reagents, so if you're not going to do many experiments, it is not cost-effective to synthesize your own reagents. The formaldehyde/borohydride reagents are relatively inexpensive and the smallest amount that can be ordered is sufficient for hundreds of labeling reactions, so this is a reasonable alternative.

6. Do you want to see the primary data, or do you trust the computers to interpret the results?

   Samuel Clemens (a.k.a. Mark Twain) popularized the saying "There are three kinds of lies: lies, damned lies, and statistics." If you are the type who likes to see the actual data and not just a computer-derived table of data and statistical calculations, then isotopic tags with manual analysis of the MS data is your best option. The visual difference between peak intensities can be extremely convincing if sufficient replicates and low variability among these replicates. For example, the data in Fig. 3 show representative results from an experiment with isotopic tags that were manually analyzed, and it is clear that there is low variability among the replicates in each group of mice, but major differences in the levels of some (but not all) peptides between the two groups of mice. However, manual interpretation of MS data from peptides labeled with isotopic tags is time-consuming. Although manual interpretation provides the most coverage of the peptidome and the highest quality of data, many investigators want higher throughput approaches, even if the quality of the data is lower. In all cases, the results require statistical calculations to determine validity, but the difference with manual interpretation of isotopic tags and the automated approaches is that the former methods provide a sense of the actual results.

## 4 Notes

1. Quantitative peptidomics reveals relative levels of peptides between two or more samples, but the interpretation of the results is not always straightforward. For example, changes in peptide levels between wild-type mice and animals lacking specific peptide-processing enzymes have been used to determine which peptides are affected by the absence of a particular enzyme [8–10, 13, 31, 62]. Peptides that are substrates of the enzyme should be elevated in tissues from

knock-out mice, while products should be decreased in the absence of the enzyme. This is generally the case for mice lacking neuropeptide processing enzymes carboxypeptidase E, prohormone convertase 1, or prohormone convertase 2—in all three of these knock-out mice, peptides that were found to be increased or decreased in the knock-out mice matched peptides known from in vitro biochemical analysis to be substrates or products of these enzymes [8–10, 13, 31, 62]. However, changes in levels of peptides in knock-out mice do not always reflect a substrate/product relationship, and it is possible that the enzyme has an indirect effect on peptides. For example, mice lacking cytosolic carboxypeptidase 1 show elevated levels of many peptides and these were initially thought to be direct substrates of this enzyme, but subsequent biochemical studies found this to be incorrect—the enzyme has a very specific requirement for peptides with C-terminal acidic residues, and therefore the increased levels of many cellular peptides (most of which did not contain C-terminal acidic residues) is therefore an indirect effect of the missing enzyme [33, 41, 46, 63, 64].

2. The "light" isotopic form is more correctly termed "natural abundance" because it is not specifically enriched for the light isotope—it simply reflects the natural abundance of the isotope. For example, carbon is 98.9% $^{12}C$ and 1.1% $^{13}C$, nitrogen is 99.6% $^{14}N$ and 0.4% $^{15}N$, and oxygen is 99.8% $^{16}O$ and 0.2% $^{18}O$. Thus, cells/organisms grown in regular media/food, or peptides labeled with isotopic tags prepared from unenriched isotopic forms will be mostly the light forms of each of these isotopes.
3. It is essential to obtain MS/MS data to determine the identity of a peptide—it is not sufficient to claim identification based on mass alone. But once a peptide has been identified by MS/MS, it is not necessary to obtain MS/MS data in every LC/MS run of sample replicates. For example, once a particular neuropeptide has been identified by MS/MS analysis of mouse brain hypothalamic extracts, then in further experiments with extracts of mouse hypothalamus it is sufficient only to match the mass, charge, and elute time from the LC columns (assuming the column and gradient were similar between the experiments). And if the peptide precursor is also known to be expressed in mouse striatum (based on mRNA analysis), then it is reasonable to allow for this peptide to be "identified" in striatal extracts based only on mass, charge, and elute time. However, many programs for automatic quantitation require MS/MS identification each time, and when using these programs there will be many peptides missing from complex samples simply because the mass spectrometer did not select the

ion for fragmentation and subsequent MS/MS analysis (or the quality of the MS/MS results were insufficient for accurate identification). Manual inspection of the MS data can determine if the peptide was present in the sample (based on *m/z* of ion, charge, and elute time) if the MS/MS data was missing or insufficient for peptide identification.

## Acknowledgments

Thanks to Jonathan Sweedler, Lingjun Li, and Michael Schrader for helpful editing of this manuscript.

### References

1. Chard T (1987) In: Burdon RH, Van Knippenberg PH (ed) An introduction to radioimmunoassay and related techniques, Elsevier, Amsterdam, p 1–255
2. Baggerman G, Verleyen P, Clynen E et al (2004) Peptidomics. J Chromatogr B Analyt Technol Biomed Life Sci 803:3–16
3. Fricker LD, Lim J, Pan H et al (2006) Peptidomics: identification and quantification of endogenous peptides in neuroendocrine tissues. Mass Spectrom Rev 25:327–344
4. Schrader M, Schultz-Knappe P, Fricker LD (2014) Historical perspective of peptidomics. EuPA Open Proteom 3:171–182
5. Che FY, Yuan Q, Kalinina E et al (2004) Examination of the rate of peptide biosynthesis in neuroendocrine cell lines using a stable isotopic label and mass spectrometry. J Neurochem 90:585–594
6. Ong SE, Blagoev B, Kratchmarova I et al (2002) Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. Mol Cell Proteomics 1:376–386
7. Southey BR, Lee JE, Zamdborg L et al (2014) Comparing label-free quantitative peptidomics approaches to characterize diurnal variation of peptides in the rat suprachiasmatic nucleus. Anal Chem 86:443–452
8. Wardman JH, Zhang X, Gagnon S et al (2010) Analysis of peptides in prohormone convertase 1/3 null mouse brain using quantitative peptidomics. J Neurochem 114:215–225
9. Zhang X, Pan H, Peng B et al (2010) Neuropeptidomic analysis establishes a major role for prohormone convertase-2 in neuropeptide biosynthesis. J Neurochem 112:1168–1179
10. Zhang X, Che FY, Berezniuk I et al (2008) Peptidǫns for neuropeptide processing. J Neurochem 107:1596–1613
11. Che FY, Vathy I, Fricker LD (2006) Quantitative peptidomics in mice: effect of cocaine treatment. J Mol Neurosci 28:265–275
12. Decaillot FM, Che FY, Fricker LD et al (2006) Peptidomics of Cpefat/fat mouse hypothalamus and striatum: effect of chronic morphine administration. J Mol Neurosci 28:277–284
13. Lim J, Berezniuk I, Che FY et al (2006) Altered neuropeptide processing in prefrontal cortex of $Cpe^{fat/fat}$ mice: implications for neuropeptide discovery. J Neurochem 96:1169–1181
14. Bantscheff M, Lemeer S, Savitski MM et al (2012) Quantitative mass spectrometry in proteomics: critical review update from 2007 to the present. Anal Bioanal Chem 404:939–965
15. Ranc V, Petruzziello F, Kretz R et al (2012) Broad characterization of endogenous peptides in the tree shrew visual system. J Proteomics 75:2526–2535
16. Old WM, Meyer-Arendt K, Aveline-Wolf L et al (2005) Comparison of label-free methods for quantifying human proteins by shotgun proteomics. Mol Cell Proteomics 4:1487–1502
17. Bures EJ, Courchesne PL, Douglass J et al (2001) Identification of incompletely processed potential carboxypeptidase E substrates from CpEfat/CpEfat mice. Proteomics 1:79–92
18. Kusebauch U, Campbell DS, Deutsch EW et al (2016) Human SRMAtlas: a resource of targeted

assays to quantify the complete human proteome. Cell 166:766–778

19. Picotti P, Clement-Ziza M, Lam H et al (2013) A complete mass-spectrometric map of the yeast proteome applied to quantitative trait analysis. Nature 494:266–270
20. Fricker LD (2012) Neuropeptides and other bioactive peptides. In: Fricker LD, Devi L (eds) Colloquium series on neuropeptides. Morgan & Claypool Life Sciences, Charleston, SC, p 107
21. Morano C, Zhang X, Fricker LD (2008) Multiple isotopic labels for quantitative mass spectrometry. Anal Chem 80:9298–9309
22. Che FY, Fricker LD (2005) Quantitative peptidomics of mouse pituitary: comparison of different stable isotopic tags. J Mass Spectrom 40:238–249
23. Che FY, Fricker LD (2002) Quantitation of neuropeptides in Cpe fat/Cpe fat mice using differential isotopic tags and mass spectrometry. Anal Chem 74:3190–3198
24. Greer T, Li L (2016) Isotopic N,N-dimethyl leucine (iDiLeu) for absolute quantification of peptides using a standard curve approach. Methods Mol Biol 1410:195–206
25. Wang J, Zhang Y, Xiang F et al (2010) Combining capillary electrophoresis matrix-assisted laser desorption/ionization mass spectrometry and stable isotopic labeling techniques for comparative crustacean peptidomics. J Chromatogr A 1217:4463–4470
26. Xiang F, Ye H, Chen RB et al (2010) N,N-dimethyl leucines as novel isobaric tandem mass tags for quantitative proteomics and peptidomics. Anal Chem 82:2817–2825
27. Wu L, Li H, Li X et al (2017) Peptidomic analysis of cultured cardiomyocytes exposed to acute ischemic-hypoxia. Cell Physiol Biochem 41:358–368
28. Gygi SP, Rist B, Gerber SA et al (1999) Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. Nat Biotechnol 17:994–999
29. Fricker LD (2010) Analysis of mouse brain peptides using mass spectrometry-based peptidomics: implications for novel functions ranging from non-classical neuropeptides to microproteins. Mol Biosyst 6:1355–1365
30. Pan H, Che FY, Peng B et al (2006) The role of prohormone convertase-2 in hypothalamic neuropeptide processing: a quantitative neuropeptidomic study. J Neurochem 98:1763–1777
31. Che FY, Biswas R, Fricker LD (2005) Relative quantitation of peptides in wild type and Cpe$^{fat/fat}$ mouse pituitary using stable isotopic tags and mass spectrometry. J Mass Spectrom 40:227–237
32. Che FY, Yuan Q, Kalinina E et al (2005) Peptidomics of *Cpe$^{fat/fat}$* mouse hypothalamus: effect of food deprivation and exercise on peptide levels. J Biol Chem 280:4451–4461
33. Gomes I, Bobeck EN, Margolis EB et al (2016) Identification of GPR83 as the receptor for the neuroendocrine peptide PEN. Sci Signal 9:ra43
34. Lyons PJ, Fricker LD (2010) Substrate specificity of human carboxypeptidase A6. J Biol Chem 285:38234–38242
35. Tanco S, Zhang X, Morano C et al (2010) Human carboxypeptidase A4: characterization of the substrate specificity and implications for a role in extracellular peptide processing. J Biol Chem 285:18385–18396
36. Berti DA, Morano C, Russo LC et al (2009) Analysis of intracellular substrates and products of thimet oligopeptidase (EC 3.4.24.15) in human embryonic kidney 293 cells. J Biol Chem 284:14105–14116
37. Che FY, Zhang X, Berezniuk I et al (2007) Optimization of neuropeptide extraction from the mouse hypothalamus. J Proteome Res 6:4667–4676
38. Dasgupta S, Yang C, Castro LM et al (2016) Analysis of the yeast peptidome and comparison with the human peptidome. PLoS One 11:e0163312
39. Dasgupta S, Fishman MA, Mahallati H et al (2015) Reduced levels of proteasome products in a mouse striatal cell model of Huntington's disease. PLoS One 10:e0145333
40. Dasgupta S, Castro LM, Dulman R et al (2014) Proteasome inhibitors alter levels of intracellular peptides in HEK293T and SH-SY5Y cells. PLoS One 9:e103604
41. Berezniuk I, Sironi JJ, Wardman J et al (2013) Quantitative peptidomics of Purkinje cell degeneration mice. PLoS One 8:e60981
42. Gelman JS, Dasgupta S, Berezniuk I et al (2013) Analysis of peptides secreted from cultured mouse brain tissue. Biochim Biophys Acta 1834:2408–2417
43. Gelman JS, Sironi J, Berezniuk I et al (2013) Alterations of the intracellular peptidome in response to the proteasome inhibitor bortezomib. PLoS One 8:e53263
44. Fricker LD, Gelman JS, Castro LM et al (2012) Peptidomic analysis of HEK293T cells: effect of the proteasome inhibitor epoxomicin on intracellular peptides. J Proteome Res 11:1981–1990

45. Gelman JS, Sironi J, Castro LM et al (2011) Peptidomic analysis of human cell lines. J Proteome Res 10:1583–1592
46. Berezniuk I, Sironi J, Callaway MB et al (2010) CCP1/Nna1 functions in protein turnover in mouse brain: implications for cell death in Purkinje cell degeneration mice. FASEB J 24:1813–1823
47. Gelman JS, Sironi J, Castro LM et al (2010) Hemopressins and other hemoglobin-derived peptides in mouse brain: comparison between brain, blood, and heart peptidome and regulation in Cpefat/fat mice. J Neurochem 113:871–880
48. Zhang R, Sioma CS, Thompson RA et al (2002) Controlling deuterium isotope effects in comparative proteomics. Anal Chem 74:3662–3669
49. Fricker LD (2015) Limitations of mass spectrometry-based peptidomic approaches. J Am Soc Mass Spectrom 26:1981–1991
50. Boersema PJ, Raijmakers R, Lemeer S et al (2009) Multiplex peptide stable isotope dimethyl labeling for quantitative proteomics. Nat Protoc 4:484–494
51. Tolonen AC, Haas W (2014) Quantitative proteomics using reductive dimethylation for stable isotope labeling. J Vis Exp. https://doi.org/10.3791/51416
52. Ross PL, Huang YN, Marchese JN et al (2004) Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. Mol Cell Proteomics 3:1154–1169
53. Dayon L, Hainard A, Licker V et al (2008) Relative quantification of proteins in human cerebrospinal fluids by MS/MS using 6-plex isobaric tags. Anal Chem 80:2921–2931
54. Zhang J, Liang D, Cheng Q et al (2017) Peptidomic analysis of fetal heart tissue for identification of endogenous peptides involved in tetralogy of fallot. DNA Cell Biol 36(6):451–461
55. Frost DC, Greer T, Xiang F et al (2015) Development and characterization of novel 8-plex DiLeu isobaric labels for quantitative proteomics and peptidomics. Rapid Commun Mass Spectrom 29:1115–1124
56. Frost DC, Greer T, Li L (2015) High-resolution enabled 12-plex DiLeu isobaric tags for quantitative proteomics. Anal Chem 87:1646–1654
57. Greer T, Lietz CB, Xiang F et al (2015) Novel isotopic N,N-dimethyl leucine (iDiLeu) reagents enable absolute quantification of peptides and proteins using a standard curve approach. J Am Soc Mass Spectrom 26:107–119
58. Hui L, Xiang F, Zhang Y et al (2012) Mass spectrometric elucidation of the neuropeptidome of a crustacean neuroendocrine organ. Peptides 36:230–239
59. Sturm RM, Lietz CB, Li L (2014) Improved isobaric tandem mass tag quantification by ion mobility mass spectrometry. Rapid Commun Mass Spectrom 28:1051–1060
60. Greer T, Hao L, Nechyporenko A et al (2015) Custom 4-plex DiLeu Isobaric labels enable relative quantification of urinary proteins in men with lower urinary tract symptoms (LUTS). PLoS One 10:e0135415
61. Lemeer S, Hahne H, Pachl F et al (2012) Software tools for MS-based quantitative proteomics: a brief overview. Methods Mol Biol 893:489–499
62. Fricker LD (2007) Neuropeptidomics to study peptide processing in animal models of obesity. Endocrinology 148:4185–4190
63. Berezniuk I, Lyons PJ, Sironi JJ et al (2013) Cytosolic carboxypeptidase 5 removes alpha- and gamma-linked glutamates from tubulin. In: J Biol Chem. https://doi.org/10.1074/jbc.M113.497917
64. Berezniuk I, Vu HT, Lyons PJ et al (2012) Cytosolic carboxypeptidase 1 is involved in processing alpha- and beta-tubulin. J Biol Chem 287:6503–6517
65. Wardman J, Fricker LD (2011) Quantitative peptidomics of mice lacking peptide-processing enzymes. Methods Mol Biol 768:307–323
66. Che FY, Yan L, Li H et al (2001) Identification of peptides from brain and pituitary of Cpe fat/Cpe fat mice. Proc Natl Acad Sci U S A 98:9971–9976

# Chapter 9

# Quantitative Peptidomics with Isotopic and Isobaric Tags

**Kurt Boonen, Wouter De Haes, Joris Van Houtven, Rik Verdonck, Geert Baggerman, Dirk Valkenborg, and Liliane Schoofs**

## Abstract

In differential peptidomics, peptide profiles are compared between biological samples and the resulting expression levels are correlated to a phenotype of interest. This, in turn, allows us insight into how peptides may affect the phenotype of interest. In quantitative differential peptidomics, both label-based and label-free techniques are often employed. Label-based techniques have several advantages over label-free methods, primarily that labels allow for various samples to be pooled prior to liquid chromatography-mass spectrometry (LC-MS) analysis, reducing between-run variation. Here, we detail a method for performing quantitative peptidomics using stable amine-binding isotopic and isobaric tags.

**Key words** Differential peptidomics, Mass spectrometry, Label-based, 4-Trimethylammoniumbutyryl (TMAB), Tandem mass tags (TMT)

## 1 Introduction

In differential peptidomics, peptide profiles are compared between biological samples, often to test which (neuro-)peptides are potentially involved in a phenotype of interest. The uses for differential peptidomics are manifold and range from investigating peptide processing enzymes [1–3] to more functional studies trying to relate peptide levels to phenotypes such as foraging or feeding behavior [4–6]. In the honey bee *Apis mellifera* for example, differential peptidomics was employed to compare brain neuropeptide profiles between foragers and nurses and between nectar and pollen foragers, through which eight peptides were identified that may regulate honey bee behavior [4].

Both label-free and label-based techniques are often employed in quantitative peptidomics, and both have their own advantages and disadvantages. The advantages of labeling methods are that they are more robust to instrumental artifacts such as ion suppression effects,

Kurt Boonen and Wouter De Haes contributed equally to this work.

Michael Schrader and Lloyd Fricker (eds.), *Peptidomics: Methods and Strategies*, Methods in Molecular Biology, vol. 1719, https://doi.org/10.1007/978-1-4939-7537-2_9, 

as can happen in experiments that add different concentrations of (nonremovable) compounds. The reason for this robustness is that multiplexed samples are influenced by the same amount of instrument variability allowing for a more efficient comparison of the signal intensities. Furthermore, the lesser time needed for one single MS analysis, due to multiplexing, can allow a more in-depth LC-MS analysis by prefractionation techniques, leading to more peptide identifications (and quantifications). In this chapter, we focus on amine-binding label-based techniques, more specifically using isotopic 4-trimethylammoniumbutyryl (TMAB) [7] or isobaric tandem mass tags (TMT) [8]. Isotopic tags differ in mass and are used to differentiate peptides from different samples observed in the precursor scan of a mass spectrometry analysis (MS1) [7]. Isobaric tags, on the other hand, have the same weight, hence isobaric, but are conceived in such a way that they each release different labels, so-called reporter ions, with different mass upon fragmentation (MS/MS) during tandem mass spectrometric analysis (MS2). This still allows for differentiation or demultiplexing of the samples by fragmentation [8] (*see* Fig. 1). Isobaric tags were developed as a label for quantitative bottom-up proteomics, which uses mostly trypsin and/or LysC to digest proteins before labeling. This proteolytic cleavage creates peptides with basic amino acid extensions, mostly lysine (K) or arginine (R), and can therefore be labeled at the primary amines of the N-terminus and lysine. Peptidomics, being focused on endogenous peptides, does not use proteolytic enzymes and the amine-binding label approaches are therefore limited to endogenous peptides with at least one primary amine. This leaves out neuropeptides with an N-terminal pyroglutamate, and without lysines. These conditions can of course be checked before choosing the appropriate quantitative strategy. Alternatively, TMT kits for cysteine and carbonyl groups are available.

Using isotopic labels such as TMAB has the advantage that it is a less expensive approach that is compatible with most mass spectrometers. Isobaric tags, on the other hand, more easily allow for comparing a multitude of samples in the single experimental run (multiplexing), although the user is more restricted in his choice of mass spectrometer due to the need to properly fragment the peptide and its bound label. For a more comprehensive discussion on the advantages and disadvantages of different quantification strategies, see the publications by Romanova et al. [9] and Fricker [10].

It is clear that both types of labels have their own set of advantages. As such, here we provide detailed protocols for performing differential peptidomics using either TMAB or TMT. For the analysis of duplex TMAB data, we utilize the open-source tool "*labelpepmatch*," which allows for semi-automated mass matching of differentially labeled peptides, as this would otherwise be a mostly manual and laborious process [11]. The *labelpepmatch* tool also includes several modules for the visualization and quality control of your peptide samples (*see* Fig. 2), which increases the interpretability of your data.

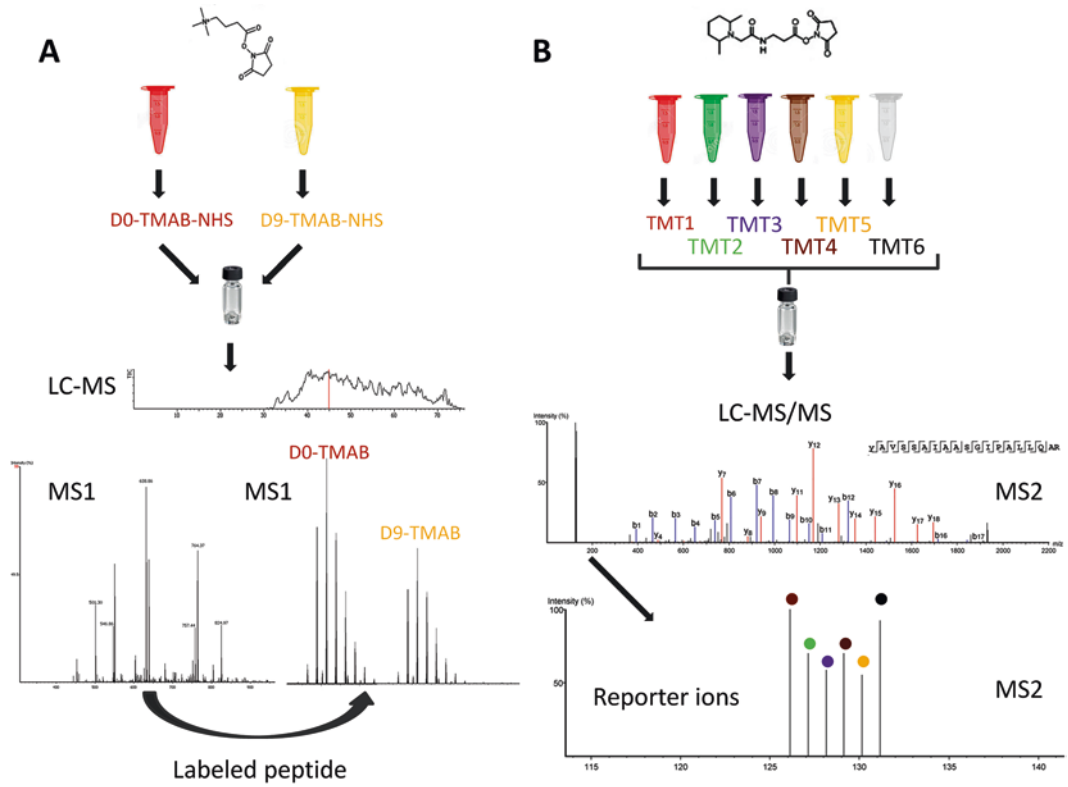



**Fig. 1** Overview of the TMAB (**a**) and TMT (**b**) labeling procedures. Peptides are labeled at their free primary amines by the isotopic TMAB or isobaric TMT labels. The different samples are pooled after labeling and analyzed by LC-MS. TMAB quantification is performed in MS1, by comparing the peptide intensities that have been labeled with a light and heavy TMAB label. Up to five different isotopic TMAB labels are available. The *labelpepmatch* tool currently uses duplex TMAB labels (usually D0 and D9). TMT quantification is performed in MS2. Reporter ions are released from the label upon fragmentation and can be used to compare the relative concentrations between multiple samples. TMT sixplex is depicted in this figure, but TMT 10-plex is also available and compatible with the CONSTAND++ workflow. TMT 10-plex analysis requires high resolution MS to distinguish between N and C isotopes

We also present a workflow for TMT-based quantitative peptidomics, based on the data-driven normalization of the data by the CONSTANd (*con*strained *stand*ardization) algorithm [12]. Data that is normalized by CONSTANd can be combined thereafter without the need of reference samples (*see* **Note 1** for experimental design and combination of experiments) and can be used prior to explorative analysis like clustering or more in-depth statistical studies (like linear mixed models). Most importantly, CONSTANd normalization allows for extracting meaningful biological data out of the combined data of multiple experiments, instead of just looking at experimentally induced differences (*see* Fig. 3). CONSTANd uses a table with (among other things) peptide/protein identifications, together with the reporter ion values as an input. The reporter ion values are normalized row wise

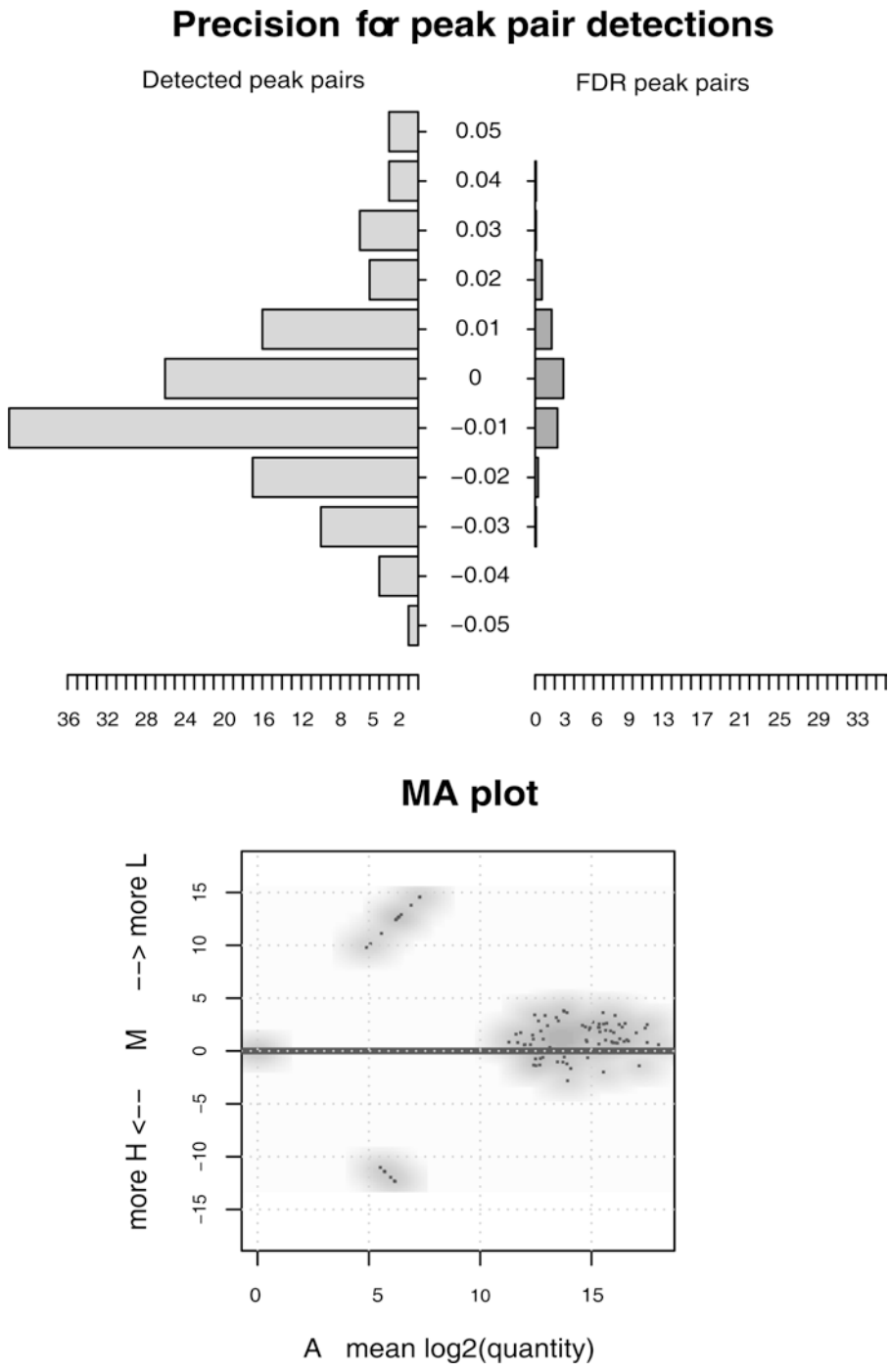



**Fig. 2** An example of the quality control graphs that are generated by *labelpepmatch* [11] for TMAB datasets. The top graph shows the distribution of the accuracy of each of the detected peak pairs (in Da), which should be a symmetrical distribution centered around an error of 0 Da. *Labelpepmatch* also contains a function to correct for false discovery rate (FDR): Semi-randomized data is also run through the labelpepmatch pipeline, which should result in much fewer detected peak pairs, as is the case here. The bottom graph shows a minus additive (MA) plot (also *see* Fig. 4) which reveals a bias toward higher intensity of light labeled peptides. A full run-through of the labelpepmatch pipeline and an explanation of each plot can be found in the vignette of the package (*see* **Note 19**)

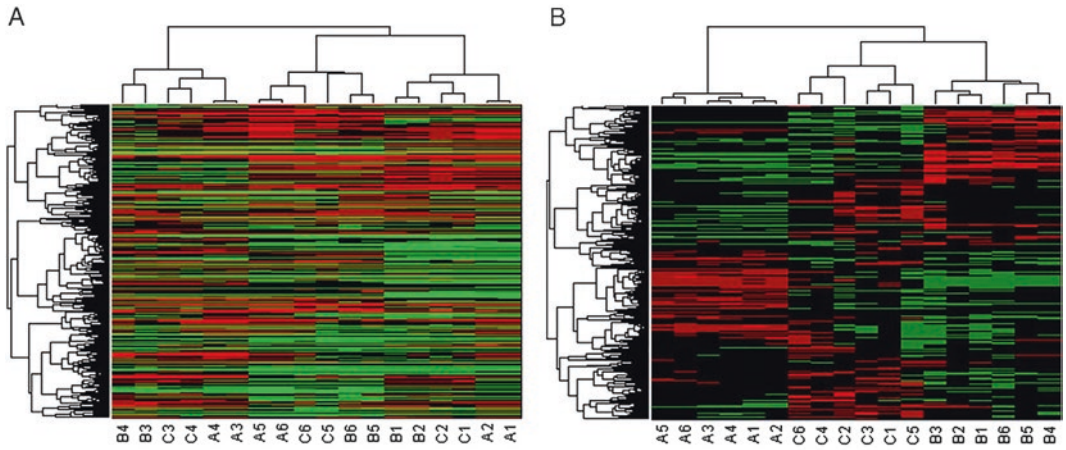



**Fig. 3** Hierarchical clustering of reporter ion intensities (dataset from [12]). 3 TMT six-plex experiments were combined (named PM, BM, and TAM) and each column represents one quantification channel. The rows represent the peptides. The color intensity indicates overexpression (red), underexpression (green), or equal amounts (black). Figure (**a**) represents all experiments after quantile normalization, (**b**) after CONSTANd workflow. Note that in figure (**a**) the conditions are more strongly correlated to the experimental conditions than to the biological conditions, while in (**b**), the biological conditions are grouped. This clearly indicates the power of the CONSTANd workflow when combining multiple experiments (*see* **Note 1**)

to sum up to 1 (transforming reporter ion intensities into "percentages"). The average value of one row of reporter ion intensities is therefore $1/n$ (for TMT6 n being 6) after scaling. A column (sample) based normalization is performed by making their sum equal to $m/n$, with m being the total number of peptides, or rows, in the tabular format. The constraint of $m/n$ represents an equimolar pool from the samples of interest. This means that after normalization, the average value of reporter ions in a row of the table is $1/n$. Sequential row and column normalization by itself would compromise each other, since the sum of the row normalization would not be equal to 1 anymore after column normalization. This problem can be addressed mathematically by the RAS procedure [12]. Most importantly, this procedure is fast and scales well with high-throughput omics data. The column normalization corrects for experimental errors like preparation of the peptide sample and labeling errors. The latter is much less of an issue since the labeling procedure is quite robust [13]. In the case of peptidomics, this means that peptide samples should roughly contain equal amounts of peptides, e.g., samples that contain much more peptides due to protein degradation (because of improper sample handling) should be discarded. It should also be noted that the eventual peptide population for normalization is smaller for peptidomics than proteomics and is therefore inherently more prone to inducing biases (*see* **Note 2** on normalization and assumptions of CONSTANd).

TMT labeling is a fast and robust approach for quantitative peptidomics and, when applied to endogenous peptides, suffers less from its major drawback: the co-isolation of peptides in MS/MS. Co-isolation of peptides tends to suppress ratio differences between conditions to unity since the reporter ions of the selected peptide are contaminated with reporter ions of the co-isolated or background peptide. Bias of peptide ratios can be solved by selecting MS/MS fragments of the labeled peptides for further fragmentation (MS3) and using the MS3 reporter ions [14]. This of course requires MS3 capabilities. Another method consists of using the complement reporter ion cluster; this is possible since the "leftovers" of the reporter ions itself contain quantitative information from the balancer group and since this cluster is specific for a single peptide [15]. Co-isolation would therefore result in more than one complement reporter ion clusters, together with the "contaminated" reporter ion cluster. The difficulty is recognizing the correct cluster since they do not have a predefined mass. The problem of co-isolation is less pertinent in peptidomic samples since these are usually less complex than shot-gun proteomics samples. Peptidomic samples can contain highly intense peptides and these do often have a tail in LC. One should be careful that such tailing is not interfering with the quantification of other peptides. This can again be checked by inspecting the raw data. The CONSTANd normalization tool is also optimized to handle proteome discoverer data (but not exclusively), and this software gives a value for co-isolation of peptides (see further).

## 2 Materials

All solutions are prepared using ultrapure water unless stated otherwise. All reactions take place at room temperature unless otherwise indicated.

### 2.1 TMAB Labeling

1. Dried peptide samples representative of two conditions you want to compare, preferably eight or more biologically independent samples per condition (*see* **Note 3**) (see other chapters for sampling and extracting peptides).
2. TMAB labels (*see* **Note 4**)—either D0 and D9, which have a mass difference of 9 Da, or all five forms that differ by 3 Da (*see* **Note 3**). Store at 4 °C and in a dry environment (*see* **Note 5**).
3. Water-free pure dimethylsulfoxide (DMSO) (*see* **Note 5**).
4. Resuspension solution: 5%$_{v/v}$ acetonitrile (ACN) and 0.1%$_{v/v}$ trifluoroacetic acid (TFA).

5. Buffering solution: 0.5 M disodium hydrogen phosphate ($Na_2HPO_4$) at pH 9.5.
6. Alkalizing solution: 1 M sodium hydroxide (NaOH).
7. Quenching solution: 2.5 M glycine. Make fresh before every experiment.
8. Hydroxylamine solution: 2 M hydroxylamine in water-free DMSO (*see* **Note 5**). Make fresh before every experiment.
9. A calibrated pH electrode capable of measuring pH in very small volumes (such as a Hamilton Biotrode).
10. Mobile Phase A: 2% ACN and 0.1% formic acid.
11. Mobile Phase B: 90% ACN and 0.1% formic acid.

### 2.2 TMT Labeling

1. Dried peptide samples.
2. TMT sixplex isobaric reagent set (Thermo Fisher) (*see* **Notes 6** and **7**).
3. 100 mM TEAB (triethyl ammonium bicarbonate) solution.
4. 5% hydroxylamine in 100 mM TEAB.
5. Anhydrous ACN.

### 2.3 Desalting

1. Hydrophilic lipophilic balance (HLB) solid phase extraction (SPE) cartridges (we use Oasis™ HLB, Waters).
2. $C_{18}$ concentrator pipette tips (we use ZipTip, Millipore).

### 2.4 LC-MS/MS Analysis and Software Packages

1. Nano liquid chromatography system (nano-LC system, *see* **Note 8**) coupled on-line to an electrospray MS/MS-capable mass spectrometer (*see* **Notes 9** and **10**).
2. CONSTANd normalization tool is available online at the website www.qcquan.net. This website also contains further information on how to use the tool for protein and peptide centered quantification approaches.
3. *labelpepmatch* [11].
4. Progenesis LC-MS (Nonlinear Dynamics) according to the instructions given by Verdonck et al. [11] (*see* **Note 11**).
5. MS/MS identification software (such as MASCOT, Matrix Science, London, or SEQUEST, Thermo Fisher).
6. De novo protein identification software such as PEAKS [16–18] or SPIDER [19].
7. R [20] (https://www.r-project.org/).

## 3 Methods

### 3.1 Differential Peptidomics Using Isotopic TMAB Labels

#### 3.1.1 Labeling Peptides Using TMAB

1. Dissolve D0 and D9 TMAB at a concentration of 1.5 M in water-free DMSO. Prepare this fresh for every day of labeling (*see* **Note 5**). Vortex briefly to properly dissolve the label. Normally, 150 μL of each label (300 μL in total) suffices for a single day of labeling (*see* **Note 12**).
2. Prepare a proper labeling design for each experiment (*see* Table 1 and **Note 12**). Always introduce a label swap, as shown in Table 1.
3. Resuspend peptide samples in 50 μL of resuspension solution (*see* **Note 13**).
4. Add 200 μL of buffering solution to each of the samples to ensure that the pH is optimal for the TMAB labeling reaction (*see* **Note 14**).
5. Start the labeling reaction by adding 4 μL of either D0 or D9 to its respective sample, according to the labeling design (Table 1). As the TMAB labeling reaction acidifies the sample, the pH is measured after 10–16 min using a pH electrode. The pH is immediately adjusted to 9–9.5 using 0.5–2 μL of the alkalizing solution. Briefly vortex and spin down after every step.
6. The previous step is repeated an additional six times in 15–16 min intervals (Table 1) to ensure complete labeling.
7. After the last pH adjustment, incubate the samples for 2 h at room temperature.
8. Afterwards, add 30 μL of the quenching solution to each sample to stop the labeling reaction. Immediately readjust the pH to 9–9.5 using alkalizing solution. Following this, incubate the samples for 40 minutes at room temperature and then readjust the pH to 9–9.5 again after. Briefly vortex and spin down after every step.
9. Add 2 μL of the hydroxylamine solution to each sample to remove labels from tyrosine residues. If necessary, immediately readjust the pH to 9–9.5 using alkalizing solution. Vortex and spin down every time you add a solution.
10. The previous step is repeated an additional two times in 10 min intervals (*see* **Note 15**).
11. Two differentially labeled peptides can be pooled into a single reaction tube (*see* **Note 15**). Always carefully note which two samples you pooled in each reaction tube.
12. Desalt the pooled samples using HLB solid phase extraction cartridges according to the manufacturer's instructions (*see* **Note 16**).
13. Vacuum-dry the desalted pooled samples (*see* **Note 16**).
14. Store the dried labeled peptides at −20 °C prior to MS analysis.

**Table 1**
**Design for a single TMAB labeling experiment**

| | Label | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | TMAB D0 (light) | | | | TMAB D9 (heavy) | | | |
| Sample | Control 1 | Experimental 2 | Control 3 | Experimental 4 | Experimental 1 | Control 2 | Experimental 3 | Control 4 |
| TMAB | 0 | 2 | 4 | 6 | 8 | 10 | 12 | 14 |
| pH | 16 | 18 | 20 | 22 | 24 | 26 | 28 | 30 |
| TMAB | 32 | 34 | 36 | 38 | 40 | 42 | 44 | 46 |
| pH | 48 | 50 | 52 | 54 | 56 | 58 | 60 | 62 |
| TMAB | 64 | 66 | 68 | 70 | 72 | 74 | 76 | 78 |
| pH | 80 | 82 | 84 | 86 | 88 | 90 | 92 | 94 |
| TMAB | 96 | 98 | 100 | 102 | 104 | 106 | 108 | 110 |
| pH | 112 | 114 | 116 | 118 | 120 | 122 | 124 | 126 |
| TMAB | 128 | 130 | 132 | 134 | 136 | 138 | 140 | 142 |
| pH | 144 | 146 | 148 | 150 | 152 | 154 | 156 | 158 |
| TMAB | 160 | 162 | 164 | 166 | 168 | 170 | 172 | 174 |
| pH | 176 | 178 | 180 | 182 | 184 | 186 | 188 | 190 |
| TMAB | 192 | 194 | 196 | 198 | 200 | 202 | 204 | 206 |
| pH | 208 | 210 | 212 | 214 | 216 | 218 | 220 | 222 |

Numbers in the table indicate time since start of labeling. As an example, here we show the labeling of four control and four experimental samples with a label swap. We recommend labeling at most eight samples simultaneously due to time constraints. It is preferable to adjust the pH of each sample after at most 16 min, and we recommend re-adding TMAB every 15–16 min thereafter. To leave sufficient time for re-adjusting pH, we leave a 2 min gap between each different reaction, thus limiting a single labeling experiment to eight samples. Following labeling and quenching, control and experimental samples are pooled to create four pools for subsequent LC-MS analysis (e.g., control 1 + experimental 1; control 2 + experimental 2). Because there are five TMAB tags that can be synthesized and used for this analysis, this table can be modified accordingly so that multiple controls and experimental samples can be labeled at the same time, pooled, and analyzed in a single LC-MS run.

*3.1.2 Labeling Peptides Using TMT*

1. Redissolve the dried peptide samples in 100 mM TEAB buffer.
2. Immediately before use, equilibrate the TMT Label Reagents to room temperature. Add 41 μL of anhydrous acetonitrile to each 0.8 mg vial. Allow the reagent to dissolve for 5 min with occasional vortexing.
3. Carefully add 41 μL of the TMT Label Reagent to each sample.
4. Incubate the reaction for 1 h at room temperature.

5. Add 8 μL of 5% hydroxylamine to the sample and incubate for 15 min to quench the reaction.
6. Combine samples at equal amounts and desalt the pooled samples using HLB solid phase extraction cartridges according to the manufacturer's instructions.
7. Vacuum-dry the desalted pooled samples (*see* **Note 16**).

### 3.1.3 LC/MS(/MS) Analysis of TMAB and TMT-Labeled Peptides

1. Dissolve the sample in a 5% ACN 0.1% formic acid solution.
2. Separate the samples on a nano-LC system (*see* **Note 8**) coupled on-line to an electrospray MS/MS-capable mass spectrometer (*see* **Notes 9** and **10**).
3. For TMAB-labeled peptides, export the MS1-mode results (*see* **Note 17**) for further analysis in Progenesis LC-MS (Nonlinear Dynamics) according to the instructions given by Verdonck et al. [11] (*see* **Note 11**).
4. TMT-labeled peptides require LC-MS/MS analysis, making sure that the low mass reporter ions are also measured (*see* **Note 10**).

### 3.1.4 Analysis of TMAB Data Using labelpepmatch

1. Install R [20] (https://www.r-project.org/) on any computer.
2. Open R and install the devtools package using the following command:
   *install.packages("devtools")*.
3. In R, install *labelpepmatch* [11] using the following command (*see* **Note 18**):
   *devtools::install_github("goat-anti-rabbit/labelpepmatch.R", build_vignettes = F)*.
4. Next, load *labelpepmatch* using the following command:
   *library(labelpepmatch)*.
5. Use the *read.progenesis()* function (*see* **Note 19**) to open the Progenesis LC-MS file containing the results of your MS runs.
6. Use the *pepmatch()* function (*see* **Note 19**) to mass match peptides with a mass difference of 9 Da (or a multiple of 9 Da) due to differential labeling.
7. Use the *pep.id()* function (*see* **Note 19**) to match peptides to one of several peptide database (*see* **Note 20**).
8. Finally, use the *lpm_linearmodel()* function (*see* **Note 19**) to look for differential peptides with statistically different abundances in different conditions.

### 3.1.5 Analysis Using CONSTANd

1. *Creation of the data files containing the peptide IDs and reporter ion intensities.* Information on the workflow and data file format is available at the website www.qcquan.net. The data is to be delivered (per experiment) as a .xlsx, .csv, or .tsv file into

the web interface, with the variable names (*see* **Note 21**) as column headers. The order is unimportant, and there can be more variables present.

2. Submit the naming scheme for your collection of experiments as a .xlsx, .csv, or .tsv file. Each label name can be substituted differently for each experiment data file, so that is easier to distinguish each experiment subject in the results.
3. Submit a data file for each experiment (*see* **Note 21**).
4. Specify some basic and/or advanced options which control the workflow. Start the job; the CONSTANd++ server now takes control:
   (a) All nonrequired variables and observations with missing values are deleted (missing reporter ions are allowed).
   (b) Only confidently identified peptides are retained.
   (c) All data with Isolation Interference [%] strictly lower than 30 are retained.
   (d) Prioritize data identified using a specific #Identifying_Node (PSM matching algorithm: e.g., MASCOT or SEQUEST).
   (e) Remove "duplicate" peptide observations with different retention times (or alternatively scan numbers) but identical Sequence, Identifying Node, Charge, and Modifications. The one with the most confident identification is selected as a representative. For each observation, a Degeneracy variable (amount of duplicates) is added.
   (f) Remove "'duplicate" observations with different Charge but identical Identifying Node and Modifications.
   (g) Isotopic Correction: for observations with no missing values in the MS2 intensities, the latter values are corrected for isotopic impurities, given the isotopic impurities matrix [21].
   (h) CONSTANd normalization [12].
   (i) Two differential expression analyses are performed using well-established statistical techniques. These will include volcano plots of the peptides/proteins, lists of differential peptides/proteins with fold changes and *p*-values, and PCA clustering plots.

When the job is finished, you can download a report with the differential expression analysis, output data as well as data removed from the workflow, and some additional workflow metadata such as warnings. By default, the report will be protein centered (peptides of same precursors grouped), but a peptide-centered analysis is also possible (*see* **Note 22**).

## 4 Notes

1. Peptidomics is peptide-centered and the goal of differential peptidomics is to compare peptide ratios instead of protein ("precursor") ratios. Quantification values of separate peptides from the same precursor are therefore usually not combined with each other to give protein level information (as biological "replicas" from the same protein), resulting in less statistical power. It is therefore highly recommended to combine several TMT experiments to increase statistical power. CONSTANd allows combining experiments without the need of reference samples. However, the experimental setup should be carefully designed and, importantly, a rigorous SOP for peptidomics must be followed to avoid biases and wrong conclusions. Conditions should be balanced over the TMT experiments (randomized block design). CONSTANd++'s output for each peptide is a percentage (percentages compared between groups in one experiment). This percentage depends on the design of the experiment, the number of replicas in each condition, and, of course, the ratio between conditions. Practically, experiments with the same layout (3 vs. 3 or 2 vs. 2 vs. 2 and same biological conditions compared) can be compared directly. Imbalanced numbers of biological replicas should be avoided since this leads to a bias in overexpressed peptides in the group with the most replicas (due to data-dependent acquisition) and to differences in percentages. Additional conditions (3 control vs. 3 condition 1 in combination with 3 controls vs. 3 condition 2) or skewed numbers of replicas (e.g., 3 vs. 3 and 2 vs. 4) can only be combined if the percentage of the controls is equalized by rescaling the percentages (this will of course lead to the sum of the percentages not being equal to 1, since they are now percentages that refer to other experiments).
2. Normalization with CONSTANd is justified when three assumptions are fulfilled. It requires first of all a reference set of peptides (or proteins) that does not vary between the samples. This condition is likely fulfilled for peptidomics when experimental manipulations do not affect the basic machinery of neuropeptide processing and secretion (e.g., prohormone convertase KOs). Second, the number of up- and downregulated peptides should roughly be equal and small compared to the not significantly altered peptides. Lastly, the systematic bias should be constant on a logarithmic scale, meaning that one normalization factor is sufficient to remove this bias (nonlinear ratio distortion at higher or lower intensities should not occur). This is explained in the MA plots in Fig. 4.
3. Here, we will solely focus on the comparison of two conditions using the most commonly used pair of TMAB labels, D0 and

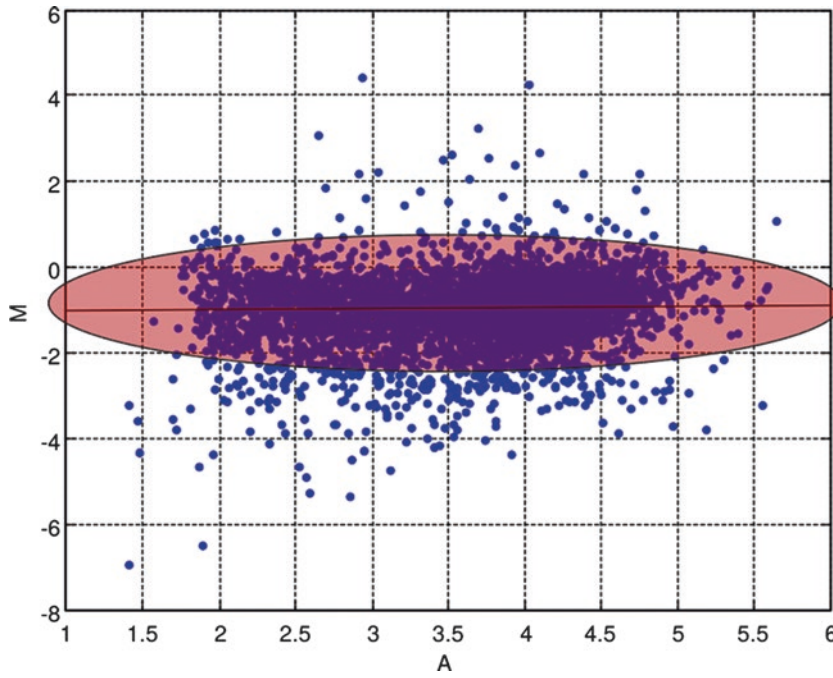



**Fig. 4** Minus Additive (MA) plots can be used to check if the assumptions for the correct application of CONSTANd are met. This is done for each combination of quantification channels. The A axis represents the average of the log intensities of the reporter ions in botch channels whereas the M axis represents the log2 ratio between the channels. Each point represents one peptide. MA plots can indicate whether all assumptions that justify using CONSTANd are met. All points should be centered around 0 if most peptides do not vary between samples. Second, there should be a roughly equal amount of points above and below 0 if there are roughly as many peptides up- or downregulated. Lastly, the average of the points on the A scale should be a straight line (no curly edges) parallel to the A-axis and the points should be contained by a rectangular box, indicating that the ratios do not correlate with low or high intensity

D9, named as such for the amount of deuteriums incorporated in either label. In total, five isoforms are available, thus also allowing multiplexing analysis [22]. The *labelpepmatch* tool is currently only compatible with duplex labeling.

4. One of the downsides to using TMAB labels is that they are not commercially available. However, their synthesis is quite basic and any organic chemistry lab should be suitably equipped to make them with relatively inexpensive reagents. For a full protocol, see the publication by Morano et al. [22].
5. Amine-reactive TMAB labels are reactive in water. This is mainly due to the fact that amine-reactive TMAB is actually a TMAB-N-hydroxysuccinimide (TMAB-NHS) ester, and the NHS moiety is sensitive to hydrolysis. As such, effort should be put in storing TMAB labels in a dry environment. Similarly, take care not to use old bottles of water-free DMSO. DMSO is hygroscopic and will accumulate water over time. Using old water-free DMSO may significantly reduce the labeling efficiency.

6. TMT six-plex is usually used, but ten-plex is also possible if the resolution of the mass spectrometer is sufficiently high. TMT 18-plex is theoretically possible.
7. It is sufficient for peptidomics to order only the labels since the isobaric mass tagging kits contain reagents only proteomic experiments use. The 0.8 mg vials are normally more than sufficient since the total protein concentration of peptidomic samples is low compared to proteomic samples.
8. For TMAB labels, we generally use a nano-LC Dionex UltiMate 3000 Dual LC System (75 μm × 15 cm Dionex Nano Series C18 column) with a 250 nL/min flow rate. For most experiments, we start the gradient at 95% mobile phase A and 5% mobile phase B and shift to 50% mobile phase B over a 45 min period. Depending on your needs and the complexity of your sample, you may need to adjust your runs.
9. We often use a microTOF-Q ESI-Q-TOF (Bruker Daltonics) for our TMAB experiments. Many other mass spectrometers work equally well (e.g., [2, 10, 22]). Do be aware that there are some limitations using MALDI-based mass spectrometers and mass spectrometers operating with an ion-trap when employing TMAB tags (*see* [10]).
10. TMT-labeled peptides can be readily analyzed by most LC-MS systems that are used for proteomics or peptidomics. However, they should be able to detect low masses since the reporter ions have masses ranging from 126 to 131 Da for TMT six-plex, ruling out some ion trap instruments. When using, e.g., a QExactive orbitrap mass spectrometer, the MS/MS fragmentation spectra should start from a fixed mass (around 100 Da). It may be necessary to optimize HCD fragmentation conditions to get both a good signal for the reporter ions and peptide fragmentation. For the QExactive, a normalized collision energy of 40% is ideal [23].
11. For quantitative analysis of TMAB-based differential peptidomics runs, data can also be analyzed in other relevant software packages (*see* refs. 24, 25). These approaches are often more laborious and require more manual input than the *labelpepmatch* approach. As such, here we specifically chose Progenesis as it links well to *labelpepmatch*.
12. Due to the strict timing requirements for labeling samples with TMAB (Table 1) it is recommended for one person to only label 8 samples simultaneously. The labeling reactions for other samples can be started after labeling of the first batch of samples has finished; or can be done by another person simultaneously.
13. It is recommended to vortex and sonicate (for 5 min in a sonication bath) your resuspended peptide samples to ensure that all peptides dissolve.

14. It is recommended to vortex and sonicate your peptide samples again after adding the buffering agent to ensure that all peptides remain in solution.
15. After this step, the samples can be flash-frozen and stored at −80 °C if your planning doesn't allow you to continue. As long as your samples remain in solution, it is recommended not to freeze them more than once.
16. Peptide samples are normally desalted during the extraction protocol. However, the TMAB labeling protocol adds large amounts of salts to the sample that may interfere with MS analysis. Therefore, we recommend desalting the TMAB-labeled samples using HLB solid phase extraction cartridges. For sensitive LC/MS systems, it can even be recommended to desalt the HLB-desalted dried peptides a second time using $C_{18}$ concentrator pipette tips according to the manufacturer's instructions.
17. We often work primarily with MS1 data when analyzing TMAB-labeled peptide spectra, because they couple well to *labelpepmatch* [11]. However, it is preferable to also analyze MS/MS data of the TMAB runs. This is especially true if little is known about the peptidome of your organism of choice, as it allows for more confident identification of peptides. Several papers show clear protocols on how to analyze MS/MS data of TMAB runs, including the book chapters by Gelman et al. [24] and Wardman and Fricker [26]. Be aware that the TMAB-label also partially fragments when using collision-induced dissociation to fragment peptides. Higher collision energies are recommended to consistently fragment the labeled peptides [24]. When using the Mascot search algorithm to identify peptides labeled with TMAB, they are listed as "GIST" rather than TMAB. Always include modifications on both N-termini and Lysine side chains. When selecting additional modifications to search for, always select oxidation of methionine as a variable modification, and bear in mind that cyanylation of cysteine and iodination of tyrosine and histidine have also recently been reported in TMAB-labeled samples [10].
18. *labelpepmatch* has several dependencies that you may need to install manually using the *install.packages()* function: lme4, multcomp, limma, bitops, brew, doParallel, foreach, influence.ME, lsmeans, plotrix, plyr, RCurl, reshape2, colorRamps, gplots, knitr, and rmarkdown.
19. A more detailed overview of the *labelpepmatch* workflow can be found in the vignette of the package. The vignette takes you through an entire *labelpepmatch* experiment using an included dataset. See the vignette at:
    https://github.com/goat-anti-rabbit/labelpepmatch.R/blob/master/vignettes/vignette.Rmd
    or
    https://perswww.kuleuven.be/~u0065551/vignette.html

20. Two peptide databases are included in *labelpepmatch* (*Schistocerca gregaria* and *Caenorhabditis elegans*) but others can be added to the package by the user.
21. The CONSTANd++ workflow is adapted to the Proteome Discoverer (PD) output and therefore follows their variable nomenclature. The following variables should be present in the table: Confidence, Identifying Node (PSM algorithm that performed this Annotated Sequence identification), **Annotated Sequence**, Modifications, # Protein Groups (number of different protein groups this peptide could be involved in), Master Protein Accessions (Uniprot identifier of the best matching protein for each protein group according to Identifying Node), **Protein Accessions**, Protein Descriptions, *m/z* [Da], Charge, Delta *m/z* [Da], Isolation Interference [%] (measure for the relative amount of co-isolation), RT [min], **First Scan** (identifier for the mass spectrometer scan in which the peptide was detected), XCorr (SEQUEST PSM score), Ions Score (Mascot PSM score), and the **names of the quantification channels**. Columns with their names in bold should be filled in (Identifying node and Xcorr can be set by default to "Sequest HT (A2)" and 0, if data are imported from other sources). There are several ways to get data files with reporter ion intensities and identifications: (A) Proteome Discoverer: the output of PD 2.0 (or higher) can be directly submitted to the CONSTANd server since it is primarily written to accept PD data. The raw reporter intensities can be obtained when the "Reporter Ions Quantifier" is incorporated in the workflow of interest (protein grouping must be disabled and visible in the peptide tab). The identification is usually done by Mascot and/or Sequest, although other modules can be implemented for peptidomics, like spectral library matching. (B) Mascot (server or in-house versions) is able to extract quantitative information from the reporter ions, although it only presents ratios compared to the intensity of a selected quantification channel. To convert these, simply add a column containing intensities for the quantification channel to which others were compared. Mascot results can be exported from the server: First choose "export search results" as format, then select the csv format and indicate that you want to include "peptide quantitation." (C) PEAKS Studio contains both de novo and database approaches, and additional modules for PTMs and mutations. The database, PTM, and mutation implement de novo information, resulting in a much-reduced computational time when allowing multiple modifications and no cleavage enzyme defined [17–19]. The Quantitation module is suited for TMT data and the results can be exported to csv files. Raw intensities can be exported instead of ratios. (D) Other options are using PeptideShaker in combination with Reporter [27] (http://compomics.github.io) or the highly flexible (but more demanding) OpenMS [28].

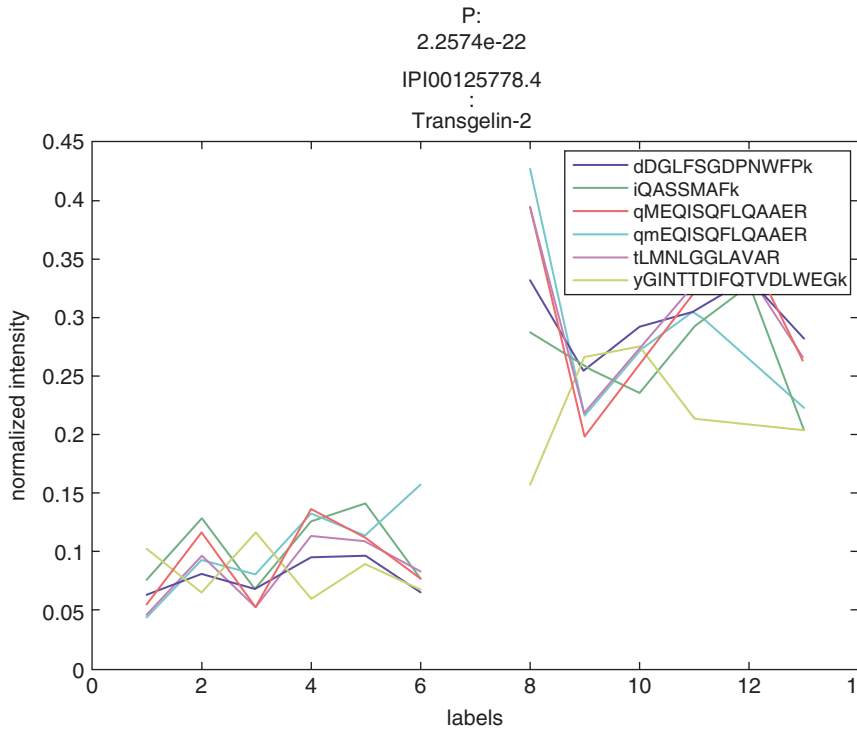



**Fig. 5** Spaghetti plot of all peptides from the Transgelin-2 protein in two conditions. This analysis is a combination of two TMT experiments. Not that all peptides from the same protein are combined in one plot. This means that separate peptides are considered as a biological replicate for the analysis of protein concentrations. The combination of all the values increases the number of data points and results in highly powered statistical test. Here, two conditions are compared by a *t*-test. This is meaningful in the case of neuropeptides if the whole precursor is upregulated. If only certain peptides have altered concentrations (due to PTMs or altered processing), the protein view obscures biologically relevant information on the peptide. However, this will become visible in the spaghetti plots of the percentages. Of course, protein inference can also be disabled and spaghetti plots (and *t*-tests or other) will only be performed on intensity percentages of a single peptide

22. Most proteomics software links the peptide spectrum matches to a protein sequence in the database (called protein inference). Protein inference in peptidomics is relevant when a precursor as a whole is up- or downregulated. Protein inference is not relevant if peptides are altered by post-translational mechanisms such as processing enzymes or have other modifications in certain conditions. Here the protein level information is obscuring the biologically relevant results. CONSTANd ++ gives a spaghetti plot of the peptide percentages (*see* Fig. 5). Deviant peptide trends can be easily observed in these plots. Alternative methods are using a peptide database for identifications (instead of the full-length proteins) or manually indexing the protein accession numbers (by adding a postfix (such as "_1", "_2", etc.) for the several peptides of one precursor). The other problem in protein inference, nonunique peptides (peptides that can be assigned to several proteins), is generally not an issue in peptidomics.

## Acknowledgments

W.D.H. is a research fellow of the FWO-Flanders. The authors wish to thank the FWO (G069713 and G095915N) and KU Leuven internal funds (C14/15/049) and the European Research Council (ERC grant 586 340318) for financial support.

### References

1. Fricker LD (2007) Neuropeptidomics to study peptide processing in animal models of obesity. Endocrinology 148:4185–4190. https://doi.org/10.1210/en.2007-0123
2. Wardman JH, Zhang X, Gagnon S et al (2010) Analysis of peptides in prohormone convertase 1/3 null mouse brain using quantitative peptidomics. J Neurochem 114:215–225. https://doi.org/10.1111/j.1471-4159.2010.06760.x
3. Miller LK, Hou X, Rodriguiz RM et al (2011) Mice deficient in endothelin-converting enzyme-2 exhibit abnormal responses to morphine and altered peptide levels in the spinal cord. J Neurochem 119:1074–1085. https://doi.org/10.1111/j.1471-4159.2011.07513.x
4. Brockmann A, Annangudi SP, T a R et al (2009) Quantitative peptidomics reveal brain peptide signatures of behavior. Proc Natl Acad Sci U S A 106:2383–2388. https://doi.org/10.1073/pnas.0813021106
5. Chen R, Hui L, Cape SS et al (2010) Comparative neuropeptidomic analysis of food intake via a multi-faceted mass spectrometric approach. ACS Chem Neurosci 1:204–214. https://doi.org/10.1021/cn900028s
6. Sterkel M, Urlaub H, Rivera-Pomar R, Ons S (2011) Functional proteomics of neuropeptidome dynamics during the feeding process of Rhodnius prolixus. J Proteome Res 10:3363–3371. https://doi.org/10.1021/pr2001012
7. Che F-Y, Fricker LD (2005) Quantitative peptidomics of mouse pituitary: comparison of different stable isotopic tags. J Mass Spectrom 40:238–249. https://doi.org/10.1002/jms.743
8. Thompson A, Schäfer J, Kuhn K et al (2003) Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. Anal Chem 75:1895–1904
9. Romanova EV, Dowd SE, Sweedler JV (2013) Quantitation of endogenous peptides using mass spectrometry based methods. Curr Opin Chem Biol 17:801–808. https://doi.org/10.1016/j.cbpa.2013.05.030
10. Fricker LD (2015) Limitations of mass spectrometry-based peptidomic approaches. J Am Soc Mass Spectrom 26:1981. https://doi.org/10.1007/s13361-015-1231-x
11. Verdonck R, De Haes W, Cardoen D et al (2016) Fast and reliable quantitative peptidomics with labelpepmatch. J Proteome Res 15:1080–1089. https://doi.org/10.1021/acs.jproteome.5b00845
12. Maes E, Hadiwikarta WW, Mertens I et al (2016) CONSTANd: a normalization method for isobaric labeled spectra by constrained optimization. Mol Cell Proteomics 15(8):2779–2790
13. Maes E, Valkenborg D, Baggerman G et al (2015) Determination of variation parameters as a crucial step in designing TMT-based clinical proteomics experiments. PLoS One 10:e0120115. https://doi.org/10.1371/journal.pone.0120115
14. McAlister GC, Nusinow DP, Jedrychowski MP et al (2014) MultiNotch MS3 enables accurate, sensitive, and multiplexed detection of differential expression across cancer cell line proteomes. Anal Chem 86:7150–7158. https://doi.org/10.1021/ac502040v
15. Wühr M, Haas W, McAlister GC et al (2012) Accurate multiplexed proteomics at the MS2 level using the complement reporter ion cluster. Anal Chem 84:9214–9221. https://doi.org/10.1021/ac301962s
16. Ma B, Zhang K, Hendrie C et al (2003) PEAKS: powerful software for peptide de novo sequencing by tandem mass spectrometry. Rapid Commun Mass Spectrom 17:2337–2342. https://doi.org/10.1002/rcm.1196
17. Han X, He L, Xin L et al (2011) PeaksPTM: mass spectrometry-based identification of peptides with unspecified modifications. J Proteome Res 10:2930–2936. https://doi.org/10.1021/pr200153k
18. Zhang J, Xin L, Shan B et al (2012) PEAKS DB: de novo sequencing assisted database search for sensitive and accurate peptide identification. Mol Cell Proteomics 11:M111.010587. https://doi.org/10.1074/mcp.M111.010587

19. Han Y, Ma B, Zhang K (2005) SPIDER: software for protein identification from sequence tags with *de novo* sequencing error. J Bioinforma Comput Biol 3:697
20. R Core Team (2016) R: a language and environment for statistical computing. Vienna, Austria. http://www.r-project.org/. Accessed 24 Oct 2016
21. Shadforth IP, Dunkley TP, Lilley KS, Bessant C (2005) i-Tracker: for quantitative proteomics using iTRAQ. BMC Genomics 6:145. https://doi.org/10.1186/1471-2164-6-145
22. Morano C, Zhang X, Fricker LD (2008) Multiple isotopic labels for quantitative mass spectrometry. Anal Bioanal Chem 80:9298–9309
23. Chiva C, Sabidó E (2014) HCD-only fragmentation method balances peptide identification and quantitation of TMT-labeled samples in hybrid linear ion trap/orbitrap mass spectrometers. J Proteome 96:263–270. https://doi.org/10.1016/j.jprot.2013.11.013
24. Gelman JS, Wardman J, Bhat VB et al (2012) Quantitative peptidomics to measure neuropeptide levels in animal models relevant to psychiatric disorders. In: Kobeissy FH (ed) Methods Mol Biol. Humana Press, pp 487–503
25. Gelman JS, Dasgupta S, Berezniuk I, Fricker LD (2013) Analysis of peptides secreted from cultured mouse brain tissue. Biochim Biophys Acta 1834:2408–2417. https://doi.org/10.1016/j.bbapap.2013.01.043
26. Wardman J, Fricker LD (2011) Quantitative peptidomics of mice lacking peptide-processing enzymes. In: Mbikay M, Seidah NG (eds) Methods Mol Biol. Humana Press, Totowa, pp 191–206
27. Vaudel M, Burkhart JM, Zahedi RP et al (2015) PeptideShaker enables reanalysis of MS-derived proteomics data sets. Nat Biotechnol 33:22–24. https://doi.org/10.1038/nbt.3109
28. Rost HL, Sachsenberg T, Aiche S et al (2016) OpenMS: a flexible open-source software platform for mass spectrometry data analysis. Nat Methods 13:741–748. https://doi.org/10.1038/nmeth.3959

# Chapter 10

# Quantitative Peptidomics Using Reductive Methylation of Amines

**Sayani Dasgupta, Leandro M. Castro, Alexandre K. Tashima, and Lloyd Fricker**

## Abstract

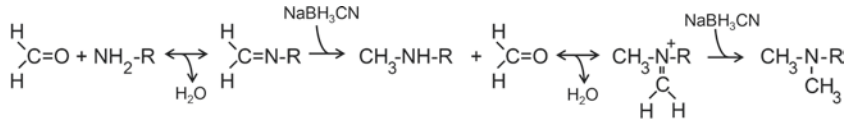
A number of different approaches have been used for quantitative peptidomics. In this protocol we describe the method in which peptides are reacted with formaldehyde and sodium cyanoborohydride, which converts primary and secondary amines into tertiary amines. By using different combinations of regular reagents, deuterated reagents ($^{2}H$), and reagents containing deuterium and $^{13}C$, it is possible to produce five isotopically distinct forms of the methylated peptides which can be quantified by mass spectrometry. Peptides with free N-termini that are primary amines incorporate two methyl groups using this procedure, which differ by 2 Da for each of the five isotopic combinations. Peptides that contain unmodified lysine residues incorporate additional pairs of methyl groups, leading to larger mass differences between isotopic forms. The reagents are commercially available, relatively inexpensive, and chemically stable.

**Key words** Isotopic labels, Formaldehyde, Cyanoborohydride

## 1 Introduction

A number of different approaches have been used to quantify the relative levels of peptides in two or more different samples [1–9]. Another chapter in this volume discusses the general considerations in choosing a quantitative peptidomics approach for a particular application (see Fricker, Quantitative Peptidomics). In this chapter, we describe a protocol for reductive methylation of amines. This general approach has been used for both proteomic and peptidomic applications as well as analysis of amine-containing metabolites and other biological applications [2, 10–15]. The labeling step is simple and uses reagents which are commercially available and inexpensive. Two reagents are needed: formaldehyde and sodium cyanoborohydride, both of which are commercially available in deuterated and nondeuterated forms, and formaldehyde is also available with deuterium and $^{13}C$. With various combinations of these reagents, one can generate five isotopic forms of labeled

Michael Schrader and Lloyd Fricker (eds.), *Peptidomics: Methods and Strategies*, Methods in Molecular Biology, vol. 1719, https://doi.org/10.1007/978-1-4939-7537-2_10, 




| Isotopic form of reagents | Product | Additional mass |
|---|---|---|
| $CH_2O$ and $NaBH_3CN$ | $(CH_3)_2$-N-R | 28.0313 |
| $CH_2O$ and $NaBD_3CN$ | $(CDH_2)_2$-N-R | 30.0439 |
| $CD_2O$ and $NaBH_3CN$ | $(CHD_2)_2$-N-R | 32.0564 |
| $CD_2O$ and $NaBD_3CN$ | $(CD_3)_2$-N-R | 34.0690 |
| $^{13}CD_2O$ and $NaBD_3CN$ | $(^{13}CD_3)_2$-N-R | 36.0757 |

**Fig. 1** Reaction scheme for the dimethylation of primary amines via the two-step process. The primary amine and formaldehyde form a Schiff base; this reaction is reversible in aqueous solutions. Reduction of the imine bond by cyanoborohydride produces a methylated secondary amine. This secondary amine reacts with another formaldehyde molecule to produce a Schiff base and is reduced by cyanoborohydride to form the tertiary amine. Peptides with N-terminal proline residues can only incorporate a single methyl group (not shown). The scheme shown at the top of the figure is for the forms of reagents containing hydrogen and $^{12}C$. The products formed using deuterated and/or $^{13}C$-containing reagents are indicated at the bottom of the figure

peptides which differ by 2 Da per primary amine in the peptide (Fig. 1). For peptides with multiple primary amines (i.e., a free N-terminus and one or more Lys residues) the mass difference between each isotopic form is sufficient for accurate quantitation. But for peptides with a single amine, there is considerable overlap in the spectra between the isotopic forms due to naturally occurring isotopes ($^{13}C$, $^{15}N$, $^{18}O$). This overlap complicates the analysis, although it is relatively easy to calculate the contribution from isotopic forms and subtract this amount from the total signal.

Peptides labeled with this protocol retain the positive charge on the amine, as the primary amines of the peptide are converted to tertiary amines (Fig. 1). Importantly, peptides labeled with deuterated and non-deuterated reagents coelute on reverse phase HPLC columns, which allows for accurate quantitation of the relative levels of each form. The reagents are stable, with a long shelf life, and the labeled peptides are also stable. This is in contrast to peptides labeled with trimethylammonium butyrate tags (a quaternary amine), which are unstable at high pH or under some mass spectrometry conditions [6, 16].

In this chapter, we also include a protocol to measure the content of amines with fluorescamine, which reacts with primary amines to form fluorescent pyrrolinone type moieties. However, fluorescamine reacts only with primary amines, not secondary or tertiary amines [17]. Reductive methylation is a two-step process, first converting a primary amine to a secondary amine, and subsequently to a tertiary amine (Fig. 1). Thus, the fluorescamine procedure can only be used to estimate the total concentration of primary amine in the sample and to test if the conversion of primary to secondary amine is complete. The ninhydrin procedure can potentially be used to monitor the completion of the reaction because it can distinguish between primary and secondary amines, thus providing a reliable means of establishing the completion of

the reaction [18]. However, ninhydrin is not very sensitive and will only be useful for procedures with very large amounts of peptides.

In this chapter, we include a protocol for peptide purification after labeling and the separation of peptides from proteins and salts that interfere with mass spectrometry. We also briefly mention specific liquid chromatography/mass spectrometry (LC/MS) methods that have been successfully used for our studies. A wide range of instruments can be used to detect the labeled peptide and it is not necessary to use the particular equipment described in this protocol. Finally, we discuss some of the key details of data analysis.

## 2 Materials

All aqueous solutions should be prepared in ultrapure water.

### 2.1 Fluorescamine Assay

1. Lysine stock solution, 0.1 mM.
2. Fluorescamine reagent, 10.8 mM (3 mg of fluorescamine dissolved in 1 mL of acetone).
3. Phosphate buffered saline (PBS), pH 7.4.
4. Plates, 96-well, compatible with fluorometer.
5. Fluorometer.

### 2.2 Reductive Methylation of Peptides

1. Biological samples containing up to 25 μg of peptide or the equivalent of 50 nmol of lysine (based on the fluorescamine assay), in a volume of ~1 mL or less in low retention 1.5 mL microfuge tubes (*see* **Notes 1** and **2**).
2. Formaldehyde solutions (4%): Prepare fresh dilution from stock on day of procedure (*see* **Note 3**). To label the amount of peptide described in this protocol requires 16 μL of formaldehyde solution per sample (*see* **Note 2**). Thus, to label two sets of five samples as described below and in Fig. 2 requires 64 μL solution of the light formaldehyde, 64 μL of the intermediate formaldehyde containing two deuteriums, and 32 μL of the heavy ($D_2{}^{13}C$) formaldehyde because the light and intermediate forms are each used for two samples in each set while the $D_2{}^{13}C$ only once per set. Commercially available forms of intermediate and heavy formaldehyde are available as 20% stock solutions, while the normal isotopic form (i.e., light) is available as a 37% stock.

   **Light**—dilute from 37 to 4%. Mix 11 μL stock with 89 μL water and split into two tubes of 50 μL (*see* **Note 4**).

   **Intermediate ($D_2$)**—dilute from 20 to 4%. Mix 20 μL stock with 80 μL water and then split into two tubes of 50 μL (*see* **Note 4**).

   **Heavy ($D_2{}^{13}C$)**—dilute from 20 to 4%. Mix 10 μL stock with 40 μL water.

| Sample | 2x8 µl of | 2x8 µl of | Set 1 | Set 2 | Mass diff per amine |
|---|---|---|---|---|---|
| 1 | $H_2CO$ | $NaBH_3CN$ | Control 1 | Treated 3 | 0 |
| 2 | $H_2CO$ | $NaBD_3CN$ | Control 2 | Treated 4 | 2 |
| 3 | $D_2CO$ | $NaBH_3CN$ | Control 3 | Treated 5 | 4 |
| 4 | $D_2CO$ | $NaBD_3CN$ | Treated 1 | Control 4 | 6 |
| 5 | $D_2{}^{13}CO$ | $NaBD_3CN$ | Treated 2 | Control 5 | 8 |

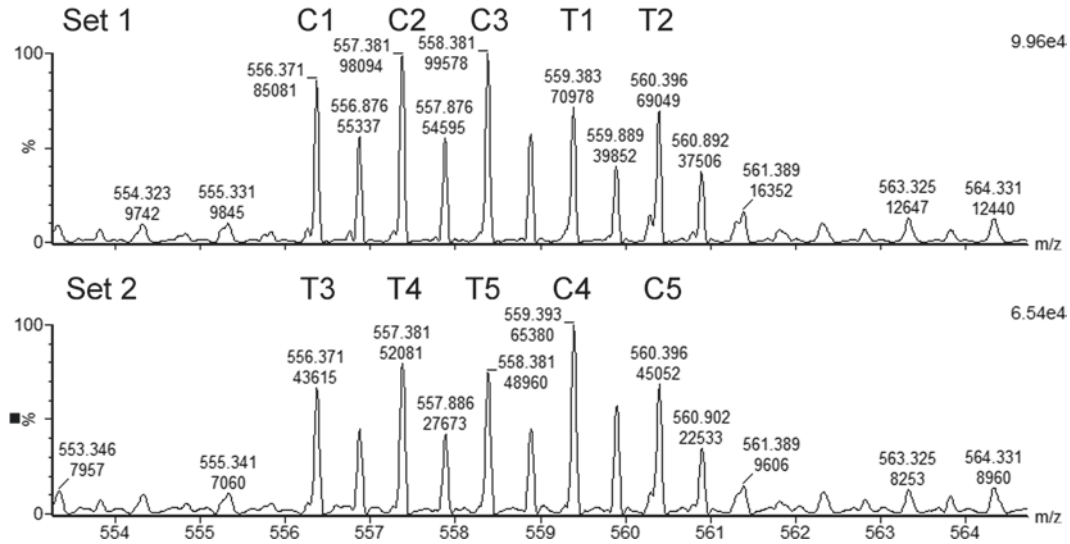



**Fig. 2** Table of reagents for a typical experiment and representative data. Using the indicated combination of heavy/light formaldehyde and borohydride, the resulting peptides differ by 2 Da per primary amine on the peptide (*see* **Note 6**). Because it is important to compare biological replicates, we typically perform two runs, one with three control replicates and two experimentally treated replicates, and the other run with two control and three treated replicates, as listed in the table. This provides $n = 5$ for each of two conditions. Note that the labels are reversed between the two sets of replicates in order to control for potential problems with the individual reagents. Spectra from a representative experiment performed on extracts of Human Embryonic Kidney 293 T cells are shown; the upper spectrum corresponds to set 1 with three control replicates (C1, C2, and C3) and the lower spectrum corresponds to set 2 with two control replicates (C4 and C5). The peptide shown in this example was subsequently found by MS/MS analysis to be the N-terminal fragment of Protein DJ-1 (Parkinson disease protein 7) with the sequence Ac-ASKRALVILA. For this example, the overlap between the peaks requires consideration of the relative abundance of naturally occurring isotopes within the peptide and subtraction of the background (*see* **Note 6**). Without adjustment, the relative levels of C1, C2, and C3 are 0.90, 1.04, and 1.06, respectively (standard deviation 0.085). After adjustment using the formulas listed in **Note 6**, the relative levels of C1, C2, and C3 are 1.03, 0.98, and 0.99, respectively (standard deviation 0.029)

3. Sodium cyanoborohydride solutions ($NaBH_3CN$ or $NaBD_3CN$; both 0.6 M): Prepare fresh right before procedure (*see* **Note 3**). Need 16 µL per sample, as shown in Fig. 2, which requires 64 µL of the light and 96 µL of heavy.

   **Light**—dissolve 3.77 mg in 100 µL water—split into 2 × 50 µL (*see* **Note 4**).

   **Heavy**—dissolve 5.93 mg in 150 µL water—split into 3 × 50 µL (*see* **Note 4**).

4. Ammonium bicarbonate solution (1%): Prepare fresh on day of reaction. Need 200 μL per sample (12.5 μmol), which requires 2 mL for the procedure described below. Dissolve 20 mg in 2 mL water.
5. Formic acid (5%). Use 100 μL per sample, which requires 1 mL for the procedure described below.
6. Sodium acetate buffer, 1 M, pH 6.0 adjusted with NaOH or acetic acid.
7. Sodium hydroxide, 1 M and 0.1 M.
8. HCl, 1 M and 0.1 M.
9. pH indicator paper, both extended range (such as 3–9) and narrow range close to 6 (such as 5.2–6.6).

### *2.3 Peptide Purification*

1. Ultrafiltration devices (we use Amicon Ultracel-10K units, with 4 mL capacity).
2. Centrifuge with swinging bucket rotor.
3. Activation solution, 50% acetonitrile in water.
4. Equilibration solution, 5% acetonitrile and 0.5% TFA in water
5. Sample solvent, 20% acetonitrile and 2% Trifluoroacetic acid (TFA) in water.
6. Wash solution, 5% acetonitrile and 0.5% TFA in water.
7. Elution buffer, 70% acetonitrile and 0.5% TFA in water.
8. Low retention 1.5 mL microfuge tubes.
9. C18 spin columns (we use spin columns from Pierce, although other manufacturer's columns should also work).
10. Benchtop microcentrifuge.
11. Vacuum centrifuge.

## 3 Methods

### *3.1 Fluorescamine Assay to Determine the Level of Primary Amine in a Sample*

1. Add aliquots of the peptide sample to wells of the 96-well plate. We typically assay 1 and 10 μL of each extract of a biological sample, which is usually 0.5 to 1 mL (*see* **Notes 1** and **2**).
2. Prepare standard curve of lysine by adding 1, 2, 5, 10, and 20 μL of 0.1 mM lysine stock solution to separate wells in a 96-well plate. Include duplicates of this standard curve. Also include a blank with PBS alone.
3. Add the appropriate volume of PBS to the standard curve and sample for a final volume of 150 μL.
4. Add 50 μL of fluorescamine reagent to the wells containing sample, standard, or blank.
5. Shake the plate for 1 min to mix the contents.

6. Record fluorescence at 400 nm excitation and 460 nm emission.
7. Estimate the amount of primary amine in the sample based on the standard curve.

### 3.2 Reductive Methylation of Peptides

1. Add 1/10th volume of 1 M sodium acetate buffer to samples (final 100 mM sodium acetate). Check pH by testing 1 μL with indicator paper, and adjust to pH 6 with HCl or NaOH using either 1 M or 0.1 M solutions, as needed.
2. In fume hood, add 8 μL of formaldehyde (light, intermediate, or heavy) according to Fig. 2. Proceed immediately to **step 3**, and work in fume hood through **step 7** (*see* **Note 3**).
3. Add 8 μL of $NaBH_3CN$ 0.6 M or $NaBD_3CN$ 0.6 M according to Fig. 2.
4. Vortex, check pH with indicator paper, and adjust with 0.1 M HCl or NaOH to 6 if necessary.
5. After 2 h incubation at room temperature, repeat **steps 2–4** and incubate overnight at room temperature.
6. Add 200 μL of ammonium bicarbonate, vortex, and incubate for 2 h at room temperature.
7. Add 100 μL of formic acid in fume hood (*see* **Note 3**) and incubate 10 min at room temperature.
8. Combine samples for each set and store at −80 °C until proceeding with microfiltration and desalting columns in Subheading 3.3.

### 3.3 Peptide Purification and Analysis

To separate proteins from the labeled peptides, use centrifugal filter units with a 10 kDa molecular weight cut-off cellulose membrane. The flow-through contains the peptides, and this is desalted on C18 reverse phase resin, dried in a vacuum centrifuge, and analyzed on LC/MS.

1. Clean the ultrafiltration devices by adding water and centrifuging at 2300 × *g* for 3 min. Repeat this step twice more.
2. Add samples from Subheading 3.2, **step 8**, into the cleaned ultrafiltration devices and centrifuge at 2300 × *g* for 30 min. The filtrate (i.e., the flow-through) contains the peptides. If the sample volume is more than 4 mL, remove and save the filtrate, load the rest of the sample, and repeat the centrifugation. If not proceeding immediately to the desalting on C18 resin, the filtrate can be stored at −80 °C.
3. Prepare the C18 spin column by tapping it to settle the resin. Remove the bottom cap and place it into a microfuge tube. Add 200 μL of activation solution (50% acetonitrile in distilled water) to the spin column and centrifuge at 1500 × *g* for 1 min. Discard flow-through and repeat once. Then add 200 μL of equilibration solution (5% acetonitrile and 0.5%

TFA in distilled water) and centrifuge at 1500 × *g* for 1 min. Discard flow-through. Repeat the addition of equilibration solution and centrifugation twice more. After this, the C18 column is ready for use.

4. Add 1/3 volume of sample solvent to the filtrate from **step 2** (i.e., for every 300 µL of filtrate from **step 2**, add 100 µL sample solvent). Load 400 µL to the top of the C18 resin bed and centrifuge at 1500 × *g* for 1 min. Discard flow-through. If the sample is larger than 400 µL, repeat this step until the entire sample has been loaded onto the C18 spin column.
5. Add 200 µL of wash solution (5% acetonitrile and 0.5% TFA in distilled water) to the C18 column and centrifuge at 1500 × *g* for 1 min. Discard flow-through and repeat this step twice more.
6. Place C18 spin column in a new microfuge tube and add 80 µL of elution buffer (70% acetonitrile and 0.5% TFA in distilled water). Centrifuge at 1500 × *g* for 1 min and save the material that elutes from the column. Add another 80 µL of elution buffer to the C18 column and repeat the centrifugation step, collecting into the same microfuge tube. The eluates can be stored at −80 °C.
7. Freeze-dry the eluates in a vacuum centrifuge. The dried eluates can be stored at −80 °C.
8. Resuspend the samples in a small volume of water (10 µL) and analyze an aliquot (2–5 µL) by liquid chromatography/mass spectrometry (LC/MS). A variety of LC/MS systems have been used for this, and many MS instruments are compatible with the methyl tags (*see* **Note 5**).
9. Spectra are analyzed using the appropriate software for the mass spectrometer used for the LC/MS (we typically use MassLynx 4.0 software, Waters). Peak groups representing peptides labeled with different isotopic labels are identified and the relative intensity of each monoisotopic peak is used for calculations. To quantify relative peptide levels, the peak intensity of each treated group is compared to the average of the control replicates in each experiment (*see* **Notes 6** and **7**).

   The monoisotopic mass of each peptide without added methyl groups or protons is calculated from the following formula:

   $$\text{Mass unmodified peptide} = (m/z_x{}^{*}z) - (C_x{}^{*}T) - (1.008{}^{*}z).$$

   $m/z_x$ is the observed mass to charge value for the monoisotopic peak for each peptide labeled with different combinations of light/heavy formaldehyde and borohydride ($x$ = 1, 2, 3, 4, or 5, corresponding to the sample number in Fig. 2).

$z$ is the charge state.

$C_x$ is the monoisotopic mass of a pair of methyl groups:

For $x = 1$, $C_x = 28.0313$ (the net addition of two $CH_3$ groups to the primary amine).

For $x = 2$, $C_x = 30.0439$ for two $CHD_2$ groups.

For $x = 3$, $C_x = 32.0564$ for two $CD_2H$ groups.

For $x = 4$, $C_x = 34.0690$ for two $CD_3$ groups.

For $x = 5$, $C_x = 36.0757$ for two $^{13}CD_3$ groups.

$T$ is the number of pairs of methyl groups incorporated into the peptide. This can be calculated from the following formula when five tags are used: $T = z * (m/z_5 - m/z_1)/8$.

For primary amines, $T$ is a whole number (e.g., 1, 2, 3) equal to the number of primary amines within the peptide. For secondary amines, such as peptides with N-terminal proline, $T = 0.5$ if there are no additional lysine residues within the peptide, or 1.5, 2.5, etc. if there are 1, 2, etc. primary amines in addition to the N-terminal proline (*see* **Note 7**).

10. To identify peptides, MS/MS data are analyzed using a database search engine (we use Mascot, Matrix Science Ltd., UK). No cleavage site is specified. Currently the Mascot server only allows for dimethylation with four of the five possible labels (the 2 Da form that contains two deuterium atoms is missing from the current version of the Mascot program, but can be manually added in the licensed version of the software). Because the N-terminal and lysine modifications are listed separately, this means that there are eight modifications listed on Mascot that need to be specified in the search parameters: Dimethyl (K); Dimethyl (N-term); Dimethyl: 2H (4) (K); Dimethyl: 2H(4) (N-term); Dimethyl: 2H(6) (K); Dimethyl: 2H(6) (N-term); Dimethyl: 2H(6)13C(2) (K); and Dimethyl: 2H(6)13C(2) (N-term). *See* **Notes 8** and **9** for additional search tips.

11. Results should be manually interpreted to eliminate false positives, using published criteria [6, 19]. Briefly, the criteria for considering a peptide identified from the MS/MS analysis are: (1) a parent mass within 0.1 Da of the theoretical mass (*see* **Note 10**); (2) the observed number of methyl groups added to the peptide matches the predicted number of free amines available (i.e., two methyl groups per Lys residue and free N-terminus unless proline, which incorporates only one methyl group; *see* **Note 7**); (3) the observed charge state(s) of the peptide is consistent with the expected number of positive charges; (4) 80% or more of the major fragments observed in MS/MS match predicted fragments (minimum five matches).

## 4 Notes

1. Peptide extraction prior to labeling is not described in detail here, and is the subject of other chapters in this volume. In brief, it is important to treat the tissue or cell extract to eliminate all enzymatic activity. One reason is to prevent the degradation of peptides by peptidases present in the sample. Another equally important reason is to prevent the formation of new peptides by the action of proteases on proteins present in the extracts. The goal of most peptidomics studies is to measure the naturally occurring peptides present in a sample, and post-extraction formation of peptides can greatly complicate the analysis. For this reason, we primarily use heat inactivation of the biological sample to 80 °C, either by microwave irradiation of animal tissues immediately after death or by heating cultured cells in hot water. Avoid hot acid, which is known to degrade peptides, especially adjacent to Asp residue [20, 21].
2. This protocol uses a vast excess of formaldehyde and sodium cyanoborohydride relative to the amine. The precise ratio depends on the amount of free amine in the peptide (determined in part from fluorescamine, although this detects only primary amines). If peptide equivalent of 50 nmol of lysine is used, as described above, this corresponds to 100 nmol of free amines. Using the amounts of formaldehyde and sodium cyanoborohydride described above, and considering that two molecules of each are required to convert the primary amine into a dimethylamine, there is a 50–100 fold excess of the two reagents relative to the amine content. When this procedure was tested with 50 nmol of lysine, there was >99% completion of both the first and second steps of the reaction, as determined using ninhydrin to measure the level of secondary amine. When a tenfold excess of reagents relative to the number of amine groups was used, the formation of monomethylamine was still >99% complete but the formation of dimethylamine was only ~98% complete. Therefore, it is recommended to use a ratio of reagents similar to the protocol described here, and if the biological samples contain a higher amine content than 100 nmol, then increase the other reagents accordingly to maintain a 50–100 fold ratio. Because fluorescamine detects only primary amines, it cannot be used to determine if the dimethylation reaction is complete because the intermediate (i.e., mono-methylation) is a secondary amine. However, it will be apparent from the mass spectra of the data whether the reaction is

complete, or whether some primary amines were only labeled with a single methyl group.

3. Formaldehyde and sodium cyanoborohydride are very toxic and all procedures with these reagents should be performed in a fume hood (including weighing the sodium cyanoborohydride). The quenching reaction and acidification may also result in the generation of hydrogen cyanide, a toxic gas. Although the amounts of these reagents are very small, it is important to work in a fume hood until after the samples are quenched.
4. To reduce the potential for human error in the addition of the wrong reagent to the sample tubes, the reagents used for multiple reactions can be split into aliquots so that each sample tube to be labeled is placed on a rack together with the two reagents that will be added (i.e., one of the three forms of formaldehyde and one of the two forms of the reducing agent). This simplifies the addition process and reduces the chance of inadvertently adding the wrong reagent.
5. Many types of mass spectrometers can be used to analyze the peptides. We typically use a Synapt G2 Q-TOF mass spectrometer coupled to a nanoAcquity capillary liquid chromatography (LC) system (Waters, Milford, MA, USA). The peptide mixture is desalted online for 5 min at a flow rate of 8 μL/min of phase A (0.1% formic acid) using a Symmetry C18 trapping column (5-μm particles, 180-μm inner diameter, 20-mm length; Waters). The trapped peptides are subsequently eluted with a gradient of 7–65% over 60 min of phase B (0.1% formic acid in acetonitrile) through a BEH 130 C18 column (1.7-μm particles, 75-μm inner diameter, 200-mm length; Waters). The data are acquired in the data-dependent mode and the mass spectra of multiple-charged protonated peptides generated by electrospray ionization are acquired for 0.2 s from $m/z$ 300–1600. The three most intense ions exceeding base peak intensity threshold of 2500 counts are automatically selected and tandem mass spectrometry (MS/MS) is performed by dissociation of the ions by 15 to 60-eV collisions with argon for 0.2 s. The typical LC and electrospray ionization conditions are a flow rate of 250 nL/min, a capillary voltage of 3.0 kV, a block temperature of 70 °C, and a cone voltage of 50 V. The dynamic peak exclusion window is set to 90 s.
6. Peak overlap can be a problem for peptides that contain a single primary amine and therefore are labeled with only two methyl groups. For most peptides, the 2 Da difference

between the labeled peptides needs to be adjusted to take into account the natural abundance of stable isotopes within the peptides: primarily $^{13}C$ and also $^{15}N$ and $^{18}O$ atoms. For example, the representative data shown in Fig. 2 should have 1:1:1 relative levels of the three control replicates, but the relative peak intensities are 0.90, 1.04, and 1.06. Peptides with a higher mass show even greater deviation from the theoretical 1:1:1 ratio. It is possible to take into account the natural isotopic content based on the known isotopic distribution of the elements and the average levels of each atom in peptides (which will vary slightly based on amino acid composition). Figure 3 shows the relative levels of the monoisotopic ion and the 2nd, 4th, 6th, and 8th isotopes of a set of peptides covering the mass range of 0.5–5 kDa—these are the ions that would interfere with the dimethylation of a single primary amine. Formulas were derived to adjust the relative peak intensity and subtract the natural isotopic composition of the peptides. For peptides that contain a single primary amine and incorporate two methyl groups, the formulas to adjust based on the mass ($M$) of the observed peptide are:

$$S_i = \frac{I_i}{k_i} - I_b$$

in which $S_i$ is the adjusted intensity of sample $i$ ($i$ = 1, 2, 3, 4, or 5), $I_i$ is the observed intensity, $k_i$ is an adjusting factor, and $I_b$ is the background intensity. The adjusting factor $k_i$ is given by:

$$k_i = a_i + b_i M + c_i M^2 + d_i M^3$$

and $a_i$ to $d_i$ are adjusted constants based on the isotopic distribution calculations (Table 1).

7. Proline is a secondary amine and therefore an N-terminal proline will only incorporate a single methyl group from the reaction with formaldehyde and sodium cyanoborohydride using the scheme described in this protocol. This means that the mass difference between each isotopic tag will be only 1 Da if there are no lysines within the peptide, or 3 Da if there is one lysine, 5 Da if two lysines, etc. [2].
8. In addition to including the modifications for the dimethylation reaction when performing Mascot searches, it is useful to include additional post-translational modifications. The modifications to include in the search parameters depend on the biological sample. Met-oxide is a common modification found in a wide range of studies. If looking for neuropeptides, search for amidation, acetylation, pyroglutamylation, and phosphorylation. But

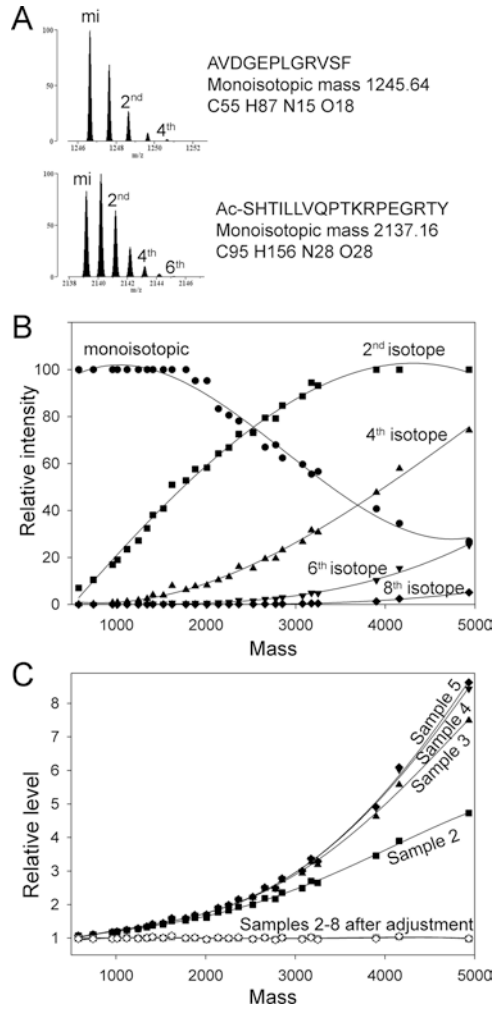



**Fig. 3** Relative intensity of the monoisotopic ion and relative abundance of naturally occurring isotopic forms. (**a**) Two representative peptides, previously identified from peptidomic analysis of Human Embryonic Kidney 293 T cells [26, 27], were analyzed using the Protein Prospector MS-Isotope calculator (http://prospector.ucsf.edu/prospector/cgi-bin/msform.cgi?form=msisotope), which takes into account the natural abundances of $^{13}C$, $^{15}N$, and $^{18}O$. The monoisotopic (mi) peak and the second, fourth, and sixth isotopic peaks are indicated—these complicate the analysis of peptides labeled with pairs of methyl groups which are 2 Da apart using the scheme in Figs. 1 and 2. (**b**) Twenty-six previously identified peptides [26, 27] were analyzed using the Protein Prospector MS-Isotope calculator. The selected peptides range in size from 578.31 to 4933.52 Da. The relative abundance of the monoisotopic ion and the ions corresponding to the second, fourth, sixth, and eighth isotope are indicated. The relative intensity (y-axis) refers to the relative value shown in ProteinProspector (*see* Panel (**a**)). (**c**) The relative level of peak intensity due to natural isotopes of lighter peaks separated by 2 Da was calculated for samples 2–5 (using the scheme listed in Figs. 1 and 2). In this graph, sample 1 would have a relative level of 1.00. For this analysis, the contribution of isotopes from each previous sample was included (e.g., for Sample 5 this involved the second isotope peak from Sample 4, the 4th isotope peak from Sample 3, the 6th isotope peak from Sample 2, and the 8th isotope peak from Sample 1). Then, the ratio of the $^{13}C$-containing peak relative to the monoisotopic peak was calculated. Filled symbols represent unadjusted results, open symbols represent values after adjustment using the formulas described in **Note 6**. Without adjustment a 4 kDa peptide labeled with two methyl groups would be off by >500% from the theoretical ratio, and a 5 kDa peptide would be off by >800%. After adjustment, all peptides were within 7% of the theoretical 1:1:1:1:1 ratio

**Table 1**
**Adjusting factor $k_i$ constants for each experimental condition ($i$ = 1–5)**

| $i$ (sample) | $a_i$ | $b_i$ | $c_i$ | $d_i$ |
|---|---|---|---|---|
| 1 | 1 | 0 | 0 | 0 |
| 2 | 0.991 | $4.54 \times 10^{-5}$ | $1.48 \times 10^{-7}$ | 0 |
| 3 | 1.089 | $-4.49 \times 10^{-5}$ | $1.29 \times 10^{-7}$ | $3.00 \times 10^{-11}$ |
| 4 | 0.993 | $1.62 \times 10^{-4}$ | $-3.95 \times 10^{-9}$ | $5.70 \times 10^{-11}$ |
| 5 | 0.958 | $2.33 \times 10^{-4}$ | $-4.60 \times 10^{-8}$ | $6.44 \times 10^{-11}$ |

if any of these are found, they need to be interpreted with caution. For example, amidation of neuropeptides is only known to occur for peptides that are produced from a precursor with a Gly in the downstream position which is converted to the amide group by the enzyme peptidyl-glycine-alpha-amidating monooxygenase [22, 23]. Likewise Ser and Thr phosphorylation of secretory pathway peptides is performed by an enzyme with a very strict consensus site, and does not phosphorylate most Ser or Thr residues [24, 25].

9. Because of the small mass difference between the isotopic forms, the program that calculates the monoisotopic mass of the observed peptide does not always pick the correct peak, and sometimes selects one of the $^{13}$C-containing peaks. Therefore, database searches with a small error tolerance (such as 0.1 Da) will not identify these mis-assigned peaks because they will be incorrectly considered to be off by 1 Da. A solution is to use a very large mass window when searching Mascot, such as 1.1 Da, and then consider potential hits that are within 0.1 Da or 0.9–1.1 Da (i.e., within 0.1 Da of a peptide with a mass 1 Da heavier/lighter). Manual analysis of the data will reveal if Mascot selected a $^{13}$C-containing peak instead of the monoisotopic peak (*see* **Note 10**).
10. Depending on mass spectrometer used for the analysis, the mass accuracy window can be even narrower than 0.1 Da. For example, with a Synapt G2 (Waters), the accuracy is often within 0.02 Da. Conversely, if performed on an instrument with lower mass accuracy, this needs to be considered in the interpretation of the data.

## References


1. Che FY, Fricker LD (2002) Quantitation of neuropeptides in Cpe fat/Cpe fat mice using differential isotopic tags and mass spectrometry. Anal Chem 74:3190–3198
2. Fu Q, Li L (2005) De novo sequencing of neuropeptides using reductive isotopic methylation and investigation of ESI QTOF MS/MS fragmentation pattern of neuropeptides with N-terminal dimethylation. Anal Chem 77:7783–7795
3. Che FY, Fricker LD (2005) Quantitative peptidomics of mouse pituitary: comparison of different stable isotopic tags. J Mass Spectrom 40:238–249
4. Rogatsky E, Balent B, Goswami G, Tomuta V, Jayatillake H, Cruikshank G, Vele L, Stein DT (2006) Sensitive quantitative analysis of C-peptide in human plasma by 2-dimensional liquid chromatography-mass spectrometry isotope-dilution assay. Clin Chem 52:872–879
5. Fricker LD, Lim J, Pan H, Che FY (2006) Peptidomics: identification and quantification of endogenous peptides in neuroendocrine tissues. Mass Spectrom Rev 25:327–344
6. Morano C, Zhang X, Fricker LD (2008) Multiple isotopic labels for quantitative mass spectrometry. Anal Chem 80:9298–9309
7. Romanova EV, Lee JE, Kelleher NL, Sweedler JV, Gulley JM (2012) Comparative peptidomics analysis of neural adaptations in rats repeatedly exposed to amphetamine. J Neurochem 123:276–287
8. Romanova EV, Dowd SE, Sweedler JV (2013) Quantitation of endogenous peptides using mass spectrometry based methods. Curr Opin Chem Biol 17:801–808
9. Southey BR, Lee JE, Zamdborg L, Atkins N Jr, Mitchell JW, Li M, Gillette MU, Kelleher NL, Sweedler JV (2014) Comparing label-free quantitative peptidomics approaches to characterize diurnal variation of peptides in the rat suprachiasmatic nucleus. Anal Chem 86:443–452
10. Hsu JL, Huang SY, Chow NH, Chen SH (2003) Stable-isotope dimethyl labeling for quantitative proteomics. Anal Chem 75:6843–6852
11. Guo K, Ji C, Li L (2007) Stable-isotope dimethylation labeling combined with LC-ESI MS for quantification of amine-containing metabolites in biological samples. Anal Chem 79:8631–8638
12. Kleifeld O, Doucet A, Prudova A, auf dem Keller U, Gioia M, Kizhakkedathu JN, Overall CM (2011) Identifying and quantifying proteolytic events and the natural N terminome by terminal amine isotopic labeling of substrates. Nat Protoc 6:1578–1611
13. Larda ST, Bokoch MP, Evanics F, Prosser RS (2012) Lysine methylation strategies for characterizing protein conformations by NMR. J Biomol NMR 54:199–209
14. Boersema PJ, Raijmakers R, Lemeer S, Mohammed S, Heck AJ (2009) Multiplex peptide stable isotope dimethyl labeling for quantitative proteomics. Nat Protoc 4:484–494
15. Melanson JE, Avery SL, Pinto DM (2006) High-coverage quantitative proteomics using amine-specific isotopic labeling. Proteomics 6:4466–4474
16. Fricker LD (2015) Limitations of mass spectrometry-based peptidomic approaches. J Am Soc Mass Spectrom 26:1981–1991
17. Udenfriend S, Stein S, Bohlen P, Dairman W, Leimgruber W, Weigele M (1972) Fluorescamine: a reagent for assay of amino acids, peptides, proteins, and primary amines in the picomole range. Science 178:871–872
18. Fitzpatrick WH (1949) Spectrophotometric determination of amino acids by the ninhydrin reaction. Science 109:469
19. Wardman J, Fricker LD (2011) Quantitative peptidomics of mice lacking peptide-processing enzymes. Methods Mol Biol 768:307–323
20. Che FY, Zhang X, Berezniuk I, Callaway M, Lim J, Fricker LD (2007) Optimization of neuropeptide extraction from the mouse hypothalamus. J Proteome Res 6:4667–4676
21. Allen G (1981) Sequencing of proteins and peptides. Elsevier, Amsterdam
22. Chufan EE, De M, Eipper BA, Mains RE, Amzel LM (2009) Amidation of bioactive peptides: the structure of the lyase domain of the amidating enzyme. Structure 17:965–973
23. Eipper BA, Mains RE (1988) Peptide alpha-amidation. Ann Rev Physiol 50:333–344
24. Tagliabracci VS, Engel JL, Wiley SE, Xiao J, Gonzalez DJ, Nidumanda Appaiah H, Koller A, Nizet V, White KE, Dixon JE (2014) Dynamic regulation of FGF23 by Fam20C phosphorylation, GalNAc-T3 glycosylation, and furin proteolysis. Proc Natl Acad Sci U S A 111:5520–5525
25. Tagliabracci VS, Xiao J, Dixon JE (2013) Phosphorylation of substrates destined for secretion by the Fam20 kinases. Biochem Soc Trans 41:1061–1065
26. Dasgupta S, Castro LM, Dulman R, Yang C, Schmidt M, Ferro ES, Fricker LD (2014) Proteasome inhibitors alter levels of intracellular peptides in HEK293T and SH-SY5Y cells. PLoS One 9:e103604
27. Gelman JS, Sironi J, Castro LM, Ferro ES, Fricker LD (2011) Peptidomic analysis of human cell lines. J Proteome Res 10:1583–1592

# Chapter 11

# Metabolic Labeling to Quantify *Drosophila* Neuropeptides and Peptide Hormones

**Thomas Otto Kunz, Jiangtian Chen, Megha, and Christian Wegener**

## Abstract

Neuropeptides and peptide hormones are involved in the regulation of most if not all body functions, ranging from physiology to neuronal processing and the control of behavior. To assess their functions, it is often vital to determine when and in which quantities they are produced, stored, and released. The latter is especially difficult to assess in small insects, such as the genetically amenable fruit fly *Drosophila melanogaster*, and cannot be achieved merely by quantifying mRNA transcripts. We have adapted and optimized methods to quantify neuropeptides and peptide hormones by metabolic labeling followed by LC-MS. In this chapter, we describe the labeling protocols used in our laboratory and discuss problems and pitfalls that we encountered.

**Key words** Quantitative peptidomics, Neuropeptides, Metabolic labeling, Insects, LC-MS

## 1 Introduction

The development of highly sensitive mass spectrometric techniques and the miniaturization of liquid chromatography enabled the biochemical characterization of neuropeptides and peptide hormones (here, for simplicity, referred to as peptides) in the small fruit fly *Drosophila* (*see* [1]). This peptidomic characterization, in turn, has fueled the use of sophisticated *Drosophila* genetic tools to dissect insect peptide functions and signaling. For a comprehensive understanding of the physiology and regulation of peptidergic signaling systems, quantitative information about the timing of peptide production, storage, and release is essential. Though it is rather straightforward to quantify peptide hormone titers in the circulation of larger animals, attempts for the fruit fly have so far been largely unsuccessful due to the small volume and the high peptide complexity of insect hemolymph [2, 3]. Measurement of peptide release within the nervous system remains a real challenge since nervous system extracts contain both stored peptides (from intracellular

Michael Schrader and Lloyd Fricker (eds.), *Peptidomics: Methods and Strategies*, Methods in Molecular Biology, vol. 1719, https://doi.org/10.1007/978-1-4939-7537-2_11, 

vesicle pools) and released peptides (from the extracellular space). Thus, prepropeptide mRNA is usually quantified to obtain information about when the production of a peptide is upregulated, indicating peptide release and active peptide signaling. However, prepropeptide mRNA quantification can only serve as a rough approximation, as peptides can be intracellularly stored in large amounts. Moreover, it is unclear whether an increased amount of prepropeptide mRNA necessarily results in increased peptide production (translation plus processing) or release, and whether release activity is always coupled to an upregulation of prepropeptide gene transcription and translation. A more realistic approach to assess peptide release also takes the amount of peptide into account and can be represented by the formula $R_{t_0-t_x} = A_{t_0} \quad A_{t_x} + P_{t_0-t_x}$, where $R$ is the amount of peptide released, $A$ is the total amount of peptide in the tissue, $P$ is the amount of peptide produced, and $t_0 - t_x$ is the time between timepoints $t_0$ and $t_x$. For example, if the peptide's mRNA level is upregulated at $t_x$ compared to $t_0$, while the total amount of peptide within the tissue is unchanged ( $A_{t_0} = A_{t_x}$ ), then it is reasonable to assume that peptide was released and hence peptide signaling was active during the physiological condition during the time $t_0 - t_x$.

In this chapter, we describe a protocol to metabolically label peptides in order to obtain quantitative information about the total peptide amount in a tissue. This allows comparison of peptide levels between flies of different genotypes or physiological states, and if $P$ is known, estimation of peptide release. The protocol is an adapted and more economical version of a protocol originally developed by Gouw and colleagues for *Drosophila* proteomics [4]. Unlike peptide quantification by direct mass-spectrometric tissue profiling (*see*, e.g., [5–7]), metabolic labeling allows the quantification of peptides in any tissue and is not restricted to isolated cells or small and anatomically well-defined brain compartments or neurohemal organs.

## 2 Materials

### 2.1 Yeast Labeling

1. Minimal medium: 1.7 g yeast nitrogen base without amino acids and ammonium sulfate (e.g., BD Difco). 20 g sucrose (analysis grade, nitrogen-free), 1 g $^{15}$N-labeled ammonium sulfate (for $^{15}$N-labeling; we use 99% $^{15}$N-labeled ammonium sulfate from Euriso-Top, Saarbrücken, Germany; *see* **Note 1**) or 1 g normal ammonium sulfate (for $^{14}$N-labeling). Fill up to 1 L with ultrapure water.
2. Phosphate-buffered saline (PBS): 8 g NaCl, 0.2 g KCl, 1.44 g $Na_2HPO_4$, 0.24 g $KH_2PO_4$, 900 mL double distilled water. Adjust to pH 7.4 with $Na_2HPO_4$ or $KH_2PO_4$ if necessary, then fill up to 1 L with double distilled water.

3. *Saccharomyces cerevisiae* (yeast, e.g., Sigma, type II YSC2).
4. 50 mL tubes (e.g., Falcon tubes).
5. Sterile 2.5 L Erlenmeyer flask.
6. Shaking incubator.
7. Large refrigerated centrifuge.

### 2.2 Fruit Fly Labeling

1. For timed larvae: 60 mm plates containing 2% agarose and 1.7% sucrose (250 mM). Prepared fresh on day of egg collection.
2. For adult flies: apple juice plates: add 24 g agar to 700 mL water and bring to boil. Then cool down to around 70 °C and add 300 mL apple juice. Pour into plastic Petri dishes and let harden. Can be stored at 4 °C.
3. Agarose and sucrose.
4. Cotton wool.
5. Medium-sized fly culture vials (Fig. 1).
6. Fly incubator.
7. Yeast-sugar solution: combine 3.4 g yeast ($^{15}$N-labeled or unlabeled) with 0.9 g sucrose and 7 mL water.

### 2.3 Tissue Dissection and Extraction and Purification of Peptides

1. Dissecting saline: e.g., hemolymph-like solution HL3.1 [8] (in mM): NaCl 128, KCl 2, $CaCl_2$ 1.8, $MgCl_2$ 4, sucrose 36, HEPES 5, pH 7.1 adjusted with 1N HCl, or alternatively: NaCl 80, KCl 5, $CaCl_2$ 1.5, $MgCl_2$ 4, $NaHCO_3$ 10, trehalose 5, sucrose 115, HEPES 5, pH 7.2, adjusted with 1N HCl [9].

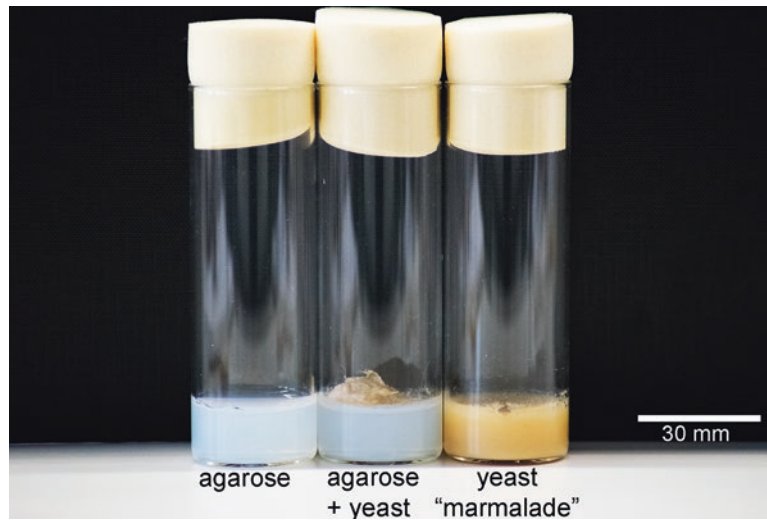



**Fig. 1** Fly culture vials prepared for metabolic labeling. Left: vial filled with pure agarose. Middle: vial filled with agarose plus loose cotton wool on top which is soaked to saturation with yeast paste. Right: vial filled with "yeast marmalade" (*see* **Note 7**)

2. A pair of fine forceps (e.g., sharpened Dumont No. 5), and a micro-needle.
3. A silicon polymer-coated preparation dish (*see* **Note 2**).
4. Dissecting microscope with high 20–40× magnification.
5. Microcentrifuge tubes, 0.5 or 1.5 mL, low-binding (*see* **Note 3**).
6. Labtop cooler for microcentrifuge tubes that keeps temperature <−10 °C.
7. Methanol extraction: methanol/ultrapure water/trifluoroacetic acid (TFA) (90:9:1 (v/v/v)), all HPLC grade.
8. Water extraction: ultrapure water, later to be added: 1 mM HCl.
9. Heating device to boil the sample (heating plate or gas burner and a beaker with water, or a thermoblock for microcentrifuge tubes).
10. Ultrasonic waterbath.
11. Ice.
12. Microcentrifuge.
13. Vacuum concentrator.
14. Pipettes and low protein-binding tips.

### 2.4 LC-MS

This protocol uses LC offline-coupled to MALDI-TOF mass spectrometry. Although many investigators use ESI-MS for peptidomic studies, MALDI-TOF MS holds several advantages (*see* **Note 4** and Fig. 2).

1. Reverse phase HPLC system, such as the Dionex UltiMate 3000, ThermoFisher.
2. C18 trap column, such as Acclaim PepMap100 C18, 5 mm, 100 Å pore size trap column.
3. C18 analytical column, such as Acclaim PepMap100 C18, 3 mm, 100 Å pore size analytical column.
4. MALDI spotter, such as SunCollect, SunChrom, GmbH, Friedrichsdorf, Germany.
5. MALDI TOF/TOF analyzer, such as 4800 Plus, ABI Sciex, Framingham, MA, USA.
6. Eluent A: 98% HPLC grade water, 2% acetonitrile (ACN), 0.05% trifluoroacetic acid (TFA) (v/v/v).
7. Eluent B: 80% ACN, 20% HPLC grade water, 0.04% TFA (v/v/v).
8. Matrix solution: half-saturated alpha-cyano-4-hydroxycinnamic acid in 60% ACN, 40% HPLC grade water, 0.1% TFA (v/v/v).

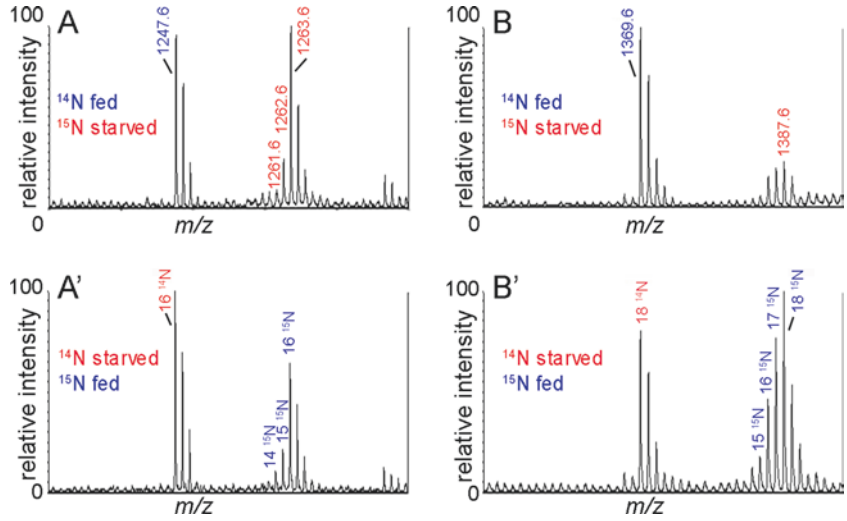



**Fig. 2** Examples of MALDI-TOF MS spectra for labeled and unlabeled peptides. (**a**) Mass spectrum showing unlabeled ($[M+H]^+$ = 1247.6 Da) and $^{15}N$-labeled ($[M+H]^+$ = 1263.6 Da) myosuppressin (containing 16 nitrogens), methanol-extracted from pooled brains of fed (unlabeled) and 24 h starved (labeled) flies raised on yeast-soaked cotton wool. (**a′**) Same as in (**a**), but here fed flies are labeled and starved flies are unlabeled. In both (**a**) and (**a′**) the unlabeled myosuppressin shows the expected isotopic pattern, while the peaks at 1262.6 and 1261.6 Da (incorporation of 1 or 2 $^{14}N$, respectively) indicate that the $^{15}N$ labeling efficiency is 98%. (**b**) Mass spectrum showing unlabeled ($[M+H]^+$ = 1369.6 Da) and $^{15}N$-labeled ($[M+H]^+$ = 1387.6 Da) corazonin (containing 18 nitrogens), methanol-extracted from pooled brains of fed (unlabeled) and 24 h starved (labeled) fed with "yeast marmalade." (**b′**) Same as in (**b**), but here fed flies are labeled and starved flies are unlabeled. In both (**b**) and (**b′**) the unlabeled corazonin shows the expected isotopic pattern, while the peaks at 1386.6, 1385.6, and 1384.6 Da indicate that the $^{15}N$ labeling efficiency is only 93%, most likely due to ammonium salt contaminations in the pectin. This low labeling efficiency is problematic for a proper ESI-MS analysis, but can easily be accounted for in the MALDI spectra. The data would suggest that corazonin levels are decreased by starvation, while myosuppressin levels are not greatly affected by starvation

## 3 Methods

### *3.1 Yeast Labeling*

1. Add 5 mL minimal medium in a 50 mL Falcon tube, and inoculate with yeast.
2. Grow overnight at 30 °C on a shaker with 230–270 rpm.
3. Use 500 μL of the overnight culture to inoculate 1 L of minimal medium in a 2.5 L Erlenmeyer flask. This is done for both the $^{14}N$- and $^{15}N$-enriched medium.
4. Grow overnight at 30 °C on a shaker with 230–270 rpm.
5. Transfer the culture to centrifuge bottles and centrifuge at 2400 × *g* for 20 min at 4 °C.
6. Transfer the supernatant back to the Erlenmeyer flask. As enough yeast are present in the supernatant, it can be used for a second overnight incubation as above, to collect further yeast.

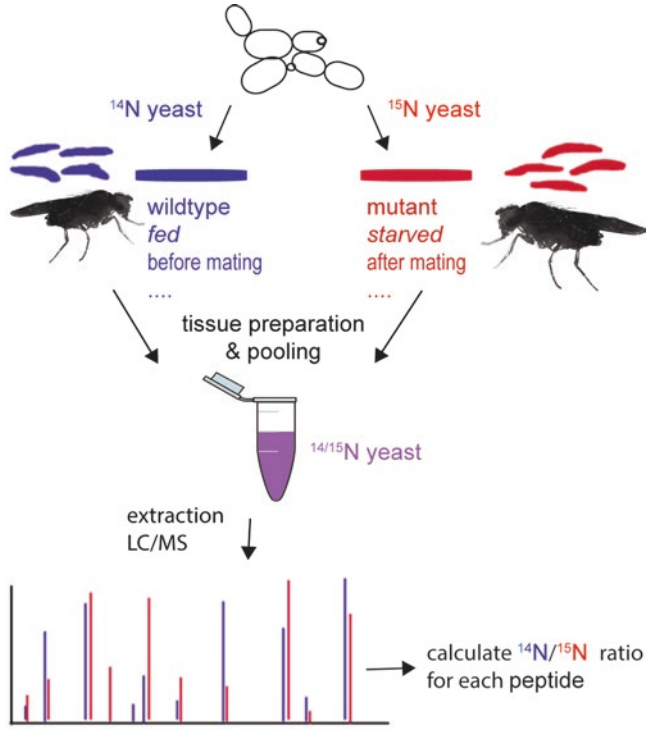



**Fig. 3** Scheme for metabolic labeling. Test and control flies are fed with $^{14}$N and $^{15}$N-labeled yeast, respectively. Then, the tissue of interest is dissected from an equal number of test and control flies directly into the same microcentrifuge tubes. After extraction, peptides can then be quantified by LC-MS. The same experiment should in parallel be conducted with reciprocally labeled flies

7. Resuspend the yeast pellet in 20 mL PBS and transfer to a 50 mL Falcon tube. Centrifuge at 2400 × *g* for 20 min at 4 °C.
8. Remove the supernatant, and use the tube to pool further yeast (i.e., from the second overnight incubation).
9. If you only need small portions of $^{14}$N- and $^{15}$N-labeled yeast, aliquot appropriately in smaller tubes.
10. Store the tubes with yeast at 4 °C for direct use or at −80 °C for long-term storage.

### 3.2 Fruit Fly Labeling

Run each set of experimental and control flies in parallel, and reciprocally on $^{14}$N or $^{15}$N-labeled yeast (Fig. 3) to minimize possible effects due to heavy-labeled yeast (*see* **Note 5**).

For timed larvae:

1. Egg collection: To ensure that a synchronized larval population is recovered, allow flies to do a pre-egg lay of 1 h on regular food. At the desired time point of transfer, allow flies (20–30 females, 10 males) to lay eggs on the freshly prepared agarose

and sugar plate. Allow egg laying to proceed depending on fecundity of flies. In our laboratory, 3 h was sufficient.

2. After egg collection, cut out the agarose from the middle, creating a 1.5 cm × 1.5 cm square. Place a small amount of cotton wool to fill the area. The cotton wool should be loosely held and not dense. Soak with 500 μL yeast-sugar solution. Place in 25 °C incubator. Hatchlings will migrate to the cotton wool by next day. Add 200 μL of fresh labeled or unlabeled media everyday till time point of collection.
3. At the desired time point of collection (we used 88 h after egg laying), larvae are collected from the cotton wool with forceps, washed in Ringer solution, and placed on ice. The dissection protocol for larvae is similar as for adults.

For adult flies:

1. Place flies (20–30 females, 10 males) on an apple juice agar plate overnight for oviposition.
2. The next morning, wash the eggs carefully off the plate. Transfer a batch of 30–35 eggs to a freshly prepared fly culture vial filled with 2–3 cm 2% agarose onto which a small layer of cotton wool was added (Fig. 1). Pipette yeast paste onto the cotton layer until the cotton is fully soaked. Soak the cotton layer daily with new yeast-sugar solution (*see* **Note 6**).
3. Keep flies at 25 °C and the desired housing conditions until dissection. If larvae or flies are to be starved, transfer them onto fresh vials filled with 2% agarose for the desired starvation time (*see* **Note 7**).

### 3.3 Tissue Dissection and Extraction

For better illustration, we here describe the dissection and extraction using an experiment with $^{14}$N-labeled fed vs. $^{15}$N-labeled starved flies as an example. The same applies for comparison of any two groups of adult flies or larvae, such as wild-type vs. mutant flies, or before vs. after mating.

1. Anesthetize flies on ice. Then quickly dissect the $^{14}$N-labeled control tissue of interest (e.g., brain, ventral ganglia, midgut) in ringer solution on ice.
2. Using a needle, transfer the tissue with as little saline as possible to the bottom of a pre-cooled and empty microcentrifuge tube kept in the labtop cooler. The tissue should freeze to the tube wall. Keep on dissecting until 20–30 $^{14}$N-labeled brains, ventral ganglia, or midguts have been collected.
3. Repeat the dissection with the $^{15}$N-labeled starved flies and collect the same amount of $^{15}$N-labeled tissue into the same tube as for the $^{14}$N-labeled tissue. You should end up with a tube containing 40–60 tissues, of which half is labeled by $^{14}$N, the other by $^{15}$N.

4. Extract directly, or store the dry frozen tissue in the tube at −80 °C until use.
5. (a) For methanolic extraction, add 30 µL methanol/ultrapure water/TFA (90/9/1, v/v/v) to the frozen tissue samples and let thaw on ice. For nervous system samples, sonicate the tissue in a cold ultrasonic water bath for 1 min until the tissue is largely disintegrated (*see* **Note 8**). Then, for all tissues, incubate on ice for 30 min, and centrifuge for 15 min at 15,000 × *g* (maximum speed in a microcentrifuge). Transfer the supernatant into a fresh tube. The pellet can be re-extracted as above, and the supernatant pooled with that of the first extraction. Then, centrifuge the supernatant at maximum speed and dry down in a vacuum concentrator to get rid of the methanol prior to the LC step (*see* **Note 9**). The dried-down sample can be either directly injected into the LC system, or kept at −80 °C.

   (b) For water extraction, add 30 µL ultrapure water to the frozen tissue samples and let thaw on ice. For nervous system samples, sonicate the tissue in a cold ultrasonic water bath for 1 min until the tissue is largely disintegrated (*see* **Note 8**). Then, for all tissues, boil immediately for 10 min, followed by cooling for 5 min on ice. Then add 0.3 µL 1 mM HCl and centrifuge at maximum speed. Inject the supernatant immediately into an LC system, or transfer the supernatant to a fresh tube and store at −80 °C (*see* **Note 10**).

### 3.4 LC-MS

We here describe the conditions optimized for our LC-MS systems as a guide line.

#### 3.4.1 capRP-HPLC and MALDI-TOF/TOF MS

1. Dissolve the dried sample in 30 µL eluent A in an ultrasonic bath for 15 min and then centrifuge at 13,000 × *g* for 15 min.
2. Load the sample onto an RP C18 trap column with eluent A at a flow rate of 20 µL/min.
3. Switch the flow through the trap column and an analytical RP column with a rate of 2 µL/min. Then run a linear gradient from 4% to 60% eluent B over 30 min.
4. Use the spotter to mix 1 µL sample fraction with 1 µL of half-saturated alpha-cyano-4-hydroxycinnamic acid matrix solution and spot for 30 s/spot onto a stainless steel MALDI target plate.
5. Analyze the samples by MALDI-TOF MS/MS. A mass window of 800–2500 Da will comprise the majority of detectable peptides.

#### 3.4.2 LC-ESI-MS

1. Dissolve the dried sample in 30 µL eluent A in an ultrasonic bath for 15 min and then centrifuge at 13,000 × *g* for 15 min (*see* **Note 11**).

2. Load the peptide extract onto a trapping column and then switch to a separation column with a 30 min linear gradient from 3% to 30% acetonitrile and 0.1% formic acid at a flow rate of 200 nL/min.
3. Acquire MS scans.

## 4 Notes

1. To obtain a sufficiently high labeling rate in flies, labeling efficiency for the ammonium sulfate should be >98%.
2. There are different brands (e.g., Sylgard®, Dow Corning, Midland, MI) of silicon polymers. We use Elastosil® (Wacker Chemie, Munich, Germany) as follows: Mix components A and B (9:1) and pour into transparent plastic dishes (300 mL is enough for ten dishes Ø 5.5 cm; 45 mL is enough for 20 dishes Ø 3.5 cm). Then place the dishes on a heat plate at 70–90 °C for 1 h, and let stand overnight at room temperature. If a black background is preferred for dissection, grind up charcoal tablets and mix the powder with the silicon polymer in the dish before putting the dish onto the heat plate.
3. To clean the tubes from methanol-soluble compound, we prewash them with 90% HPLC-grade methanol and let them air-dry.
4. We typically use LC offline-coupled to MALDI-TOF mass spectrometry. Although today increasingly displaced by ESI-MS for peptidomic studies, MALDI-TOF MS holds several advantages: (a) only single-charged ions occur and peptide peak intensities can intuitively be quantified and checked manually from the mass spectrogram as labeled and unlabeled fractionated peptides appear in the same mass spectrogram (Fig. 2); (b) variability in labeling efficiency can be compensated to a substantial degree (Fig. 2). Peptide quantification is also possible with any other LC-MS systems. For example, we have successfully used ESI-MS (LTQ Orbitrap Velos Pro with EASY-Spray Ion Source coupled to an EASY-nLC 1000 (Thermo Scientific), equipped with a PepMap C18 3 μm particles, 100 Å pore size trap column and an EASY-Spray separation column (25 cm × 75 μm ID, PepMap C18 2 μm particles, 100 Å pore size) to quantify peptides, but here it is critical that the software algorithms quantify only the processed bioactive peptides and do not integrate over all peptides originating from propeptide processing.
5. Though chemically identical, $^{15}N$ could have effects especially when situated in a catalytic center of enzymes. In our laboratory, $^{15}N$ labeled larvae and flies are on average somewhat smaller than $^{14}N$-labeled flies. Therefore, it is important to

perform replicates in which the control and experimental groups are labeled with the opposite combination of $^{14}N$- and $^{15}N$-isotopes.

6. If additional water supply or humidity is needed, add moistened nitrogen-poor analysis filter paper (e.g., MN 620 from Macherey-Nagel, Düren, Germany) but not normal filter or other paper.
7. Alternatively, add $^{14}N$ or $^{15}N$ "yeast marmalade": 3.5 g yeast, 0.9 g sucrose, 7 mL double distilled water, 84 μL 10% nipagine dissolved in ethanol, 0.2 g pectin (biochemical grade or better). Yeast marmalade gives an improved survival rate especially in starvation experiments, but also leads to a reduced labeling efficiency due to nitrogen contaminations of pectin. This is little problematic if MALDI-TOF MS is used (*see* Fig. 2), but presents a problem when quantifying with other LC-MS methods.
8. Keep the water bath in a cold room, or add ice to the water.
9. A complete dry-down results in larger peptide loss than if the sample is only concentrated to 1–2 μL. Methanol predominantly evaporates first.
10. We do not recommend drying down as this always results in peptide loss.
11. To reduce clogging of the nano-columns, samples can be cleaned-up prior to injection using StageTips [10].

## Acknowledgments

We thank Gertrud Gramlich for excellent technical help, and Jörg Kahnt (MS facility, Max-Planck-Institute of Terrestrial Microbiology, Marburg, Germany), Jens T. Vanselow, and Andreas Schlosser (MS facility, Virchow Center, University of Würzburg, Germany) for MS measurements and helpful discussions. Original work was supported by the Deutsche Forschungsgemeinschaft (DFG WE 2652/4-2, to CW) and The Wellcome Trust/DBT India Alliance Early Career Fellowship (Award # IA/E/12/1/500742, to Megha).

### References

1. Pauls D, Chen J, Reiher W, Vanselow JT, Schlosser A, Kahnt J, Wegener C (2014) Peptidomics and processing of regulatory peptides in the fruit fly Drosophila. EuPA Open Proteom 3:114–127. https://doi.org/10.1016/j.euprot.2014.02.007
2. Fastner S, Predel R, Kahnt J, Schachtner J, Wegener C (2007) A simple purification protocol for the detection of peptide hormones in the hemolymph of individual insects by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. Rapid

Commun Mass Spectrom 21:23–28. https://doi.org/10.1002/rcm.2800

3. Zeng Q, Avilov V, Shippy SA (2016) Prefractionation methods for individual adult fruit fly hemolymph proteomic analysis. J Chromatogr B Analyt Technol Biomed Life Sci 1015–1016:74–81. https://doi.org/10.1016/j.jchromb.2016.02.014
4. Gouw JW, Tops BBJ, Krijgsveld J (2011) Metabolic labeling of model organisms using heavy nitrogen (15N). Methods Mol Biol Clifton NJ 753:29–42. https://doi.org/10.1007/978-1-61779-148-2_2
5. Jiménez C, ter Maat A, Pienemann A, Burlingame A, Smit AB, Li K (2004) Spatio-temporal dynamics of egg-laying-inducing peptides during an egg-laying cycle: a semi-quantitative matrix-assisted laser desorption/ionization mass spectrometry approach. J Neurochem 89:865–875
6. Jiménez CR, Li KW, Dreisewerd K, Mansvelder HD, Brussaard AB, Reinhold BB, Van der Schors RC, Karas M, Hillenkamp F, Burbach JP, Costello CE, Geraerts WP (1997) Pattern changes of pituitary peptides in rat after salt-loading as detected by means of direct, semiquantitative mass spectrometric profiling. Proc Natl Acad Sci U S A 94:9481–9486
7. Wegener C, Herbert H, Kahnt J, Bender M, Rhea JM (2011) Deficiency of prohormone convertase dPC2 (AMONTILLADO) results in impaired production of bioactive neuropeptide hormones in Drosophila. J Neurochem 118:581–595. https://doi.org/10.1111/j.1471-4159.2010.07130.x
8. Feng Y, Ueda A, C-F W (2004) A modified minimal hemolymph-like solution, HL3.1, for physiological recordings at the neuromuscular junctions of normal and mutant Drosophila larvae. J Neurogenet 18:377–402. https://doi.org/10.1080/01677060490894522
9. Jan LY, Jan YN (1976) Properties of the larval neuromuscular junction in Drosophila melanogaster. J Physiol 262:189–214
10. Rappsilber J, Mann M, Ishihama Y (2007) Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. Nat Protoc 2:1896–1906. https://doi.org/10.1038/nprot.2007.261

# Chapter 12

# Data Preprocessing, Visualization, and Statistical Analyses of Nontargeted Peptidomics Data from MALDI-MS

**Harald Tammen and Rüdiger Hess**

## Abstract

Mass spectrometric (MS) comparative analysis of peptides in biological specimens (nontargeted peptidomics) can result in large amounts of data due to chromatographic separation of a multitude of samples and subsequent MS analysis of numerous chromatographic fractions. Efficient yet effective strategies are needed to obtain relevant information. Combining visual and numerical data analysis offers a suitable approach to retrieve information and to filter data for significant differences as targets for succeeding MS/MS identifications.

Visual analysis allows assessing features within a spatial context. Specific patterns are easily recognizable by the human eye. For example, derivatives representing modified forms of signals present are easily identifiable due to an apparent shift in mass and chromatographic retention times. On the other hand numerical data analysis offers the possibility to optimize spectra and to perform high-throughput calculations. A useful tool for such calculations is R, a freely available language and environment for statistical computing. R can be extended via packages to enable functionalities like mzML (open mass spectrometric data format) import and processing. R is capable of parallel processing enabling faster computation using the power of multicore systems.

The combination and interplay of both approaches allows evaluating the data in a holistic way, thus helping the researcher to better understand data and experimental outcomes.

**Key words** MALDI, Mass spectrometry, R, Data analysis, Parallel computing, MALDIQuant, Peptidomics, Statistics

## 1 Introduction

Analysis of biological samples via mass spectrometry is a valuable tool to detect differences between peptide levels in biological samples. These differences can serve as a starting point for biomarker discovery or to determine effects related to specific experimental settings (e.g., pharmaco-dynamic studies). Typical studies may be comprised of 20–100 samples. After separation of samples via liquid chromatography (LC) and subsequent mass spectrometric (MS) measurements thousands of MS spectra need to be statistically analyzed to extract relevant information. The aim of data analyses is to reveal significant

Michael Schrader and Lloyd Fricker (eds.), *Peptidomics: Methods and Strategies*, Methods in Molecular Biology, vol. 1719, https://doi.org/10.1007/978-1-4939-7537-2_12, 

differences in mass spectrometric data that are the base for interpretation of given effects and provide guidance for sequencing efforts. Prior to data analysis a preprocessing of data is beneficial. Data preprocessing may involve data conversion (conversion of proprietary data formats to standard formats), baseline corrections and smoothing procedures, *m/z*-recalibrations of mass spectrometric data, data reduction (binning) to enable fast processing without losing relevant information, and peak recognition (feature detection).

To determine differences, analyses based on visualized mass spectrometric data and exported signal intensities of detected features can be carried out. The visualization of mass spectrometric data is beneficial to comprehensively assess data to determine various aspects of the measurement. It allows for a wide-ranging assessment of features within a spatial context (the distribution of signals based on hydrophobicity and molecular mass) and an assessment of interrelations between signals. For example, conspicuous features like polymeric components or chromatographic elution shifts are easily recognizable. On the other hand univariate and multivariate analyses applied to extracted peak/intensity values from spectra complement the visual inspection since these procedures allow for high-throughput determination of statistically different signals. Both types of analyses are evaluated for consistency and serve as a starting point for data interpretation and subsequent sequencing efforts.

## 2 Materials

### 2.1 Data Conversion

To convert data from a vendor-specific format to an open standard format like mzML or mzXML various tools are available. On the one hand vendors may provide suitable tools (e.g., Bruker's CompassXport) or open-source projects like Proteowizard [1] (http://proteowizard.sourceforge.net/downloads.shtml) or T2D converter/raw2MS (pepchem.org) [2] can accomplish this task. It is mandatory to convert the file format in order to process data in programs not capable of reading vendor's proprietary file formats.

### 2.2 Data Preprocessing, Peak Extraction, and Statistical Analysis

For comprehensive data analysis of mzML/mzXML files R [3], a freely available language and environment for statistical computing (www.r-project.org), is a suitable tool. R is highly extensible via additional packages to implement additional functionality. The following R-packages are suggested for processing of mass spectrometric data:

1. Parallel (allows for parallel computation on UNIX systems).
2. pROC (calculation of receiver-operating characteristics) [4].
3. stringr (string manipulation) [5].

4. MALDIquantForeign (import and export of mass spectrometric data into R) [6].
5. MALDIQUANT (analysis of mzML/mzXML files) [6].
6. RMySQL (integration of MySQL databases) [7].

In addition to R an IDE (integrated development environment) is recommended (e.g. rstudio.com) [8]. This provides a user interface to R, facilitates the interaction with R, has built-in source control, and offers the possibility for server installation.

### 2.3 Data Visualization

Data visualization communicates a quantitative message. It might refer to presentation of statistical results (e.g., box-and-whisker plots) or raw data to make complex data more accessible. Within the scope of this article visualization refers to display of mass spectrometric data. Visualization of MALDI mass spectra is important to mine basic structures in the data. For visualization of complex 2D LC-MALDI-MS datasets, a two-dimensional (chromatographic elution and molecular mass) map view has been introduced resembling the picture of 2D gel electrophoresis [9]. It allows assessing high density information in a natural and convenient way. Often spectra are ordered by the chromatographic retention time ($y$-coordinate); thus spectra are sorted by hydrophobicity. The $m/z$ values are represented by the $x$-coordinate whereas the intensity is reflected by a $z$-scale (coloring). Applications that allow for comprehensive visualization of mass spectrometric data include MSight (www.web.expasy.org/MSight/) [10], mMass (www.mmass.org/), and Spectromania (www.spectromania.com).

## 3 Methods

The following chapter can provide guidance of analysis of MALDI-MS data, only. Due to the complexity of the subject, a detailed description would exceed the scope of this book. However by using examples the basic principles will be explained.

The analysis of MALDI-MS data basically consists of the following steps:

### 3.1 Preprocessing of MS Data

The preprocessing of raw MS data consists of conversion steps, smoothing operations, $m/z$ re-alignments, and data reduction to prepare spectra for peak detection and visualization.

#### 3.1.1 Conversion of Raw Data to mzML

To convert MS data from propriety file format to mzML data [11] vendor provided or third-party software solutions (cf. Chapter 2) can be used. The conversion is necessary for further processing of mass spectra.

### 3.1.2 Import mzML Files into R

For optimization of mass spectrometric data prior to peak detection and export, spectral data can be imported into R. Therefore the R-package MALDIquantForeign can be used. To import data use the following commands in R:

```
library("MALDIquantForeign") #load and attach add-on packages.
spectra = importMzMl(pathToFile) #load mzML File and store it in the variable spectra
```

An mzML file can contain single or multiple spectra, e.g., from one sample after chromatographic separation and subsequent MS measurement of individual fractions. After import, spectral data can be assessed using the @-notation:

```
massValues = spectra[[1]]@mass #assess mass values of first spectrum
intensityValues = spectra[[1]]@intensity #assess intensity values of first spectrum
metadata = spectra[[1]]@metaData #assess metadata like file information or ms level
```

The numeral within the square brackets refers to spectrum number within the mzML file.

### 3.1.3 Optimize Spectra (Baseline Correction, Mass Recalibration, and Optional Intensity Calibration)

The optimization of spectra includes baseline correction, *m/z* re-alignment to a reference spectrum, smoothing, and optionally intensity calibration.

The corresponding functions in the R-package MALDIQUANT are:

```
removeBaseline() # for baseline correction
transformIntensity()/calibrateIntensity() #for normalizing intensities of a mass spectrum
smoothIntensity() #for smoothing intensity using Savitzky-Golay or Moving Average
```

To align mass spectra (phase correction) a warping has to be determined and applied. The functions *determineWarpingFunctions()* and *warpMassSpectra()* can achieve the re-alignment.

Since proteomics experiments usually provide hundreds to thousands of individual mass spectra, it is advisable to use parallel processing on multicore systems. Unix-based systems are capable of providing this functionality out of the box. Parallelization (*see* **Note 1**) vastly reduces processing times. For Unix-based systems, the parallels package can be used:

```
library(parallel) #load and attach add-on packages.
resultList = mclapply(c(1:length(spectra)),functionName, mc.cores = 4)
```

This function (mclapply) applies a given function (functionName) for each spectrum using four processor cores and stores results in a list (resultList).

### 3.1.4 Rebinning of Spectra for Data Reduction

It might be beneficial to reduce the mass spectrometric data by binning the data to a suitable value. In MALDI-MS for peptidomics analyses a value of 1 Dalton can be regarded as sufficient, since the isotopic resolution significantly decreases with $m/z$ ratio of larger 2000 for spectra acquired in linear mode. The binning of data significantly improves visualization speeds due to data reduction by more than 90% without losing relevant information. Figure 1 shows the same mass spectrum before (1) and after (2) processing. The lower part of the figure shows the transformation of both spectra into a 2D-gel like view. Spectrum 1 consists of 190,028 data points and spectrum 2 of 13,951 data points. The numerals below the spectra refer to the signal-to-noise ratios of selected signals (SNR > 7). The processed spectrum exhibits better signal-to-noise ratios yet consists of 1/10 of data points facilitating subsequent processing.

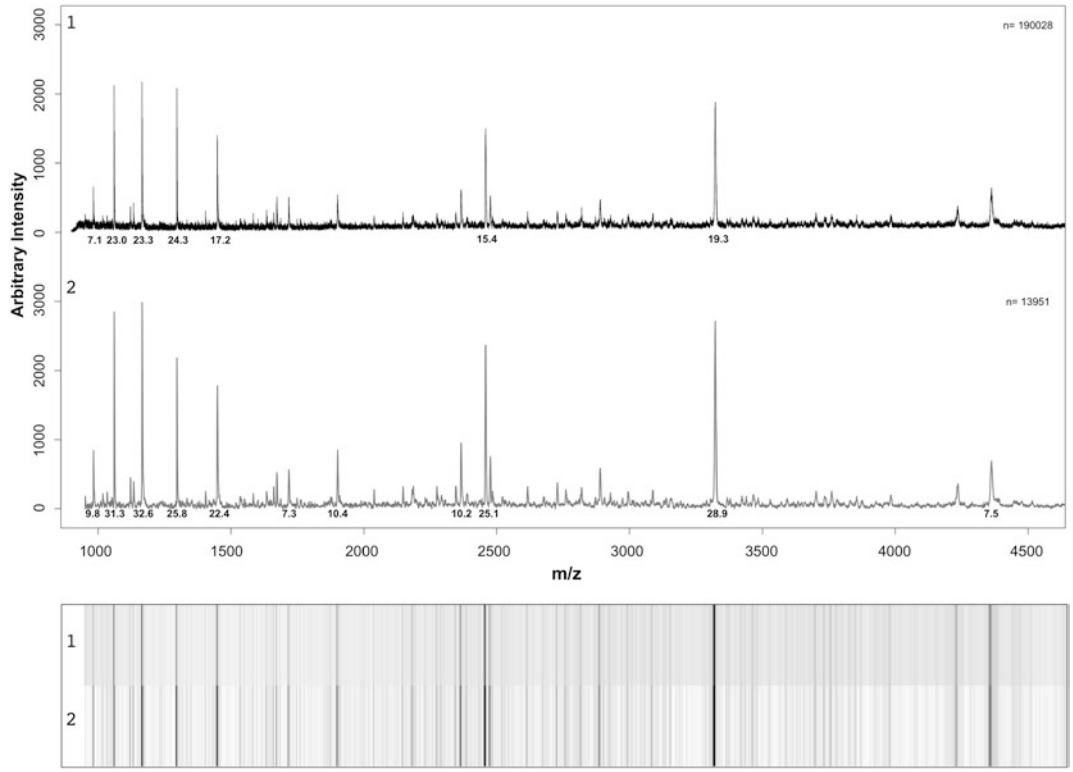



**Fig. 1** The figure depicts the very same mass spectrometric data before (1) and after (2) data reduction as graph views (above) and 2D gel like views (below). The numerals below peaks refer to the signal-to-noise ratios of corresponding signals. Spectrum 1 consists of 190,028 data points whereas spectrum 2 consists of 13,951 data points. To obtain the optimized spectra, a baseline correction, smoothing, and rebinning operation were applied to the original spectrum

### 3.2 Peak Detection and Statistical Analysis of Exported Signal Intensities

The objective for most analyses is to detect specific significant differences between datasets, i.e., signals that are regulated (increased or decreased signal intensities) in dependence of sample metadata (descriptive characteristics like classification of samples). Therefore, signals have to be defined and the corresponding intensities extracted from the preprocessed mass spectrum. Subsequently, peak data and adjacent experimental metadata can be subjected to statistical analysis.

#### 3.2.1 Peak Detection

For peak detection, associated mass spectra are averaged using the function *averageMassSpectra()*. Subsequently, a peak detection *using detectPeaks()* can be performed. After determination of peaks, their characteristics (signal-to-noise ratio, intensity values) can be stored and subjected to statistical analysis.

#### 3.2.2 Import of Peak Data and Metadata into R

Various statistical tests can be performed in R. As a prerequisite a description of samples is required that lists the sample names and provides metadata like groups or other classifiers (*see* **Note 2**). Metadata can be imported from an SQL database (e.g., mySQL (*see* **Note 3**)) or tab-delimited file (*see* **Notes 4** and **5**).

#### 3.2.3 Univariate and Multivariate Analyses in R to Reveal Differences

In R both datasets (signal data and metadata) are used to perform the appropriate univariate statistical calculations like *t*-tests (*see* **Note 6**). In the following example peak data and metadata are stored in two separate data frames (SI and metadata) that are comprised of rows and columns. The rows (observations) represent the sample names whereas the columns (variables) represent the peak or metadata variable. To assess an item stored in a data frame the *[row-number, column-number]* notation can be used. For example to calculate the *p*-value using a *t*-test the following function can be applied:

```
resultTTest = t.test(SI[,1]~metadata$Classifier, var.equal = TRUE,
    alternative = "g")
pValue = resultTTest$p.value #store the p-value of the t-test in the
    variable pValue
```

For each variable, appropriate statistical tests (e.g., statistical significance of discriminatory power, comparison of SNR ratios, correlations, or receiver-operating characteristics (*see* **Note 7**)) can be performed. This process can be parallelized leading to fast processing of a high number of variables. Besides univariate tests to unravel specific signals that match certain criteria also multivariate analyses can be carried out. One well-established multivariate method is the classical principal component analysis (PCA), where the information is represented by a new set of latent variables, termed components. These components summarize variances within a dataset and are especially useful to detect global biases. Biases might be related to technical procedures (e.g., chromatographic shift) or might represent an experimental parameter with

significant impact on sample composition (e.g., plasma vs. serum samples, [12]). It is important to understand that the PCA primarily detects effects that affects a high number of variables and therefore is prone to unravel biases within datasets. To calculate the principal components the build-in function *prcomp()* can be used. Prior to calculating the PCA, signal data with no variations (signal intensities with a standard deviation of zero) have to be removed:

```
if(length(which(as.numeric(apply(SIselect,2,sd))==0))) SI = SI[,-which
    (apply(X = SI,MARGIN = 2,FUN = sd) == 0)] #remove columns
    with a standard deviation of zero from the data.frame SI.
PCAresults = prcomp(x = log(SI), center = TRUE, scale = TRUE)#
    calculate PCA
Scores = PCAresults$x # retrieve the scores (principal components)
```

The scores of the PCA can be plotted against each other (e.g., PC1 vs. PC2) to visualize sample clustering or individual scores can be analyzed by univariate methods to determine potential underlying causations for observed variations. The interpretation of principal component analyses can sometimes be challenging but it is a very important tool for data analysis.

## 3.3 Visual Data Analysis

Visual analysis of mass spectrometric data can deliver important clues for data interpretation since the human brain possesses a strong aptitude for pattern recognition thus enabling the researcher to observe effects present in datasets that might otherwise be difficult to determine using uni- or multivariate methods.

### 3.3.1 Visual Inspection of Mass Spectrometric Data

The visualization of mass spectrometric data is beneficial to comprehensively assess data to determine various aspects of the measurement. For example, the distribution and intensity of signals is accessible and conspicuous features like polymeric components are recognizable. The visualization allows for wide-ranging assessment of signals within a spatial context (e.g., the distribution of signals based on fraction and mass) and interrelations between signals like post-translational modifications become apparent (cf. Subheading 3.2).

It is also recommended to verify results from numerical analysis (*see* Subheading 3.2.3) by evaluating corresponding mass spectrometric data. This can be achieved by grouping all mass spectrometric data from one chromatographic fraction (same fraction of each individual sample) to assess the regulation of a specific signal.

In order to illustrate possibilities for data inspection by visualization, two examples are presented and detailed.

### 3.3.2 Visual Assessment of MS Data

Figure 2a depicts mass spectrometric data obtained from cerebrospinal fluid. In the upper part five chromatographic fractions (fractions 26–30 corresponding to 16.2–17.0% Acetonitrile) in the mass range from 6200 to 7000 Dalton are depicted. The mass spectra are

processed as depicted in Fig. 1. A main signal cluster around 6.5 kDa is visible. The spectral trace in the lower part shows the summed spectra from fraction 27–29. Letters A and B mark two signals with an $m/z$ difference of 30 Da. A1 and B1 mark signals with a mass difference of 80 Dalton to A and B, respectively. The small table on the right side depicts the intensity of signal A and B in fraction 28 and 29. Signal B corresponds to human secretogranin-1 (UniProt accession number P05060) representing amino acid 217–275. Signal A represents the same peptide but with an amino acid exchange (T243A, rs236151). Signal A1 and B1 represent phosphorylated forms of the same peptides. The amino acid exchange leads to a shift in $m/z$ of 30 Dalton that corresponds to the difference between Threonine and Alanine and a chromatographic shift due to the higher hydrophobicity of Threonine due to change from medium size and polar (T) to small size and hydrophobic (A). The chromatographic shift is depicted by the intensity values. However the sum of intensities from both fractions for each signal is equal since both alleles contribute equally to the expression of the precursor protein. Additionally the phosphorylation leads to a mass shift of 80 Da and a more hydrophilic elution pattern.

### 3.3.3 *Visual Assessment of Statistical Values*

In contrast to direct representation of mass spectrometric data, it is also possible to first calculate statistical properties and subsequently visualize results in a spatial context. Figure 2b shows visualized correlation values [13] in the fraction range from 49 to 53 and

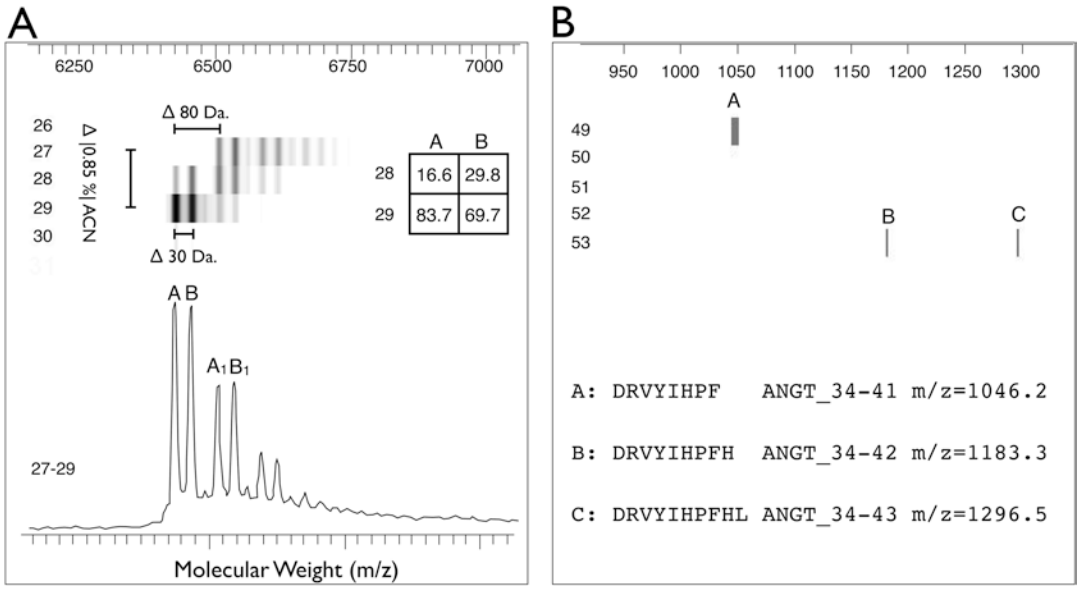



**Fig. 2** (**a**) The figure shows the detection of a polymorphism (T243A) of secretogranin-1 (aa 217–275) in cerebrospinal fluid. The figure depicts visualized mass spectrometric data and its relation to the chromatographic elution profile. A detailed description is given in Subheading 3.2. (**b**) The figure shows the detection of Angiotensin-2 and its derivatives. The figure is based on calculated correlation values obtained from mass spectrometric analysis of 28 plasma samples. A detailed description is provided in Subheading 3.3

mass range from 950 to 1300 Dalton. The correlation was calculated based on the intensity of signal A (Angiotensin II, P01019) from 28 samples. Intensities with correlating intensities (threshold 0.8) are visible (Signal B and C). Based on the amino acid sequence of Signal A, the corresponding sequences of both Signal B and C can be predicted with high probability. The mass differences and chromatographic shifts indicate that both signals correspond to longer variants of Signal A.

Both examples demonstrate that careful visual inspection of mass spectrometric data can yield valuable information for data interpretation and may aid in sequence determination.

## 4 Notes

1. For parallel processing (Parallel Package) it is advisable to use the number of physical cores and not the number of virtual cores since a performance hit can otherwise occur. To determine the number of cores the function *detectCores()* can be used.
2. To select data in R it is possible to use the %in% notation. For example to select items from one list based on properties stored in a second list use the following command:

   *list1[which(list1$property1%in% list2$property2)]*
3. The type of a field of a mySQL database should represent the type used in R; otherwise errors can occur.
4. For performance reasons the number of columns (variables) should always be greater than the number of rows. If necessary, use the transpose function: *t()*.
5. It is advantageous to always use tab-delimited text files instead of comma-separated text files to ensure international compatibility (in some regions commas serve as delimiter for decimal numbers).
6. When using Welch Two Sample *t*-test, it is advisable to test both alternatives (less and greater) separately and not two-sided. This is related to the signal dynamic of mass spectrometry.
7. The program R-base does not offer functions for calculation of receiver-operating characteristics. However the pROC package adds this functionality.

### References

1. Chambers MC, Maclean B, Burke R, Amodei D et al (2012) A cross-platform toolkit for mass spectrometry and proteomics. Nat Biotechnol 30:918–920. https://doi.org/10.1038/nbt.2377
2. T2d/RAW converter. http://www.scripps.edu/new_yates/sites/ygao//index.html
3. R Core Team (2016) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/
4. Robin X, Turck N, Hainard A et al (2011) pROC: an open-source package for R and S+ to analyze and compare ROC curves.

BMC Bioinform 12:77. https://doi.org/10.1186/1471-2105-12-77

5. Wickham H (2015) stringr: simple, consistent wrappers for common string operations. R package version 1.0.0. https://CRAN.R-project.org/package=stringr
6. Gibb S, Strimmer K (2012) MALDIquant: a versatile R package for the analysis of mass spectrometry data. Bioinformatics 28: 2270–2271
7. Ooms J, James D, DebRoy S et al. (2016). RMySQL: Database Interface and 'MySQL' Driver for R. https://CRAN.R-project.org/package=RMySQL
8. RStudio Team (2015) RStudio: Integrated Development for R. RStudio, Inc., Boston, MA. http://www.rstudio.com/
9. Zucht HD, Lamerz J, Khamenia V et al (2005) Datamining methodology for LC-MALDI-MS based peptide profiling. Comb Chem High Throughput Screen 8:717–723
10. Palagi PM, Walther D, Quadroni M et al (2005) MSight: an image analysis software for liquid chromatography-mass spectrometry. Proteomics 5:2381–2384
11. Deutsch EW (2010) Mass spectrometer output file format mzML. Methods Mol Biol (Clifton, NJ) 604:319–331. https://doi.org/10.1007/978-1-60761-444-9_22
12. Tammen H, Schulte I, Hess R et al (2005) Peptidomic analysis of human blood specimens: comparison between plasma specimens and serum by differential peptide display. Proteomics 5:3414–3422
13. Lamerz J, Crameri R, Scapozza L, Mohring T et al (2005) Peptide sequence prediction supported by correlation-associated networks in human cerebrospinal fluid. Comb Chem High Throughput Screen 8:789–799

# Part III

## Diverse Applied Protocols and Outlook

# Chapter 13

# Affinity Purification of Neuropeptide Precursors from Mice Lacking Carboxypeptidase E Activity

Lloyd Fricker

## Abstract

Peptidomic techniques are powerful tools to identify peptides in a biological sample. This protocol describes a targeted peptidomic approach that uses affinity chromatography to purify peptides that are substrates of carboxypeptidase E (CPE), an enzyme present in the secretory pathway of neuroendocrine cells. Many CPE products function as neuropeptides and/or peptide hormones, and therefore represent an important subset of the peptidome. Because CPE removes C-terminal Lys and Arg residues from peptide-processing intermediates, organisms lacking CPE show a large decrease in the levels of the mature forms of most neuropeptides and peptide hormones, and a very large increase in the levels of the processing intermediates that contain C-terminal Lys and/or Arg (i.e., the CPE substrates). These CPE substrates can be purified on an anhydrotrypsin-agarose affinity resin, which specifically binds peptides with C-terminal basic residues. Not all peptides with basic C-terminal residues within a cell are CPE substrates, and these other peptides will also be purified on the anhydrotrypsin affinity column. However, a comparison of peptides purified from wild-type mice and from mice lacking CPE allows for the rapid identification of CPE substrates based on their large increase in the absence of CPE.

**Key words** Anhydrotrypsin, Affinity chromatography, Carboxypeptidase E, Neuropeptide

## 1 Introduction

Neuropeptides and peptide hormones are the largest group of intercellular chemical messengers and are found in all multicellular organisms that have been studied [1]. Even *Saccharomyces cerevisiae* (yeast) produce peptides that signal between cells and serve as mating factors [2]. The first peptides involved in intercellular signaling were discovered over 100 years ago using bioassays to detect the peptides and follow their activities during purification, which often took many years [3]. The development of mass spectrometry-based peptidomics techniques eliminated the need to purify peptides to homogeneity prior to sequencing and led to a rapid expansion of the number of identified peptides [4–6]. However, many of the peptides in tissue extracts do not represent neuropeptides, and instead are proteasomal products [6, 7].

Michael Schrader and Lloyd Fricker (eds.), *Peptidomics: Methods and Strategies*, Methods in Molecular Biology, vol. 1719, https://doi.org/10.1007/978-1-4939-7537-2_13, 

Although some of these cellular peptides may be functional and affect cellular processes [8], there is a need for targeted studies that are focused on peptides that primarily function in intercellular signaling. The protocol described in this chapter describes a one-step approach to enrich for precursors of peptides that function in cell-cell signaling.

The production of most neuropeptides and peptide hormones involves the selective cleavage of a precursor, usually a small protein of ~200–300 amino acids, but on occasion a larger protein [1]. The precursor is targeted for the secretory pathway by the presence of an N-terminal signal peptide that directs the nascent protein to the rough endoplasmic reticulum where the protein is translocated into the lumen during translation. The signal peptide is removed in the endoplasmic reticulum and the resulting protein (termed a prohormone or proneuropeptide) is transported through the Golgi to the trans Golgi network where it is sorted into secretory vesicles (Fig. 1). Proteolytic cleavage of the prohormone begins in either the late Golgi or the immature secretory vesicles, depending on the precursor [1]. Most precursors are cleaved by both endo- and exo-peptidases [1]. The endopeptidases cleave to the C-terminal side of basic residues (Lys, Arg), typically producing several peptide-processing intermediates containing C-terminal extensions, and one peptide that represents the C-terminal portion of the precursor (this C-terminal peptide does not usually have Lys or Arg extensions). The peptide-processing intermediates are substrates for a carboxypeptidase that removes the C-terminal basic residues [1, 9]. The carboxypeptidase is the final processing step for many bioactive peptides, although some peptides require additional modifications before the peptide is bioactive. One common additional modification is the formation of a C-terminal amide residue, which is enzymatically produced if the peptide contains a C-terminal Gly residue [10].

Peptidases in the trans Golgi include endopeptidases such as furin and other furin-like enzymes, and the exopeptidase carboxypeptidase D (CPD) which removes the basic residues [1]. However, furin and CPD are thought to be responsible for only a minor fraction of the cleavages, and the vast majority of the endopeptidase cleavages occur in maturing secretory vesicles, mediated by prohormone convertase 1 (PC1, also known as PC3) and prohormone convertase 2 (PC2) [1]. A single carboxypeptidase, CPE, is present in the mature secretory vesicles [1]. Prior to the discovery of CPD, it was predicted that the absence of CPE would be lethal to mice due to the large numbers of peptides that require removal of basic residues for their activity, and the broad role of these peptides in many physiological functions. However, in 1995 we found mice with the *fat* mutation, a naturally occurring mutation discovered in the early 1970s, had a point mutation in the coding region of CPE that changed a Ser into a Pro, causing the enzyme to be inactive [11, 12].

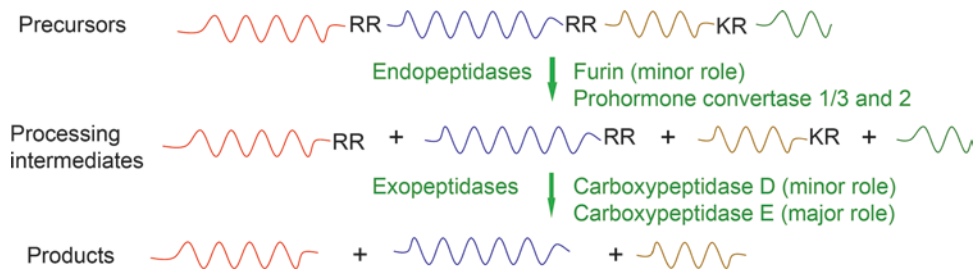



**Fig. 1** General scheme for the production of neuropeptides and peptide hormones from their precursor proteins. A representative example is shown in which four distinct peptide products are formed from a single precursor. Some precursors produce only two peptides, while others produce dozens of peptides. Typical cleavage sites are basic amino acids lysine (K) and arginine (R), most commonly RR and KR, and on occasion RK, KK, and single basic residues. Endopeptidases cleave the precursor to the C-terminal side of these basic residues, forming peptide-processing intermediates that contain C-terminal basic residues. These basic residues are then efficiently removed by a carboxypeptidase (unless the penultimate residue is a proline, at which point the peptide is not efficiently processed by the carboxypeptidase). The primary peptide-producing endopeptidases are prohormone convertases 1/3 (PC1/3) and 2; these enzymes are present in the secretory vesicles. In addition to these, some peptide precursors are processed in endopeptidases that are present in the trans Golgi network such as furin and furin-like enzymes. Carboxypeptidase D (CPD) is present in the trans Golgi network and carboxypeptidase E (CPE) is enriched in the secretory vesicles

The mutation was subsequently renamed *Cpe*$^{fat}$. Mice homozygous for this mutation (i.e., *Cpe*$^{fat/fat}$ mice) are obese, infertile, and show depressive-like and anxiety-like behaviors [12, 13]. Peptidomic analysis using a quantitative approach to compare peptides in various brain regions of wild-type and *Cpe*$^{fat/fat}$ mice found that CPE is responsible for the production of >90% of the neuropeptides detected in the analysis [14]. Peptides unaffected by the absence of CPE activity were not CPE substrates. This latter group included neuropeptides that corresponded to the C-terminal fragments of their precursors (Fig. 1), as well as numerous peptides derived from cytosolic proteins [14].

This chapter describes the one-step purification of peptides with Lys and/or Arg on the C-terminus. When used with extracts of tissue from *Cpe*$^{fat/fat}$ mice, this approach allows for the isolation of all peptide-processing intermediates that are CPE substrates (Fig. 2). The vast majority of cytosolic proteasome fragments (or other cellular peptides) do not contain C-terminal basic residues, and these will not be enriched by this procedure (Fig. 2). A small fraction of cytosolic peptides that contain C-terminal basic groups will be purified along with the neuropeptide intermediates, but these cytosolic peptides will also be present in wild-type mice. Therefore, a comparison of the spectra from LC/MS analysis of *Cpe*$^{fat/fat}$ and wild-type mice processed using this approach will reveal those peptides common to both genotypes (non-CPE substrates) and those peptides which are greatly enriched in the *Cpe*$^{fat/fat}$ mice (likely CPE substrates). Using this approach, we detected many known neuropeptides or predicted

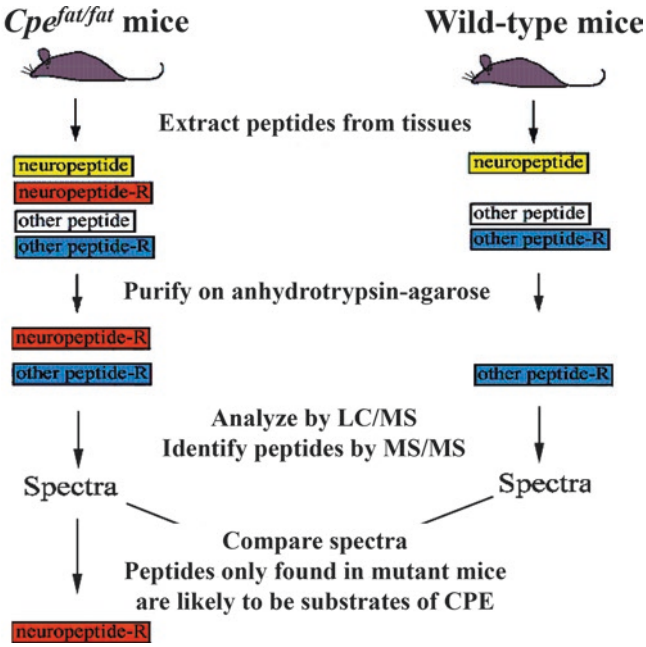



**Fig. 2** The *Cpe*$^{fat/fat}$ mice have a mis-sense mutation in which Ser in position 202 is replaced with Pro, causing the enzyme to misfold and be degraded. In the absence of CPE activity, peptide-processing intermediates with C-terminal Lys and/or Arg residues accumulate (only peptides with Arg, or R, are shown in the figure). These peptides are usually present in wild-type mice at extremely low levels, less than 1% of the levels of the mature forms lacking the basic residues (unless the penultimate residue is Pro, which reduces the activity of CPE for these peptides and results in high levels of the Lys- or Arg-containing peptides). In addition to neuropeptides, tissue extracts contain other peptides such as proteasome products; most of these do not contain C-terminal basic residues, but some of these other peptides contain C-terminal Lys or Arg residues (indicated in figure as "other peptide-R"). The levels of these other cellular peptides are generally similar in *Cpe*$^{fat/fat}$ and wild-type mice. Extracts of *Cpe*$^{fat/fat}$ and wild-type mice are purified on anhydrotrypsin-agarose columns, which enriches peptides with C-terminal Lys and/or Arg residues. These are analyzed by MALDI (not shown) and LC/MS, with tandem mass spectrometry (MS/MS) to determine their sequence. A comparison of signal intensities in spectra from *Cpe*$^{fat/fat}$ compared to wild-type mice reveals those peptides greatly enriched in the mutant mice, which are putative CPE substrates

fragments of neuropeptide precursors, thus validating the technique [15]. Several completely novel neuropeptides were detected; some of these peptides are major neuropeptides that are abundant in mouse brain [16]. In addition to being useful for *Cpe*$^{fat/fat}$ mice, this same strategy can be used to isolate CPE substrates in other organisms, such as the EGL-21 mutant of *Caenorhabditis elegans* (EGL-21 is the *C. elegans* ortholog of CPE) [17, 18]. In addition to purifying precursors of bioactive neuropeptides or peptide hormones, CPE also produces peptides that perform other functions, such as cytokines and antibacterial peptides [19], and this technique can potentially be used to isolate these types of peptides. Finally, one can use a similar strategy (albeit with change of affinity column) to capture substrates of other enzymes.

## 2 Materials

High purity water is used for all solutions (*see* **Note 1**).

### 2.1 Peptide Extraction

1. *Cpe*$^{fat/fat}$ mice (such as B6.HRS(BKS)-*Cpe*$^{fat}$/J from The Jackson Laboratory, Bar Harbor, ME, USA).
2. Wild-type mice (ideally littermates from breeding pairs of heterozygote *Cpe*$^{fat}$ mice; the homozygous mice are infertile, so it is necessary to breed heterozygote *Cpe*$^{fat}$ mice, which results in wild-type littermates).
3. Conventional microwave oven.
4. Dissecting scissors, spatula, and razor blades.
5. Microfuge tubes, low retention.
6. 0.1 M HCl solution.
7. Centrifuge.
8. Tissue disruptor or sonicator, such as the Polytron (Brinkman) or probe-type sonicator.

### 2.2 Affinity Purification

1. Anhydrotrypsin-agarose (*see* **Note 2**).
2. Small columns to hold anhydrotrypsin-agarose. We use Poly-Prep Chromatography Columns (BioRad) that can accommodate bed volumes of 0.2–2 mL and have a 10 mL reservoir, but columns from other sources should also be suitable.
3. Lyophilizer or vacuum centrifuge concentrator.
4. Microfiltration units with 10 kDa cut-off (such as Centriplus-10 membrane from Amicon).
5. Sodium acetate (NaAc) buffer, 50 mM and 100 mM, pH 5.0 (pH adjusted with acetic acid or sodium hydroxide).
6. 3-[(3-Cholamidopropyl)dimethylammonio]-1-propanesulfonate (CHAPS), 5% stock solution (to prepare 0.5% final).
7. $CaCl_2$ solution, 200 mM as a stock (to prepare 20 mM final).
8. 0.5 M NaCl in 0.1 M NaOH.

### 2.3 MS Analysis

1. MALDI-TOF-MS or LC/ESI-MS (various suppliers possible).
2. Software for database searches like Mascot or Peaks, but will depend on the programs available to your research group.

## 3 Methods

### 3.1 Stabilization of Tissue and Peptide Extraction

A number of different protocols have been used for the preparation of the tissue and extraction of peptides (*see* **Note 3**). The following describes a procedure that minimizes postmortem changes in neuropeptides and prevents the formation of additional peptides from degradation of cellular proteins.

1. Mice are sacrificed by cervical dislocation followed by decapitation. Heads are immediately microwaved in a conventional oven, using settings that will heat the brain to an internal temperature of 80 °C [20] (*see* **Note 4**).
2. After cooling to room temperature, the brain is removed from the scull and a razor blade is used to dissect out brain regions of interest (*see* **Note 5**).
3. Dissected brain regions are placed in microfuge tubes (*see* **Note 1**). If not proceeding with the extraction, the tissue can be stored at −80 °C for up to 1 year. If proceeding with **step 4**, place the tubes on ice to chill the brain regions.
4. Add 5 μL of cold water per μg of tissue (*see* **Note 1**) and homogenize the tissue in either a sonicator or tissue homogenizer. A minimum of 200 μL needs to be used for a typical sonicator, but this will depend on the device that is used (*see* **Note 6**).
5. Incubate the homogenates in a 70 °C water bath for 20 min and then cool on ice for 15 min.
6. Add ice-cold 0.1 M HCl to a final concentration of 10 mM HCl. Mix each tube and incubate on ice for 15 min.
7. Centrifuge the homogenates at 13,000 *g* for 40 min at 4 °C. Transfer the supernatants to new microfuge tubes and store at −80 °C if not proceeding to the next step.

### 3.2 Purification of Peptides

1. Filter the tissue extracts (from Subheading 3.1, **step 7**) through 10 kDa microfiltration devices (*see* **Note 1**). Use the material that passes through the filter (i.e., the peptides <10 kDa) for the next step.
2. Adjust pH to 5.0 by the addition of NaAc buffer (50 mM final).
3. Add the detergent CHAPS (0.5% final) and $CaCl_2$ solution (20 mM final) (*see* **Note 7**).
4. Apply the mixture containing peptides, buffer, CHAPS, and $CaCl_2$ to a column containing 0.5 mL of anhydrotrypsin-agarose previously equilibrated in the same buffer.
5. Collect the flow-through and pass it back through the column. This will increase the amount of peptide bound. This step can be repeated 2–3 times.

6. Wash the column with 5 mL of 0.5% CHAPS in 100 mM NaAc pH 5.0 buffer.
7. Wash the column with 5 mL of 10 mM NaAc buffer without detergent.
8. Elute the peptides from the column with 2 mL of water (*see* **Notes 1** and **8**).
9. Elute the peptides from the column with 4 mL of 5 mM HCl (*see* **Note 8**).
10. Freeze the eluates and remove the solvent in a vacuum concentrator.
11. The anhydrotrypsin-agarose resin is re-generated by washing with 1 mL of 0.5 M NaCl in 0.1 M NaOH followed by 10 mL of water, and 5 mL of 50 mM NaAc pH 5.0. Sodium azide (0.02%) can be added for long-term storage (be careful with sodium azide—it is very toxic).
12. After drying, the water and HCl elutes are resuspended in a small volume of water (typically 10–20 μL) and analyzed by MS (*see* **Note 9**).

### 3.3 MS Analysis

1. A number of different MS instruments can be used to analyze the peptides present in the eluates from the anhydrotrypsin-agarose column. Matrix-assisted laser desorption ionization time-of-flight (MALDI-TOF) will give a general idea of the most abundant peptides present in the sample and reveal if there are major differences between the peptides isolated from $Cpe^{fat/fat}$ mice and wild-type mice [15]. Liquid chromatography mass spectrometry (LC/MS) is useful to detect a larger number of peptides than can be detected by MALDI. We have used a variety of MS instruments and LC systems, and all have successfully identified many peptides using this approach [15, 16].
2. Determine which peptides are enriched in $Cpe^{fat/fat}$ mouse tissues by comparing spectra. There are multiple ways for this process. If using MALDI, it is simple to compare the spectra from each sample. For LC/MS data, the results can be divided into 1 min windows (or less if the samples are complex) and the ion profiles compared between the $Cpe^{fat/fat}$ and wild-type mouse extracts. Automated software can also be used with "label-free" quantitative approaches (see the chapter in this volume on quantitative techniques by Fricker).
3. Identify peptides from MS/MS data, using standard procedures. Typically, these involve database searches using programs like Mascot or Peaks, but will depend on the programs available to your research group.

## 4 Notes

1. Avoiding contaminants is important. Small molecules and polymers can substantially interfere with the MS analysis. Clean water is essential. We use distilled water that is additionally purified through columns to remove trace contaminants. In addition to clean water for all steps of the procedure, it is also important to use clean tubes and microfiltration units. Some brands of microfuge tubes and filtration devices have polymeric contaminants that appear as polyethylene glycol-related compounds on MS; these contaminants can overwhelm the signals from the tissue-derived peptides. Rinse and dry tubes with ultrapure deionized water before use. Clean filtration units by passing water through them before filtering peptides. Use low-retention microfuge tubes to reduce the loss of peptides through binding to the plastic surface.
2. We previously used anhydrotrypsin-agarose from Panvera LLC, but this reagent does not appear to be currently available (Panvera no longer exists as a separate company). Other manufacturers of anhydrotrypsin-agarose can be found on the internet, and should be suitable. Alternatively, it is possible to custom-synthesize this reagent from trypsin, agarose, and simple reagents, as described [21].
3. In the original report of this technique, mice were sacrificed by exposure to $CO_2$, and no specific steps were taken to inactivate proteases prior to dissection of the tissue [15]. Furthermore, the peptides were extracted from tissue using boiling acetic acid. Subsequent to the original report using this procedure, it was found that these techniques led to numerous postmortem changes [20, 22, 23]. If brain tissue is not heat-stabilized within 1 min of death, general protein degradation greatly increases levels of peptides that are derived from cytosolic, mitochondrial, and nuclear proteins [23, 24]. Also, levels of neuropeptides decrease within minutes of death, possibly due to secretion of peptide-containing vesicles and subsequent degradation of the peptides by extracellular proteases. These processes can be prevented by rapid heat inactivation of the brain tissue, as described in this protocol. Hot acid is not recommended because it causes cleavages of peptides, primarily adjacent to Asp residues [21]. These changes from the original report of the technique are reflected in this protocol.
4. Temperatures >65 °C appear to be sufficient to inactivate proteases in mouse brain, but temperatures of ~80 °C produced more uniform results in an experiment testing microwave conditions [20]. Each microwave oven needs to be calibrated to ensure that the proper internal temperature of the brain is achieved. Mice can also be sacrificed by microwave

irradiation, although this requires an expensive piece of equipment that is not available in most laboratories [22]. Alternatively, brain tissue can be rapidly dissected and heat-inactivated in a specialized device, as described [24] and also reported in another chapter in this book (see Fridjonsdottir et al.).

5. After heating to 80 °C, the brain is no longer elastic and cannot be dissected by standard approaches used for fresh tissue. We have found the best method of dissection uses a razor blade held in a coronal orientation to make slices that include the desired brain regions, and these slices are then further cut to isolate the regions of interest. Coordinates for dissecting some of the major brain region are described in the mouse brain atlas [25]. The frontal cortex is obtained by removing the olfactory bulb and taking the slice that is anterior to Bregma 1.94. To obtain cortex and striatum, make another coronal cut at bregma 0.00; the resulting slice of Bregma 1.94 to 0.00 contains cortex and the striatum. The cortex is isolated from the dorsal side of the slice. The remainder of the slice includes the caudate putamen, nucleus accumbens, septum, and ventral palladium. Another coronal cut is made at Bregma −3.00; this coronal section of Bregma 0.00 to −3.00 is dissected into the hippocampus, thalamus, amygdala, and hypothalamus. Cortex can also be dissected from this section.
6. Use settings that completely disrupt the tissue (based on visual inspection). Typical settings are 20 s of sonication (1 pulse per s) at duty cycle 3, with 50% output. A small additional volume of water is used to rinse the tip of the sonicator into the tube containing the brain extracts to maximize recovery. The sonicator tip should be washed with water between tissue extractions to prevent contamination between samples.
7. Most detergents cause problems for subsequent MS analysis. Even though the detergent is not present in the final elution buffer, residual amounts of some detergents can interfere with the MS signals. CHAPS is one of the better detergents for subsequent MS analysis.
8. To elute the peptides, we apply 1 mL of water (Subheading 3.2, **step 8**) or 5 mM HCl (Subheading 3.2, **step 9**) to the top of the column and collect the eluate into a microfuge tube, then repeat with another 1 mL and collect into a fresh tube. This is done for two reasons. First, it is easier to lyophilize the samples in a vacuum concentrator. Second, the addition of 2 × 1 mL provides a slower flow rate than simply adding 2 mL to the top of the column, and in theory this should allow for higher recovery of peptides, although this has not been tested.
9. The water and HCl elutes can be combined or analyzed separately. We found many differences in the peptides that eluted with water versus HCl, so by analyzing separately we presumably detected more peptides than would have been identified by pooling the samples.

## References


1. Fricker LD (2012) In: Fricker LD, Devi L (eds) Neuropeptides and other bioactive peptides. Morgan & Claypool Life Sciences, Charleston, SC
2. Fuller RS, Sterne RE, Thorner J (1988) Enzymes required for yeast prohormone processing. Ann Rev Physiol 50:345–362
3. Bayliss WM, Starling EH (1902) The mechanism of pancreatic secretion. J Physiol 28:325–353
4. Andren PE, Caprioli RM (1999) Determination of extracellular release of neurotensin in discrete rat brain regions utilizing in vivo microdialysis/electrospray mass spectrometry. Brain Res 845:123–129
5. Garden RW, Shippy SA, Li L et al (1998) Proteolytic processing of the Aplysia egg-laying hormone prohormone. Proc Natl Acad Sci U S A 95:3972–3977
6. Fricker LD (2010) Analysis of mouse brain peptides using mass spectrometry-based peptidomics: implications for novel functions ranging from non-classical neuropeptides to microproteins. Mol Biosyst 6:1355–1365
7. Fricker LD, Gelman JS, Castro LM et al (2012) Peptidomic analysis of HEK293T cells: effect of the proteasome inhibitor epoxomicin on intracellular peptides. J Proteome Res 11:1981–1990
8. Ferro ES, Hyslop S, Camargo AC (2004) Intracellular peptides as putative natural regulators of protein interactions. J Neurochem 91:769–777
9. Fricker LD, Snyder SH (1982) Enkephalin convertase: purification and characterization of a specific enkephalin-synthesizing carboxypeptidase localized to adrenal chromaffin granules. Proc Natl Acad Sci U S A 79:3886–3890
10. Eipper BA, Mains RE (1988) Peptide alpha-amidation. Ann Rev Physiol 50:333–344
11. Varlamov O, Leiter EH, Fricker LD (1996) Induced and spontaneous mutations at Ser202 of carboxypeptidase E: effect on enzyme expression, activity, and intracellular routing. J Biol Chem 271:13981–13986
12. Naggert JK, Fricker LD, Varlamov O et al (1995) Hyperproinsulinemia in obese fat/fat mice associated with a point mutation in the carboxypeptidase E gene and reduced carboxypeptidase E activity in the pancreatic islets. Nat Genet 10:135–142
13. Rodriguiz RM, Wilkins JJ, Creson TK et al (2013) Emergence of anxiety-like behaviours in depressive-like Cpe(fat/fat) mice. Int J Neuropsychopharmacol 16:1623–1634
14. Zhang X, Che FY, Berezniuk I et al (2008) Peptidomics of Cpe(fat/fat) mouse brain regions: implications for neuropeptide processing. J Neurochem 107:1596–1613
15. Che FY, Yan L, Li H et al (2001) Identification of peptides from brain and pituitary of Cpe fat/Cpe fat mice. Proc Natl Acad Sci U S A 98:9971–9976
16. Fricker LD, McKinzie AA, Sun J et al (2000) Identification and characterization of pro-SAAS, a granin-like neuroendocrine peptide precursor that inhibits prohormone processing. J Neurosci 20:639–648
17. Jacob TC, Kaplan JM (2003) The EGL-21 carboxypeptidase E facilitates acetylcholine release at Caenorhabditis elegans neuromuscular junctions. J Neurosci 23:2122–2130
18. Husson SJ, Janssen T, Baggerman G et al (2007) Impaired processing of FLP and NLP peptides in carboxypeptidase E (EGL-21)-deficient Caenorhabditis elegans as analyzed by mass spectrometry. J Neurochem 102:246–260
19. Zasloff M (1992) Antibiotic peptides as mediators of innate immunity. Curr Opin Immunol 4:3–7
20. Che FY, Lim J, Biswas R et al (2005) Quantitative neuropeptidomics of microwave-irradiated mouse brain and pituitary. Mol Cell Proteomics 4:1391–1405
21. Refojo D, Kovalovsky D, Young JI et al (2002) Increased splenocyte proliferative response and cytokine production in beta-endorphin-deficient mice. J Neuroimmunol 131:126–134
22. Svensson M, Skold K, Svenningsson P et al (2003) Peptidomics-based discovery of novel neuropeptides. J Proteome Res 2:213–219
23. Skold K, Svensson M, Norrman M et al (2007) The significance of biochemical and molecular sample integrity in brain proteomics and peptidomics: stathmin 2-20 and peptides as sample quality indicators. Proteomics 7:4445–4456
24. Scholz B, Skold K, Kultima K et al (2011) Impact of temperature dependent sampling procedures in proteomics and peptidomics—a characterization of the liver and pancreas post mortem degradome. Mol Cell Proteomics 10:M900229MCP900200
25. Paxinos G, Franklin KBJ (2001) The mouse brain in stereotaxic coordinates. Academic, San Diego

# Chapter 14

# Mass Spectrometry Based Immunopeptidomics for the Discovery of Cancer Neoantigens

**Michal Bassani-Sternberg**

## Abstract

Recent data indicate that endogenous mutated cancer proteins can be processed and presented as HLA binding peptides, leading to their recognition in vivo as "non-self." Targeting such neoantigens would enable immune cells to distinguish between normal and cancerous cells, avoiding the risk of autoimmunity. So far, discovery of such neoantigens relies mainly on prediction-based interrogation of the "mutanome" using genomic information as input, followed by highly laborious and time-consuming T cell screening assays. Currently, mass spectrometry is the only unbiased methodology to comprehensively interrogate the naturally presented repertoire of HLA binding peptides, including peptides derived from tumor-associated antigens and post-translational modified peptides. This chapter describes a detailed protocol for in-depth and accurate mass spectrometry based immunopeptidomics, enabling the direct identification of tissue-derived neoantigens extracted from human tumors.

**Key words** HLA binding peptides, Neoantigens, Immunopeptidomics, Immunoaffinity purification, Mass spectrometry, Cancer immunotherapy

## 1 Introduction

Cancer immunotherapy reprograms the inherent capacity of immune cells to eliminate tumors by virtue of recognizing molecular entities expressed specifically on tumors but not on normal cells. The entities are short peptides that are presented on the cell surface on human leukocyte antigen (HLA) molecules, named the HLA binding peptides (HLAp), and collectively the immunopeptidome. Recent clinical data provide clear evidence that human tumor cells express and present antigenic HLAp that can be recognized by the patients' own (autologous) T cells. The enhancement of such reactivity, for example by using immune checkpoint blockade therapies, can lead to cancer regression in patients with advanced high mutational load tumors [1–6], [6–11] highlighting the involvement of mutated neo-antigens as the main targets [12–15]. The remarkable clinical results of the immune checkpoint blocking therapies have

Michael Schrader and Lloyd Fricker (eds.), *Peptidomics: Methods and Strategies*, Methods in Molecular Biology, vol. 1719, https://doi.org/10.1007/978-1-4939-7537-2_14, 

motivated researchers to discover the immunogenic cancer-specific antigens that mediate T cell-based killing and long-lasting disease control [8]. Such antigens may be further exploited in the development of personalized vaccines to enhance the reactivity of the checkpoint blocking therapies, especially in the cohort of nonresponding patients.

The HLA immunopeptidome is a highly dynamic, rich, and complex repertoire of peptides that inform T cells about abnormalities within diseased cells [16]. Dedicated cellular machinery for antigen processing and presentation ensures continuous sampling of intracellular proteins on the HLA class I molecules. Additional parallel machinery is responsible for processing of exogenous antigens on HLA class II molecules expressed on professional antigen presenting cells. This way, alterations in the genome, transcriptome, and the proteome are presented to T cells. In cancer, HLAp may be derived from tumor-associated (over)expressed self-proteins, from oncogenic viruses, endogenous retroviral elements, or mutated tumor proteins. Currently, mass spectrometry based immunopeptidomics is the only unbiased methodology to interrogate the repertoire of naturally presented HLAp in tissues [17]. Of interest, mass spectrometry has been successful in identifying neoantigens on cancer cell lines and in melanoma tissues [18–23]. This requires previous knowledge of the nonsynonymous genetic alterations in the investigated tissue. Since mutations are mostly private, the resulting MS data should be searched in a personalized manner against a customized reference database that includes the patient's mutated protein sequences. In addition to the direct identification of neoantigens, this discovery approach reveals the repertoire of the thousands of self peptides among them peptides derived from tumor-associated antigens and post-translational modified peptides. Of note, once more advanced computational algorithms are available, re-interrogating such immunopeptidomics dataset could lead to the identification of yet unforeseen novel antigens.

Here I describe a protocol enabling sequential immunoaffinity purification of HLA-I and HLA-II binding peptides directly from tissue samples as shown schematically in Fig. 1. Details regarding the analyses of these peptidome samples by mass spectrometry are included together with suggestions for the downstream computational pipeline using a new module in MaxQuant. Some of these procedures have previously been described [24, 25], and others are being prepared for publication (P. Sinitcyn et al., manuscript in preparation). While the extraction of DNA from tissues and PBMCs and the exome sequencing analysis are beyond the scope of this chapter, I describe here how to incorporate existing next-generation sequencing data into this computational pipeline to directly identify patient's specific tissue-derived neoantigens as described in [18].

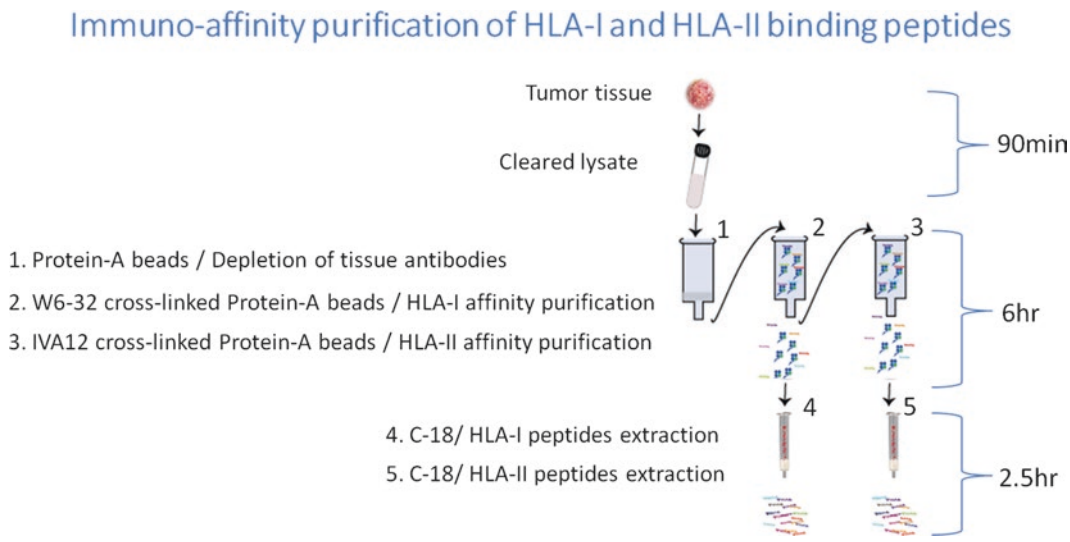



**Fig. 1** General overview of immunoaffinity-based purification of HLA-I and HLA-II binding peptides from one tissue sample for mass spectrometry analysis

## 2 Materials

Prepare all solutions with ultrapure water and analytical grade reagents.

### *2.1 Affinity Chromatography*

1. 5 mL of growth medium from HB95 hybridoma cells secreting the W6-32 anti-pan HLA-I and from HB145 hybridoma cells secreting the IVA12 anti-pan-HLA-II (*see* **Note 1**).
2. Protein A-Sepharose 4B Conjugate.
3. Wash solutions for purification of antibodies: 100 mM Tris–HCl pH 8, 20 mM Tris–HCl pH 8.
4. 0.1 N Acetic acid pH 3.
5. 1 M Tris–HCl pH 8.
6. Preservation buffer: PBS with 0.02% $NaN_3$.
7. 0.2 M Sodium Borate buffer pH 9: Prepare solution A: dissolve 12.4 g of Boric acid in 1 L water, and solution B: dissolve 19.05 g Borax (Sodium tetraborate) in 1 L water. In a new bottle mix 50 mL of solution A, 59 mL of solution B, and 91 mL water to receive 200 mL of 0.2 M Sodium Borate buffer pH 9.
8. Dimethylpimelimidate.
9. Ethanolamine solution: add 600 µL of ethanolamine to 50 mL of water and adjust to pH 8 with HCl. Keep in the dark at room temperature.
10. Poly Prep 9 cm columns (Bio-Rad) with a stand (*see* **Note 2**).
11. Nano-drop.
12. SDS-PAGE system.

### 2.2 HLAp Sample Preparation

1. Tumor tissues (flash frozen stored at −80 °C).
2. Lysis buffer: 0.25% sodium deoxycholate, 0.2 mM iodoacetamide, 1 mM EDTA, 1:200 Protease Inhibitors Cocktail (we use the cocktail reagent from Sigma-Aldrich), 1 mM Phenylmethylsulfonyl fluoride, 1% octyl-beta-D glucopyranoside in PBS. Prepare fresh (about 13 mL per tissue). Keep on ice.
3. Wash solution A: 150 mM NaCl and 20 mM Tris–HCl pH 8. Keep on ice.
4. Wash solution B: 400 mM NaCl and 20 mM Tris–HCl pH 8. Keep on ice.
5. Wash solution C: 20 mM Tris–HCl pH 8. Keep on ice.
6. 0.1 N Acetic acid pH 3.
7. Sep-Pak C-18 wash solution: 0.1% TFA.
8. Sep-Pak C-18 peptide elution solution: 0.1% TFA with 30% acetonitrile.
9. Sep-Pak C-18 protein elution solution: 0.1% TFA with 80% acetonitrile.
10. SepPak cartridges vac 1 cc tC-18 (Waters) and dedicated adaptor (Waters).
11. 10 mL plastic syringe
12. ~30 mL centrifuge tubes.
13. ULTRA-TURRAX homogenizer (IKA, or comparable equipment). Keep blade at 4 °C.
14. Laminar flow.
15. High speed centrifuge.
16. Vacuum centrifuge.

### 2.3 Mass Spectrometry

1. LC-MS/MS. We use a Q Exactive mass spectrometer coupled on-line to nano-Easy 1200 UHPLC system fitted with Nanospray Flex™ source (Thermo Fisher Scientific).
2. Column oven.
3. Analytical column with 8 μm tip opening and a diameter of 50 cm × 75 μm (we use the column from New Objectives).
4. ReproSil-Pur C18 1.9 μm particles 120 Å pore size resin (we use the resin from the company Dr. Maisch GmbH).
5. MaxQuant computational proteomics platform.

## 3 Methods

The protocol described below provides instructions on how to purify and cross-link antibodies sufficient for the immunoaffinity purification of HLA-I and HLA-II complexes from four distinct

tissue samples of 1 $cm^3$ each. The procedures can be scaled up or down according to the amount of available tissues.

### 3.1 Purification of Anti-HLA Antibodies from Hybridoma Growth Medium

Five milliliters of the hybridoma growth medium yields about 10–15 mg of antibodies (*see* **Note 1**). Purified antibodies can be stored at −20 °C for future use.

1. Into two distinct empty Bio-Rad columns add a final volume of 2 mL of beads per column. Allow the preservation buffer to drain and wash the beads with 10 mL of 100 mM Tris–HCl pH 8. Close the tip of the columns with the caps.
2. Label the columns as "W6-32" and "IVA12." Load 5 mL of each of the growth medium containing the monoclonal antibodies onto the corresponding labeled columns. Close the columns with the upper caps and slowly rotate at room temperature for 15 min.
3. Remove the caps and let the liquid flow through. Wash the columns with 10 mL 100 mM Tris–HCl pH 8 and then with 10 mL of 20 mM Tris–HCl pH 8.
4. Elute the antibodies with serial elutions each of 1 mL using acetic acid 0.1 N pH 3 into Eppendorf tubes containing 300 μL of 1 M Tris–HCl pH 8. Collect 6–7 elutions in total per antibody. Vortex the Eppendorf tubes for 3 s (*see* **Note 3**).
5. Quantify the amount of antibodies in each tube using a NanoDrop (*see* **Note 4**).
6. Wash the beads with 10 mL of 100 mM Tris–HCl pH 8 and then with 2 mL of PBS containing 0.02% $NaN_3$. Cover the beads with 2 mL of PBS with 0.02% $NaN_3$ and close the columns with the caps. Keep the columns at 4 °C for future use.

### 3.2 Preparation of Anti-HLA Affinity Columns

Cross-link sufficient amount of antibodies to the beads as described below and use cross-linked antibodies of same batch for parallel purifications of tissue sections from the same tumor. Beads cross-linked with antibodies may be stored at 4 °C for several months and therefore may be prepared in advanced.

1. Into two Bio-Rad columns labeled as "W6-32-ProteinA" and "IVA12-ProteinA" add a final volume of 2 mL of beads per column. Allow the preservation buffer to drain and wash the beads with 10 mL of 100 mM Tris–HCl pH 8. Close the tip of the columns with the caps.
2. Load 10 mg of each of the antibodies onto the corresponding labeled column. Close the columns with the upper caps and slowly rotate at room temperature for 30 min. Remaining unused antibodies may be stored at −20 °C for future use.
3. Remove caps and allow the liquid to drain. Wash the columns with 10 mL of 0.2 M Sodium Borate buffer pH 9. Close the

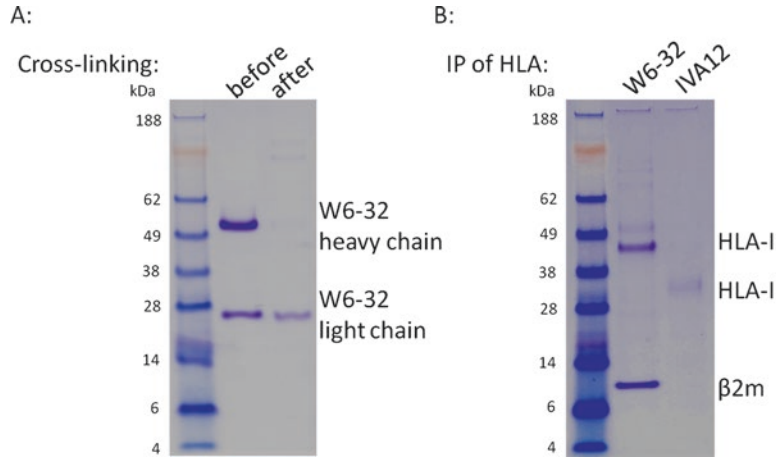



**Fig. 2** Examples of SDS-PAGE analyses of W6-32 antibody before and after cross-linking to Protein-A beads (**a**) and of HLA-I and HLA-II complexes isolated by immunoaffinity purification (IP) from human ovarian cancer tissue (**b**)

tips of the columns with a cap. Resuspend the beads with 2 mL of 0.2 M Sodium Borate buffer pH 9 and transfer into new Eppendorf tubes 30 μL of the resuspended beads from each column for SDS-gel analysis to evaluate the efficiency of the cross-linking.

4. Add 20.8 mg Dimethylpimelimidate (final concentration of 20 mM) directly onto the column. Close the column with the upper cap and rotate at room temperature for 30 min (*see* **Note 5**).
5. Open the upper cap. Resuspend the beads and transfer into a new microfuge tube 30 μL of the resuspended beads from each column for SDS-PAGE analysis to evaluate the efficiency of the cross-linking as shown in Fig. 2 (*see* **Note 6**).
6. Open the cap of the tip and let the liquid flow through. Wash the column with 5 mL of 0.2 M ethanolamine pH 8.
7. Close the tip with the cap and add 5 mL of ethanolamine pH 8. Close the upper cap and rotate at room temperature for 2 h.
8. Open the caps. Allow the liquid to flow through. Wash the columns with 10 mL of PBS with 0.02% $NaN_3$. Cover the beads with 2 mL of PBS with 0.02% $NaN_3$ and close the columns with the caps. Keep the columns at 4 °C for future use.

### *3.3 Purification of HLA Class I and HLA Class II Peptides*

Affinity columns may be reused, preferably using biological samples from the same donor and can be stored at 4 °C for several months. Keep all tubes and solutions on ice.

1. Prepare and label 3 new and prewashed Bio-Rad columns for each tissue sample. Label the columns as "pre-column" (ProteinA only), "W6-32-ProteinA" and "IVA12-ProteinA"

including the tissue identification number. Distribute into the columns labeled as pre-column a final volume of 400 μL of ProteinA beads. Into each of the remaining columns add a final volume of 400 μL of either the W6-32 or IVA12 cross-linked beads. Remove the preservation liquid and wash all columns with 1 mL of acetic acid 0.1 N, followed by 10 mL of 100 mM Tris–HCl pH 8. Place the columns in a stand at 4 °C.

2. Place the ULTRA-TURRAX homogenizer inside a laminar flow in order to protect the user from aerosols that may be generated during the homogenization of the tissues.
3. Process the tissues one at a time as indicated here. Add into the centrifuge tubes 5 mL of freshly prepared cold lysis buffer. Place the tube in a cup filled with ice. With a pair of tweezers transfer the first frozen tissue directly into the tube containing the cold lysis buffer and wait 1–2 min. Homogenize the tissue at high speed for 5 s using the Ultraturax homogenizer (*see* **Notes 7** and **8**).
4. Add to the tube additional 5 mL of cold lysis buffer and mix the tube gently.
5. Disassemble the blade and wash it with ethanol and then with water. Wipe it with a paper towel and assemble it again.
6. Repeat **steps 3–5** again with the following tissue.
7. Keep the tubes on ice for 1 h. Mix the tubes gently every 10 min.
8. Balance the centrifuge tubes and centrifuge at 40,000 *g* at 4 °C for 30 min using the high speed centrifuge.
9. Place the stand with the prepared Bio-Rad columns (from **step 1**) in a cold room where **steps 9–12** should be performed. Align the columns on top of 15 mL collection tubes. Transfer the supernatant fractions from each of the samples from **step 8** and load them onto the corresponding pre-columns. Remove the tip caps and collect the flow-through fractions in the 15 mL tubes. Once the lysates passed completely, add 1 mL of lysis buffer to the columns and collect also the dead-volume captured in the beads.
10. Transfer the flow-through fractions from the 15 mL tubes and load them directly on the corresponding W6-32-ProteinA columns. Collect the flow-through fractions and load them again. Once the lysates passed completely, add additional 1 mL of lysis buffer to the columns and collect also the dead-volume captured in the beads and collect the flow through.
11. Finally, load the flow-through fractions on the IVA12-ProteinA columns. Collect the flow-through fractions and load them again. Store the flow-through fractions at −20 °C.
12. Wash each of the W6-32-ProteinA and IVA12-ProteinA columns with 10 mL of wash solution A, then with 10 mL of wash

solution B, followed by 10 mL of wash solution A, and lastly with 10 mL of wash solution C. In the meantime continue to **step 13**.

13. Wash the pre-columns first with 5 mL of acetic acid to remove bound antibodies, and then wash the columns with 10 mL of 100 mM Tris–HCl pH 8. Cover the beads with 2 mL of PBS with 0.02% $NaN_3$ and close the columns with the caps. Keep the pre-columns at 4 °C for future use.
14. Once the last wash has drained completely, place a stand containing seven labeled microfuge tubes under the tip of each column. Add 0.5 mL of acetic acid 0.1 N pH 3 into the column. Collect the elution in the first tube. Repeat this step six more times and collect the elutions one at a time into the remaining six tubes. Repeat this for all the columns. Keep the tubes containing the elutions on ice (*see* **Note 9**).
15. Wash all columns with 10 mL of 100 mM Tris–HCl pH 8. Cover the beads with 2 mL of PBS with 0.02% $NaN_3$ and close the columns with the caps. Keep the columns at 4 °C for future use.
16. Optional: Analyze 10 μL of each elution by SDS-PAGE to evaluate the yield and the purity of the HLA complexes as shown in Fig. 2. The HLA-I heavy chain and the β2m should appear as a 43 kDa and a 12 kDa bands, respectively, starting from the second or third elution. The two subunits of the HLA-II should appear as broad faint band that is spread around the 28–30 kDa. Only fractions that contain HLA should be used for the following purification of the HLA binding peptides.

### *3.4 Purification of HLA Binding Peptides*

1. Prepare 2 Sep-pak cartridges for each tissue sample—for the HLA class I and HLA class II peptides samples, and label them accordingly.
2. Using a syringe and the dedicated adaptor wash the cartridges first once with 1 mL of 80% ACN in 0.1% TFA, then twice with 1 mL 0.1% TFA.
3. Load each of the biological samples (HLA class I or HLA class II, per tissue) on a separate SepPak cartridge (load only elution fractions that contain HLA and β2m as detected by SDS-PAGE onto the cartridges). Pass them through slowly (speed of about 1 mL in 20 s).
4. Wash the cartridges twice with 1 mL of 0.1% TFA.
5. Elute the HLA binding peptides into a collection tube with 300 μL of 30% ACN in 0.1% TFA (*see* **Note 10**).
6. Elute the higher molecular weight proteins into another collection tube with 300 μL of 80% ACN in 0.1% TFA.
7. Dry the samples containing the peptides and the samples containing the proteins using vacuum centrifugation at 30 °C

for about 1.5 h and 45 min, respectively, until completely dried.

8. Resuspend the peptides with 17ul 0.1% TFA and store at −20 °C until mass spectrometry analyses.
9. Resuspend the proteins with 50ul 0.1% TFA or with PBS and store at −20 °C (*see* **Note 11**).

### 3.5 Mass Spectrometry Analyses

This part of the protocol describes the analysis of HLA peptides using the Easy 1200 UHPLC system coupled online to a Q Exactive HF mass spectrometer, both from Thermo Fisher Scientific. Other comparable instruments could be used. Analytical column with 8 μm tip opening and a diameter of 50 cm × 75 μm may be self-packed with the ReproSil-Pur C18 1.9 μm particles having 120 Å pore size resin. Such long analytical column should be heated to 50 °C using a column oven. Make sure that the analytic column is clean and measure a blank sample before the HLA peptides samples to exclude carry over from previously measured samples. Preferably, two technical replicates of 5 μL each should be injected sequentially. The remaining sample may be used in the future to validate targets of interest by spiking into these remaining samples isotopically heavy-labeled synthetic peptides. Alternatively, remaining samples may be used for further fractionation. Use the same MS settings for HLA class I and HLA-II peptidomes, unless indicated otherwise.

1. Place 11 μL of each HLA peptides samples in the auto-sampler. Inject first 5 μL of the first sample.
2. Apply a linear gradient of 2–30% of 80% ACN and 0.1% formic acid at a flow rate of 250 nL/min over 90 min.
3. Suggested settings for the mass spectrometer are: MS scan range can be set to 300 to 1650 *m/z* with a resolution of 60,000 (200 *m/z*) and a target value of 3e6 ions. Data may be acquired with data-dependent "top10" method, which isolates the ten most intense ions and fragments them by higher-energy collisional dissociation (HCD) with an applied normalized collision energy of 27% and with an AGC target value of 1e5 with a maximum injection time of 120 ms and MS/MS resolution was 15,000 (200 *m/z*). For HLA class I peptides, in case of unassigned precursor ion charge states, or charge states of four and above, no fragmentation should be performed. For HLA class II peptides, in case of unassigned precursor ion charge states, or charge states of one or of five and above, no fragmentation should be performed. Dynamic exclusion of fragmented *m/z* values from further selection can be set for 20 s.
4. Inject again 5 μL of the first sample (technical duplicate).
5. Wash the system with two standard wash runs before loading the next biological sample.

### 3.6 Direct Identification of Neoantigens Using MaxQuant

A special module in the MaxQuant computational proteomics platform has recently been developed that enables the search for peptides based on genomic variations (Sinitcyn P. et al., in preparation). MaxQuant takes as input aligned reads from exome sequencing data and calls variants. Variants increase the peptide search space by either including or excluding them on each peptide. To account for different a priori probabilities of different peptide classes the posterior error probability is calculated depending on the type of the peptide (such as unmodified peptides without variants, unmodified peptides resulting from a variant). A common PSM-FDR threshold is then applied based on this peptide class dependent posterior error probability.

Employ the following computational analysis for each patient separately.

1. Use the latest version of the MaxQuant that includes the *Mutation Extraction* tool. In the *Mutation Extractor* tool, upload the exome sequencing data as .bam format for both the healthy tissue and the tumor tissue from one patient and specify the genome used for the alignment and its annotation. Default settings for promiscuous mutation calling with the following thresholds should be applied: total read depth of the position >10 reads, number of reads which support a variant >5 reads, and at the same time the minimum variants frequency = 5%. This tool generates a FASTA file that includes information about nonsynonymous alterations between both tissues and the reference genome, and a VCF file that lists the alterations including genomic locations and the number of reads in both the healthy tissue and the tumor (*see* **Note 12**).
2. Load the corresponding MS raw files.
3. Use the default parameters except for parameters indicated here: In the *Group-Specific Parameters* window enable N-terminal acetylation (42.010565 Da) and methionine oxidation (15.994915 Da) as variable modifications. The enzyme specificity should be set as "unspecific." In the *Global Parameters* window upload the generated FASTA file for the database search. Do not specify any fixed modifications. Change the mutation mode from "none" to "from file." A false discovery rate (FDR) of 0.05 is recommended for peptides and no FDR should be applied on the proteins level. Enable the "match between runs" option, which allows matching of identifications across different replicates that belongs the same patient, in a time window of 0.5 min and an initial alignment time window of 20 min. Launch MaxQuant (*see* **Note 13**).
4. The list of all identified unique HLA peptides can be found in the "Peptides" output table. Two columns in this table indicate the presence of peptides containing alterations from the

reference genome and the ID number of the mutations as listed in the VCF file that was generated by MaxQuant *Mutation Extractor* tool (in **step 1** above). Most of these hits will be peptides containing single nucleotide polymorphism (SNP) compared to the reference genome and therefore the corresponding read counts in the VCF file for such SNPs will be similar in the tumor and in the healthy tissues. In case alterations can only be detected in the exome sequencing data of the tumor and not in the healthy tissue then the corresponding peptides are true neoantigens (*see* **Note 14**).

## 4 Notes

1. Grow the hybridoma cells in CELLine350 or CELLine1000 flasks (Integra Biosciences) according to the manufacturer's instructions. Spin down at maximum speed using bench-top centrifuge to remove cell debris. Collect the growth medium that contains the highly concentrated monoclonal antibodies and store it at −20 °C until use.
2. Bio-Rad columns should be washed with 10 mL 1% SDS and 40 mL of water to remove possible PEG contaminants. Procedures performed with Bio-Rad columns are simply based on gravity. Place the columns in a stand that can fit 15 mL collection tubes.
3. If needed, adjust pH of the eluted antibodies to pH 8 with 1 M Tris–HCl pH 8.
4. Purity of antibodies may be estimated by loading 1% of each of the eluates on SDS-PAGE under denaturing conditions. The light and heavy chains of the antibodies should be visible.
5. Dimethylpimelimidate is stored at −20 °C. It should be allowed to warm to room temperature before use.
6. Add sample buffer to the beads samples, boil for 5 min, spin down the tubes, and collect 15 μL of the supernatant. Analyze the supernatant by SDS-PAGE. Both the antibodies heavy and light chains should be visible in the samples before the cross-linking, while the heavy chains should not be visible in the samples after cross-linking.
7. It is highly recommended to weigh the tissues before use.
8. A longer homogenization might be needed depending on the tissue content. Avoid long homogenization as this might lead to overheating of the tissues and will generate excess of foam.
9. At this stage the samples can be stored at −20 °C.
10. HLA peptides are purified using a Sep-pak c18 cartridge. Peptides are less hydrophobic than the HLA heavy chain and

the β2m. The peptides are eluted first from the SepPak cartridge using 30% ACN in 0.1% TFA. Afterwards the proteins may be eluted using 80% ACN in 0.1% TFA.

11. A fraction of the samples containing the HLA and β2m proteins may be used for SDS-PAGE or Western-blot analyses to estimate their recovery.
12. Mutation calling and the generation of the FASTA file should take approximately 1 day using a standard computer.
13. The MaxQuant search should be completed within approximately 2 days using a standard computer.
14. The number of identified peptides is in correlation with the expression of the corresponding HLA complexes in the investigated tissue. Using this protocol approximately 5000–15,000 unique HLA class I peptides may be identified from 1 $cm^3$ tumor tissue sample. The number of HLA class II peptides may vary from a few peptides to several thousands, depending on the tissue type. This protocol may lead to the identification of a few neoantigens, however, not in all tissue samples. High mutational loads and HLA expression will favor MS-based detection of neoantigens.

## References


1. Grupp SA, Kalos M, Barrett D et al (2013) Chimeric antigen receptor-modified T cells for acute lymphoid leukemia. N Engl J Med 368(16):1509–1518
2. Hodi FS, O'Day SJ, McDermott DF et al (2010) Improved survival with ipilimumab in patients with metastatic melanoma. N Engl J Med 363(8):711–723
3. Kantoff PW, Higano CS, Shore ND et al (2010) Sipuleucel-T immunotherapy for castration-resistant prostate cancer. N Engl J Med 363(5):411–422
4. Rosenberg SA, Yang JC, Sherry RM et al (2011) Durable complete responses in heavily pretreated patients with metastatic melanoma using T-cell transfer immunotherapy. Clin Cancer Res 17(13):4550–4557
5. Walter S, Weinschenk T, Stenzl A et al (2012) Multipeptide immune response to cancer vaccine IMA901 after single-dose cyclophosphamide associates with longer patient survival. Nat Med 18(8):1254–1261
6. Wolchok JD, Kluger H, Callahan MK et al (2013) Nivolumab plus ipilimumab in advanced melanoma. N Engl J Med 369(2):122–133
7. Larkin J, Chiarion-Sileni V, Gonzalez R et al (2015) Combined Nivolumab and Ipilimumab or monotherapy in untreated melanoma. N Engl J Med 373(1):23–34
8. Schadendorf D, Hodi FS, Robert C et al (2015) Pooled analysis of long-term survival data from phase II and phase III trials of Ipilimumab in unresectable or metastatic melanoma. J Clin Oncol 33(17):1889–1894
9. Borghaei H, Paz-Ares L, Horn L et al (2015) Nivolumab versus docetaxel in advanced nonsquamous non-small-cell lung cancer. N Engl J Med 373(17):1627–1639
10. Powles T, Eder JP, Fine GD et al (2014) MPDL3280A (anti-PD-L1) treatment leads to clinical activity in metastatic bladder cancer. Nature 515(7528):558–562
11. Le DT, Uram JN, Wang H et al (2015) PD-1 blockade in tumors with mismatch-repair deficiency. N Engl J Med 372(26):2509–2520
12. Castle JC, Kreiter S, Diekmann J et al (2012) Exploiting the mutanome for tumor vaccination. Cancer Res 72(5):1081–1091
13. Brown SD, Warren RL, Gibb EA et al (2014) Neo-antigens predicted by tumor genome meta-analysis correlate with increased patient survival. Genome Res 24(5):743–750
14. Rizvi NA, Hellmann MD, Snyder A et al (2015) Cancer immunology. Mutational land-

scape determines sensitivity to PD-1 blockade in non-small cell lung cancer. Science 348(6230):124–128

15. Snyder A, Makarov V, Merghoub T et al (2014) Genetic basis for clinical response to CTLA-4 blockade in melanoma. N Engl J Med 371(23):2189–2199
16. Neefjes J, Jongsma ML, Paul P et al (2011) Towards a systems understanding of MHC class I and MHC class II antigen presentation. Nat Rev Immunol 11(12):823–836
17. Bassani-Sternberg M, Coukos G (2016) Mass spectrometry-based antigen discovery for cancer immunotherapy. Curr Opin Immunol 41:9–17
18. Bassani-Sternberg M, Braunlein E, Klar R et al (2016) Direct identification of clinically relevant neoepitopes presented on native human melanoma tissue by mass spectrometry. Nat Commun 7:13404
19. Bassani-Sternberg M, Pletscher-Frankild S, Jensen LJ et al (2015) Mass spectrometry of human leukocyte antigen class I peptidomes reveals strong effects of protein abundance and turnover on antigen presentation. Mol Cell Proteomics 14(3):658–673
20. Carreno BM, Magrini V, Becker-Hapak M et al (2015) Cancer immunotherapy. A dendritic cell vaccine increases the breadth and diversity of melanoma neoantigen-specific T cells. Science 348(6236):803–808
21. Gubin MM, Zhang X, Schuster H et al (2014) Checkpoint blockade cancer immunotherapy targets tumour-specific mutant antigens. Nature 515(7528):577–581
22. Kalaora S, Barnea E, Merhavi-Shoham E et al (2016) Use of HLA peptidomics and whole exome sequencing to identify human immunogenic neo-antigens. Oncotarget 7(5):5110–5117
23. Yadav M, Jhunjhunwala S, Phung QT et al (2014) Predicting immunogenic tumour mutations by combining mass spectrometry and exome sequencing. Nature 515(7528):572–576
24. Tyanova S, Temu T, Cox J (2016) The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. Nat Protoc 11(12):2301–2319
25. Cox J, Mann M (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat Biotechnol 26(12):1367–1372

# Chapter 15

# Milk Peptidomics to Identify Functional Peptides and for Quality Control of Dairy Products

**David Dallas and Søren Drud Nielsen**

## Abstract

Human milk and dairy products are important parts of human nutrition. In addition to supplying nutrients, milk proteins contain fragments—peptides—with important biological functions that are released during processing or digestion. Besides their potential functional relevance, peptides released during processing can be used as markers of ripening stage or product deterioration. Hence, identification and quantification of peptides in milk can be used to assay potential health benefits or product quality. This chapter describes how to extract, identify, and analyze peptides within breast milk, dairy products, and dairy digestive samples. We describe how to analyze extracted peptides with liquid chromatography-mass spectrometry, to use software to identify peptides based on database searching, and to extract peak areas for relative quantification of each peptide. We describe methods for data analysis, including predicting which enzymes are responsible for protein cleavage, identifying the site specificity of protein breakdown, mapping identified peptides to known bioactive peptides, and applying models to predict novel functional peptides.

**Key words** Peptidomics, Milk, Mass spectrometry, Peptide, Bioactive, Identification, Quantification, X! Tandem, Skyline, Proteome Discoverer

## 1 Introduction

Human milk has evolved to match the newborn infant's nutritional needs. Beyond providing basic nutrients, milk also serves as a source of biologically active molecules with major impacts on the infant's health and development. Fragments of milk proteins produced by enzymatic cleavage—peptides—have an array of functions, including immunomodulation [1, 2], opioid-like activity [3, 4], and antimicrobial [5–7], antihypertensive [8], and antioxidant actions [9]. These functions are often different from the functions of the parent protein. In effect, these peptides are functional units encrypted within the parent sequence. Most previous studies of bioactive milk peptides are based on peptides released from in vitro digestion of isolated milk proteins; thus they neither represent in vivo digestion nor identify which peptides are biologically relevant.

Michael Schrader and Lloyd Fricker (eds.), *Peptidomics: Methods and Strategies*, Methods in Molecular Biology, vol. 1719, https://doi.org/10.1007/978-1-4939-7537-2_15, 

Recently, we employed peptidomics to examine the in vivo gastric digestion of milk in infants. For this study, we used milk and gastric samples collected from tube-fed infants and demonstrated that milk protein digestion begins within the mother's mammary gland, where milk proteases release hundreds of peptides [10, 11]. Within the infant gut, milk proteins are further digested into smaller fragments, producing hundreds of additional peptides [12]. Our bioinformatic analyses demonstrated that these peptides derived from specific milk proteins at specific sites within the proteins, and that many matched known functional peptides with antimicrobial and immunomodulatory actions.

Peptidomics is used not only to study digestion of proteins—it has many uses in dairy science. Dairy scientists apply peptidomics to identify peptides in dairy products, including cheeses [13–16], yogurts [17], and kefir [18]. These peptidomics studies revealed that differences in starting materials, production techniques, and storage time resulted in major differences in the peptide profile. For example, peptidomic analysis of lactose-hydrolyzed ultra-high temperature heat-treated milk indicated that residual proteases in the lactase preparation caused protein degradation that resulted in bitter flavor development with storage time [19–21]. Peptidomics can also be used to monitor inter-individual variation among cows [22] and as a marker for their health status—we demonstrated that peptide profiles differed in milk from healthy and mastitic quarters (teats) within the same cow [23].

This chapter describes how to extract peptides from dairy products and their digesta, to analyze them by liquid chromatography (LC) coupled with mass spectrometry (MS), and to identify and quantify them by label-free relative quantification (overview in Fig. 1). Several different MS instruments are appropriate for peptidomics, including Orbitrap, Q-TOF, and others [24]. Label-free relative quantification can be accomplished using peptide signal intensities. The ion-signal intensity approach uses extracted chromatographic areas to compare peptide abundances across samples. The software used for identification and quantification depends on the instrument used as the output file type differs among instruments from different vendors. This chapter describes two methods for identification and label-free relative quantification of peptides. The first method uses X! Tandem and Skyline, which are two popular noncommercial programs (Subheadings 3.7 and 3.8). Within Skyline, we create a library of X! Tandem-identified peptides across all samples. This library is used to match peaks in other samples based on retention time and mass. This approach identifies peptides that may be present in the MS data as parent molecule masses but not identified in that sample by X! Tandem. The second method uses the Proteome Discoverer software from Thermo

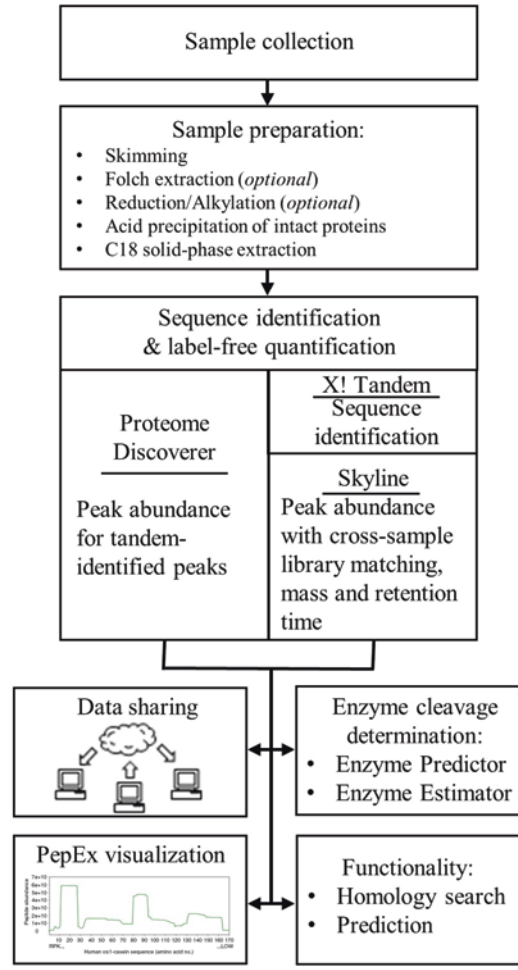



**Fig. 1** Overview of the steps described in this chapter for analysis and identification of milk peptides

Fisher (Subheading 3.9). This method determines only peptide abundance from tandem-identified peptides within the sample as no library-based peak searching is available. Selecting which program to use depends on the specific research needs, sample type, and instrument combination. This chapter includes protocols for further data analysis using bioinformatics techniques. The first tool visualizes from where identified peptides derive across the parent protein sequence. A homology search tool identifies the peptides that match previously identified functional peptides. Several tools use structure-function relationships to predict the likelihood that each identified peptide is functional. Two tools predict which proteases in a sample were responsible for protein cleavages based on the peptidomic data. Finally, we suggest that the peptidomics results of any study should be uploaded to ProteomeXchange and shared with the public.

## 2 Materials

All solutions should be prepared in glassware using nanopure water (18 MΩ) and analytical-grade reagents. Be careful to avoid any residual detergents and other contaminants as these will deteriorate the MS signal. Solutions can be stored at room temperature.

### 2.1 Sample Preparation

1. Many types of samples from milk or dairy products can be analyzed, including fresh milk, milk digesta, in vitro enzymatic digests, cheese, yogurt, and kefir.
2. Protease inhibitor cocktail, complete mini EDTA-free 50× stock solution (we use product 4693124001 from Roche).
3. Folch solution: 2:1 chloroform:methanol (MeOH).
4. Freeze drier or vacuum centrifuge.
5. 200 g/L trichloroacetic acid (TCA) in nanopure water.
6. C18 solid-phase extraction 96-well plate (we use Glygen, Columbia, MD).
7. Activation solution: 99% acetonitrile (ACN), 0.1% trifluoroacetic acid (TFA), 0.9% water (v/v).
8. TFA solution: 0.1% TFA in nanopure water (v/v).
9. Equilibration/wash solution: 1% ACN, 0.1% TFA, 98.9% water (v/v).
10. Peptide elution solution: 80% ACN, 0.1% TFA, 19.9% water (v/v).
11. Peptide rehydration solution: 2% ACN, 0.1% TFA, 97.9% water (v/v).

### 2.2 Mass Spectrometry

1. Solvent A: 0.1% formic acid (FA).
2. Solvent B: 100% ACN.
3. Liquid chromatography coupled to nanoelectrospray (we use a Waters Nano Acquity UHPLC (Waters, Milford, MA) with a Proxeon nanospray source).
4. A reversed-phase trap column (we use a 100 μm × 25 mm Magic C18 100 Å 5U column) for online desalting and a reversed-phase analytical column for peptide separation (we use a 75 μm × 150 mm Magic C18 200 Å 3U column, Waters, Milford, MA).
5. Mass spectrometry instrument (we use a Q Exactive Plus hybrid quadrupole-Orbitrap mass spectrometer from Thermo Fisher Scientific, Waltham, MA, but many other instruments can be used).

## 3 Methods

### 3.1 Liquid Milk Sample Collection and Handling

1. For human milk collection, nipples should be cleaned prior to collection. For bovine milk collection, teats should be washed in water and then dipped in an antiseptic solution. Collection volumes can vary, but we typically collect at least 1 mL of sample.
2. Milk samples must be frozen as soon as possible after collection and kept at −80 °C to limit milk protease activity, which could alter the peptide profile (*see* **Note 1**).

### 3.2 Peptide Extraction and Sample Preparation

1. Thaw samples and bring to 4 °C (*see* **Note 2**).
2. Pipette 100 μL of each liquid milk sample into a 1.5-mL tube. Add 5 μL of protease inhibitor stock solution and mix with a vortex on low speed for 10 s.
3. For solid samples, like cheese, weigh out 100 mg of sample and cut into fine pieces. Combine with 1 mL of nanopure water. Shake at 40 °C for 1 h. Agitate in an ultrasonic bath for 15 min at 40 °C. Shake at 40 °C for another 1 h. To eliminate large particles, centrifuge solution at 3080 × *g* at 4 °C for 30 min. Collect liquid supernatant (approx. 0.9 mL), skip **steps 4–6** and proceed to **step 7**.
4. Skim milk samples by centrifugation at 16,000 × *g* for 15 min at 4 °C (*see* **Note 3**).
5. Carefully insert a thin pipette tip below the upper lipid layer (cream) and collect the infranate (skim milk). If necessary, repeat the centrifugation procedure to remove any residual cream.
6. Take 25 μL of the skim milk sample. Add 100 μL of nanopure water (*see* **Note 4**).
7. For some sample types, like gastric digesta and cheese, removing any remaining lipids by applying a Folch liquid-liquid extraction avoids LC column clogging and signal deterioration (*see* **Note 5**). Add four times the sample volume of Folch solution to the sample. Mix with a vortex for 10 s. Centrifuge at 16,000 × *g* for 15 min at 4 °C. If the upper phase is not clear, centrifuge again until clear. Collect the top layer (MeOH and water). Be careful not to collect any of the middle protein layer. Discard the middle protein layer and bottom chloroform/lipid layer. Dry collected upper layer by centrifugal evaporation at 44 °C. Rehydrate samples in 100 μL nanopure water and mix using a vortex until completely solubilized.
8. Precipitate intact proteins from the samples by acid precipitation. Add 1:1 v/v (sample to solution) 200 g/L trichloroacetic acid, mix using a vortex for 5 s, then centrifuge at 4000 × *g* for 10 min at 4 °C. Collect the supernatant and discard the protein pellet.

9. Clean up extracted peptides via C18 solid-phase extraction. This step eliminates sugars, salts, and TCA. To prepare the 96-well plates, first wash them by adding 200 μL of the activation solution to each well and centrifuging at 2800 × *g* for 30 s. Repeat this step for a total of three times. Next, equilibrate the columns by washing three times with equilibration solution using 200 μL each time at the same centrifuge speeds and times. Next, add all of each sample to a well (200 μL) and centrifuge. To wash off salts, sugars, and TCA, wash three times with 200 μL wash solution. Finally, elute three times using 200 μL peptide elution solution and centrifuge each time. Collect this fraction.
10. Transfer eluted samples to 1.5-mL tubes and dry with a vacuum concentrator or freeze drier. Rehydrate peptides by adding 40 μL of 2% ACN, 0.1% TFA in nanopure water, and mix with a vortex for 1 min. Store at −80 °C until use for MS analysis.
11. *Optional step*: At this point, you may want to measure the extracted peptide concentration to determine how much sample to inject for LC-MS (*see* Subheading 3.3). We commonly used the Bradford assay, the bicinchoninic assay, or absorbance at 280 nm.

### 3.3 Mass Spectrometry

1. Add samples to sampling vials.
2. Set up analysis parameters for LC-MS. Set up the elution gradient using solvent (A) 0.1% FA and solvent (B) 100% ACN. Design a 60-min gradient as follows: 5–35% B over 50 min, 35–80% B over 3 min, 80% B for 1 min, 80–5% B over 1 min, and then hold at 5% B for 5 min. Flow rate: 300 μL/min. Set to collect mass spectra in data-dependent mode with one MS precursor scan followed by 15 MS/MS scans, using dynamic exclusion of 20 s. Set MS spectral acquisition resolution to 70,000 and a target of $1 \times 10^6$ ions or a maximum injection time of 30 ms. Set MS/MS spectral resolution to 17,500 with a target of $5 \times 10^4$ ions or a maximum injection time of 50 ms. Apply higher-energy collision dissociation with a normalized collision energy value of 27% for peptide fragmentation. Exclude unassigned charge states as well as ions >+7 from MS/MS fragmentation.
3. Load approximately two micrograms of each sample onto the enrichment column for online desalting and then onto the analytical column for analytical separation.
4. After each sample, perform a column wash using a blank sample containing 0.1% FA. Wash gradients will depend on sample type (*see* **Note 6**). We use a saw-tooth gradient as follows: 5–80% B over 6 min, 80% B for 7 min, 80–5% B over 2 min, 5% B for 3 min, 5–80% B over 6 min, 80% B for 7 min, 80–5% B over 2 min, 5% B for 7 min.

### 3.4 Build a Custom Protein Database

1. To perform a database search for peptide identification, several premade databases are available in the various software programs. However, searching for peptides with no cleavage specificity (as is required in peptidomics) greatly increases the number of possible peptides, increasing the search space and slowing the search from minutes or hours to days when searched against a full organism-level protein library. We recommend creating a custom milk-specific and species-specific protein library for your searches to decrease search time. This library can be made by gathering the fasta format sequences from www.uniprot.org based on available proteomics literature. At www.dallaslab.org, we have human, cow, and pig milk protein databases available for use (select the tab "Resources").

### 3.5 Convert Data to .mgf Format with MSconvert

1. If proceeding with X! Tandem analysis for peptide identification, raw data needs to be converted to mgf files. In order to do this, install ProteoWizard (http://proteowizard.sourceforge.net/tools.shtml). Open MSConvert, select your raw files and the location for saving the new mgf files. Select output format as mgf. Use all other settings as the defaults, making sure not to include any filters.

### 3.6 Peptide Identification with X! Tandem Database Search

1. For peptide identification via database searching (*see* **Note 7**), we use the downloadable GUI version of X! Tandem, called GPM Manager (http://www.thegpm.org/TANDEM/instructions.html). Within this program, we work inside the "advanced" tab in the "directory" section.
2. Select the files for analysis. Select "yes" to Skyline data file annotation to allow the results to import correctly into Skyline.
3. Select the protein library you wish to search against. If the library you want to search against is unavailable, this file must be added to the folder GPM Fury/thegpm/fasta and added to the files taxonomy.xml and species.js within the folder GPM (/Fury/thegpm/tandem).
4. Apply a peptide log($e$) value of less than −2. The "e" stands for "expectation value." A − 2 log($e$) corresponds to an e-value threshold of ≤0.01 (a 99% confidence level threshold).
5. For measurement errors, in case of Orbitrap data, allow 10–20 ppm for the fragment mass and 10–20 ppm for the parent mass. Allow isotope error, which allows for identification when the instrument fragments an isotope rather than the parent ion. Select "fragment type" as monoisotopic.
6. For signal processing, select a maximum parent charge of 7, a minimum parent M+H of 275, a minimum fragment $m/z$ of 50, total peaks for evaluation as 50, the minimum number of peaks for evaluation as 15 and fragment types b and y (*see* **Note 8**).

7. For protein modifications, disallow all complete (required) modifications, as the procedure did not make any chemical changes to the peptides. Allow for oxidation of methionine as a potential modification. Select "no" for using sequence annotations and "yes" to allowing protein N-terminal acetylation.
8. For refinement, select "yes" to "refine model," "no" to "use sequence annotations," "no" to "use point mutations," "no" to "use single amino acid polymorphisms" and allow round 1 modifications to include serine and threonine phosphorylation and possibly asparagine and glutamine deamidation. Select "no" for "semi-cleavage," "no" for "use mods throughout." Use a valid expectation of <−2.
9. For protein cleavage specification, select "No Enzyme [X]|[X]," which must be added into the source code (*see* **Note 9**). Leave the other settings as defaults.
10. Search spectra.

### *3.7 Label-Free Peak Quantification with Skyline*

1. Install Skyline [25] at https://skyline.gs.washington.edu/labkey/project/home/begin.view and open. Create a blank document and save as a .sky skyline file.
2. Go to Settings > Peptide Settings and click the "Modifications tab." Select the structural modification you would like to include: typically, we use oxidation (M), deamidation (NQ), and phosphorylation (S,T) (*see* **Note 10**).
3. Click the "Library tab" and then click "Build …." Insert a name for your library and specify a save location. Click "next" and then click "Add files" to select the files you wish to analyze. For our protocol, we use the X! Tandem output files (.xml), but other file types are possible. These files must be manually renamed to have the file extension .xtan.xml instead of .xml. To find your .xml files, go to GPM Fury/thegpm/gpm/archive. After importing your .xtan.xml files, place a checkmark in the field "include ambiguous results," click "finish," and wait for the library to build. Select your new library from the list and select "pick peptides matching library" and leave "rank peptides by" blank. On the filter tab, set min length as 4 and max length as 50. Click "OK."
4. Click View → Spectral libraries. Click "Add all…" and click "add all/include all" in the following boxes.
5. In Settings → Transition settings, go to the Prediction tab and select precursor mass: monoisotopic; product ion mass: monoisotopic; collision energy: Thermo TSQ Vantage; declustering potential: none; optimization library: none.
6. In the filter tab, fill in precursor charges 1–7; ion charge 1; ion type p (for precursor) (*see* **Note 11**).

7. For the library tab, select ion match tolerance 0.5 *m/z*; check the box "if a library spectrum is available, pick its most intense ions"; pick 6 product ions from filtered ions charges and types.
8. In the instrument tab, select min *m/z*: 50; max *m/z*: 1600; method match tolerance *m/z*: 0.055.
9. For the Full-Scan tab, select isotopic peaks included: count; precursor mass analyzer: Orbitrap; peaks: 3; resolving power: 60,000 at 400 *m/z*; MS/MS filtering acquisition method: none; retention time filtering: use only scans within 1 min of MS/MS IDs.
10. Go to File > Import > Results to add your RAW data file to peak volume extraction. Select "add single-injection replicates in files" and select your .raw files. After confirming, Skyline will extract the peak areas for each identified peptide in all samples.
11. If possible (depending on the number of samples and number of peptides identified), manually inspect for appropriate peak picking of the MS1-filtered peptides. Peaks that do not match criteria or are too close to the noise level to be visually discernable can be deleted from the data set. The criteria we typically use for a match are a ≤10 ppm mass error and an idotp (isotope dot product) score ≥80. This filter step also can be done within Excel in the exported results from the following step.
12. To export the results, go to File > Export > Report. Select "edit list…" and in the new window click "import" and load the file "Peptidomics_standard_output.skyr." Click "OK." This standard file can be downloaded from our website dallaslab.org under Resources. Select export report as "Peptidomics_standard_output" and export as a .tsv file (tab-separated).

### *3.8 Alternate Peptide Identification and Label-Free Quantification Approach with Proteome Discoverer (v2.1.0.81)*

1. Open Thermo Proteome Discoverer.
2. Go to file > New study/analysis. Fill in Study Name and Study Root Directory.
3. Select a consensus workflow. You can create or modify a new workflow by dragging new nodes into your workflow tree. Our consensus workflow can be viewed in Fig. 2.
4. In the consensus workflow, keep settings not described as default.
5. In the peptide and protein filter node, select "Peptide Confidence at Least": High. Set "Minimum Peptide Length" to 4.
6. In the Peptide and Protein Annotation node, set the "Annotate Flanking Residues of the Peptide" to "True." Set "Protein Modifications Reported" to "For All Proteins." Set "Protein Position for Peptides" to "For All Proteins."

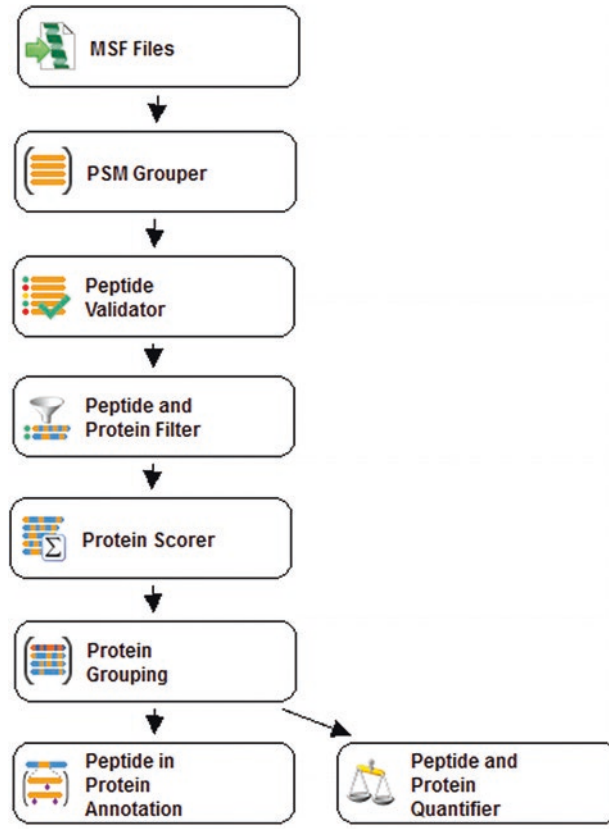



**Fig. 2** Overview of nodes in Proteome Discoverer consensus workflow to conduct milk peptidomics

7. Select a processing workflow. Our processing workflow is shown in Fig. 3.
8. In the processing workflow, keep settings not described as default.
9. In the Event Detector node, set "Mass Precision" to 2 ppm.
10. In the Spectrum Selector node, set "Min. Precursor Mass" to 300 Da. Set the "Max Precursor Mass" to 5000 Da.
11. In the Sequest HT node, set the protein database to your custom database (examples for human, cow, and pig milk can be found on www.dallaslab.org under "Resources"). Custom databases can be added through the "maintain fasta file" icon in the menu bar then press add and select your .fasta file. Back in the Sequest HT node, set Enzyme Name to No-Enzyme (Unspecific). Set "Min. Peptide Length" to 4. Set "Max Peptide Length" to 144 (the maximum allowed by Proteome Discoverer). Set "Precursor Mass Tolerance" to 10 ppm. Set "Fragment Mass Tolerance" to 0.8 Da. Allow oxidation of Met and phosphorylation of Ser and Thr as "Dynamic Modifications" 1 and 2, respectively.

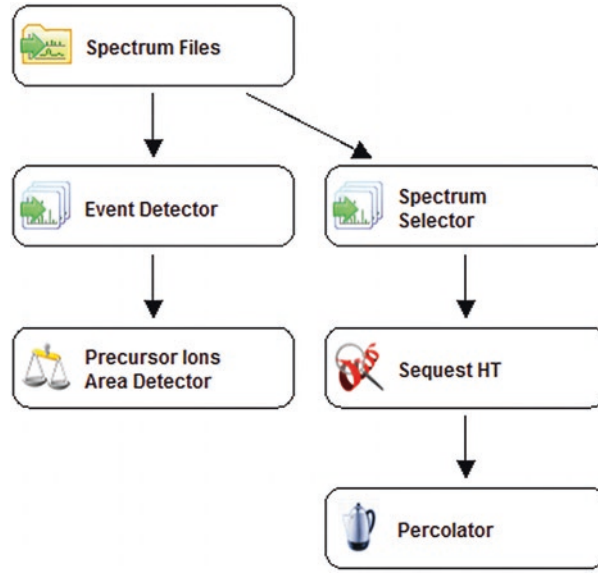



**Fig. 3** Overview of nodes in Proteome Discoverer processing workflow to conduct milk peptidomics

12. In the Percolator node, set "Validation based on" to q-Value. Quantification method is "(no quantification)" for label-free quantification. Click "OK."
13. In the "Input Files" tab > add files > and select your files (*see* **Note 12**). Select all your files and drag them to the "Input Files" field on the right. Click "Run."
14. When the search is finished, go to your study tab > then analysis results tab and double click on the result from your run to open it. In the new window you can examine your results.
15. To export your results, in the results window, go to file > export > to Microsoft Excel… > Choose a path for your file and save your results. Export a file with protein groups and one with peptide groups.

### 3.9 Peptide Mapping

1. To better visualize peptides identified, map the peptides to the protein sequences. Mapping where the fragments from a protein derived in relation to the overall sequence can support biological insight into the enzymatic processes occurring in a system. Previously we used PepEx to map the endogenous peptides in human milk and revealed that the release of peptides was highly specific to regions of the parent protein [11].
2. To access the tool to map the peptides, go to http://mbpdb.nws.oregonstate.edu and choose PepEx.
3. Click "PepEx Add Fasta files." Insert the .fasta file that you used for your X!Tandem search.

| | A | B | C | D | E | F | G | H | I | J | K | L |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Protein | Peptide | Precursor | Precursor Charge | Precursor Mz | Peptide Modified Sequence | Peptide start | Peptide end | Modifications | Library Name | Sample1 Total Area | Sample2 Total Area |
| 2 | sp\|P02666\|CASB_BOVIN | DMPIQAF | 821.4+ | 1 | 821.4 | DMPIQAF | 199 | 205 | | Milk | 1.98E+10 | 2.15E+09 |
| 3 | sp\|P02662\|CASA1_BOVIN | DQAMEDIKQM | 612.8++ | 2 | 612.8 | DQAMEDIKQM[+16] | 66 | 75 | 1O | Milk | 3.51E+10 | 2.53E+08 |
| 4 | | | | | | | | | | | | |

**Fig. 4** Example of a PepEx input file (.tsv) made inside Excel

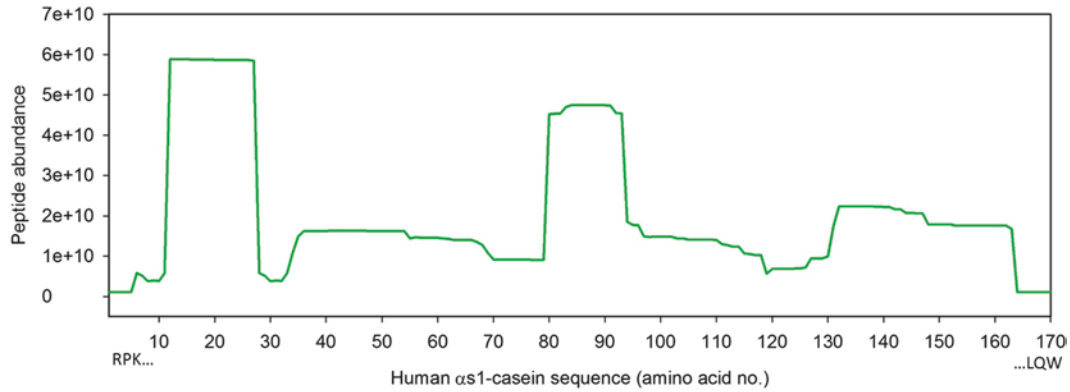



**Fig. 5** Example of how peptidomics data can be used in PepEx to map the relative abundance of peptides identified in an LC/MS run onto the sequence of human alpha s1-casein (not real data)

4. A .tsv file should be made with the following columns and information on your peptides (Fig. 4). Column A, Protein; column B, Peptide; column C, Precursor; column D, Precursor Charge; column E, Precursor Mz; column F, Peptide Modified Sequence; column G, Peptide start; column H, Peptide end; column I, Modifications; column J, Library Name; column K, Sample1 Total Area; column L, Sample2 Total Area (Fig. 4). Only columns A, G, H, and K need to be filled. You can add more samples adjacent to column L.
5. This program will output a report in Excel where, for each protein, the total abundances of peptides at each amino acid in the sequence are shown. Remember the result file comes as a .txt file, which needs to be opened with comma separation. Use Excel to make line graphs plotting the abundances (*y*-axis) against the peptide sequence (*x*-axis) to allow visualization (Fig. 5).

### *3.10 Enzyme Predictor*

1. Enzyme predictor (http://bioware.ucd.ie/~enzpred/Enzpred.php) is an online bioinformatic tool [26] that predicts which enzymes most likely cleaved the sample proteins into the identified peptides. The predictions are based on a comparison of the protease cleavage specificity with the cleavage sites at both ends of each identified peptide.

2. Upload a tab-delimited text file with two columns ("prot_acc" and "pep_seq") containing the protein accession number from www.uniprot.org and the peptide amino acid one letter sequence in each row, respectively (*see* **Note 13**).
3. The output file shows how many identified cleavages mapped to each protease.

### 3.11 Peptidomics Enzyme Estimator

1. Peptidomics Enzyme Estimator [11] is another program for estimating protease activity based on peptidomics data. This program allows more analysis options than Enzyme predictor. For example, you can use either peptide count or abundance, and you can add new proteases (and their specificity patterns) as needed.
2. Download Peptidomics Enzyme Estimator from the eparker05 Github repository (https://github.com/eparker05/Peptidomics-enzyme-estimator) and save it to your python directory. Peptidomics Enzyme Estimator requires that Python 2.7 and Biopython are installed (*see* **Note 14**).
3. Create a .csv file with your identified peptide information. Row 1 in Column A should be "Sequence," in column B "Intensity," in Column C "Protein_id," and in column D "Sample_id." In the Sequence column, add the single amino acid sequence of your peptide. In the Intensity column, add the quantification. In the Protein_id, add the protein information (should be in fasta format, i.e., sp.|P05814|CASB_HUMAN). In the Sample_id column, add the name of your sample. The program output distinguishes between the results from each sample_id so that many samples can be analyzed simultaneously.
4. In Python IDLE, use the shell window and import the Peptidomics Enzyme Estimator module using the code:

   *import PeptidomicsEnzymeEstimator as pee*
5. Open the relevant library files:

   *fastaFile = open("YourProteinLibrary.fasta", "rU")* # (*see* **Note 15**)

   *csvFile = open("YourPeptidomics.csv", "rU")*

   *inputEnzymeList = ["Enzyme1," "Enzyme2"]* # (*see* **Note 16**)
6. Load and preprocess the peptides and retrieve the results as a .csv file:

   *peptideList = pee.import_peptides_and_preprocess(csvFile, fastaFile, inputEnzymeList)*

   *resultCSV=pee.extract_data_from_processed_peptides(peptideList, inputEnzymeList, result = "list")*

   *import csv*

   *with open('output.csv', 'wb') as csvfile:*

   *resultWriter = csv.writer(csvfile)*

   *resultWriter.writerows(resultCSV)*

The results will show how much of the peptide profile (by count or abundance) was mapped to the proteases you searched against. These results will be output to a .csv file "output.csv" located in your Python directory.

### 3.12 Database Search

1. Milk contains a large number of functional peptides with a wide range of biological activities. One method for predicting peptide function is through a simple homology search against a database of known functional peptide sequences to determine whether the peptides identified in a sample have been previously associated with a biological action.
2. We have collected a complete database of human milk- and dairy-derived bioactive peptides. This database can be accessed at http://mbpdb.nws.oregonstate.edu. Select the Resources tab, which will direct you to a list of peptide tools. Choose the milk bioactive peptide database.
3. The search function in this database includes a search for the specific peptide sequence, specific protein, species, or function. It is possible to search a single peptide sequence or multiple peptide sequences. Furthermore, the search function for peptide sequence includes three search options. One option searches for bioactive peptides matching the input peptide sequence. The second option searches for bioactive peptides that contain the input peptide sequence. The third option searches for bioactive peptides contained within the input peptide sequence. These three search options can be combined with a similarity search option that allows identification of peptides with minor sequence modifications. This option is useful as homologous peptides may retain the functionality of the original peptide.

### 3.13 Antimicrobial Prediction

1. Antimicrobial activity of peptides is one of the main biological functions studied from milk-derived peptides. We apply CAMPR3 [27], a collection of antimicrobial peptide (AMPs) prediction tools (http://www.camp.bicnirrh.res.in/) to identify potential antimicrobials. Under the Tools menu, select AMP prediction. Choose predict antimicrobial peptides in the opening window. Input your peptide list in a fasta file format or upload it. Click "submit."
2. The program will produce a list with a score for the likelihood that each peptide is antimicrobial. A score above 0.5 is a positive score.
3. The database of antimicrobial activity and structure of peptides (http://dbaasp.org/home.xhtml) is another database of antimicrobial peptides that contains a tool for AMP prediction under the prediction tab.

4. For this search, paste in your fasta file and click "submit." The program determines whether or not each peptide is potentially antimicrobial.

### 3.14 Data Sharing

1. After peptides are identified, we suggest that you upload your data to ProteomeXchange [28] or a similar program for public use. Public data sharing is very important for omics data in particular.

## 4 Notes

1. If possible, add the sample directly to a mixture of protease inhibitors to prevent any proteolytic action, mix, and then freeze.
2. Unless extensively heat-treated or previously treated with protease inhibitors, breast milk and dairy samples will contain active proteases, so avoid keeping the samples at room temperature as it may change the peptide profiles.
3. Cold centrifugation aids separation of the cream layer.
4. We added 100 μL of nanopure water to assist in dissolving thick samples, such as digestive samples, kefir, yogurt, or cheese. If only milk is being examined, this addition is not necessary. However, in studies where we have compared milk to digestive samples, we have added 100 μL of water to keep sample volumes constant for relative quantification.
5. The Folch step may not be necessary, depending on the sample type. For example, we did not employ this step for kefir peptidomics [18].
6. Samples vary in how difficult they are to fully elute from the column. Therefore, it is essential to test different blank method gradients and verify that all residual peptides are removed during the blank assessment to avoid contaminating the next sample.
7. Database searching matches tandem spectra by comparison with theoretical spectra derived from predicted peptides in a protein library. Several software tools are available for identification and quantification of peptides from the spectra obtained with mass spectrometry.
8. For collision-induced dissociation (CID) and higher-energy collision dissociation (HCD) commonly used in peptide and protein mass spectrometry, "b" and "y" type ions are the most common.
9. Add the X|X enzyme in the file g_pcs.js in the folder GPM Fury/thegpm/tandem. Bottom-up proteomics employs proteolytic enzymes, such as trypsin, with high specificity, which

allows searching against only peptides matching those specificity patterns. In peptidomic analysis, peptides are cleaved by an array of often-unknown endogenous proteolytic enzymes, which requires searching against all possible peptide fragments.

10. Our saved settings file ("PeptideRelativeQuantSettings.skys") for peptide relative quantification can be downloaded at www.dallaslab.org and imported via Settings > Import.
11. Selecting "p" means we are searching based only on precursor ions and isotopes and not on the products in the particular spectra. Since we are only searching the precursor ions, the product ions box can be ignored.
12. You may want to note the ID of your sample as this ID will appear in your exported result file.
13. An example input file can be viewed at our homepage.
14. Biopython is a package for Python, which is required for PeptidomicsEnzymeEstimator. To install Biopython using pip for Windows users, open the command prompt. Locate your python directory using the change directory command (cd). When there, type in the command to install Biopython "*python -m pip install biopython*."
15. This file is the same protein library used for the original database search in X!Tandem or Proteome Discoverer.
16. The proteases already defined in Peptidomics Enzyme Estimator are: Arg-C proteinase, Asp-N endopeptidase, BNPS-Skatole, Chymotrypsin specific, Chymotrypsin low-spec, Pepsin, Plasmin, Cathepsin D, Cathepsin B, Thrombin, Elastase, Trypsin, and Thrombin-OSP. In the *inputEnzymeList = ["Enzyme1", "Enzyme2", "*etc...*"]* code, substitute Enzyme1 and Enzyme2 with the desired enzyme (e.g., Elastase or Plasmin). Continue adding as many enzymes as needed from those defined in the database. The source code can also be modified to add additional enzyme cleavage patterns.

## Acknowledgments

The authors thank C. J. Dillard for editing this manuscript. The authors gratefully acknowledge funding from the National Institutes of Health, Eunice Kennedy Shriver Institute of Child Health and Development (4R00HD079561) R00 Pathway to Independence Career Award.

## References


1. Migliore-Samour D, Floch F, Jollès P (1989) Biologically-active casein peptides implicated in immunomodulation. J Dairy Res 56:357–362
2. Jørgensen ALW, Juul-Madsen HR, Stagsted J (2010) Colostrum and bioactive, colostral peptides differentially modulate the innate immune response of intestinal epithelial cells. J Pept Sci 16:21–30
3. Brantl V (1984) Novel opioid peptides derived from human beta-casein: human beta-casomorphins. Eur J Pharmacol 106:213–214
4. Kampa M, Loukas S, Hatzoglou A et al (1996) Identification of a novel opioid peptide (Tyr-Val-Pro-Phe-Pro) derived from human alpha S1 casein (alpha S1-casomorphin, and alpha S1-casomorphin amide). Biochem J 319:903–908
5. Aniansson G, Andersson B, Lindstedt R, Svanborg C (1990) Antiadhesive activity of human casein against Streptococcus pneumoniae and Haemophilus influenzae. Microb Pathog 8:315–323
6. Stromqvist M, Falk P, Bergstrom S et al (1995) Human-milk k-casein and inhibition of Helicobacter pylori adhesion to human gastric mucosa. J Pediatr Gastroenterol Nutr 21:288–296
7. Liepke C, Zucht H-D, Forssmann W-G, Ständker L (2001) Purification of novel peptide antibiotics from human milk. J Chromatogr B Analyt Technol Biomed Life Sci 752:369–377
8. Yamada A, Sakurai T, Ochi D et al (2015) Antihypertensive effect of the bovine casein-derived peptide Met-Lys-Pro. Food Chem 172:441–446
9. Suetsuna K, Ukeda H, Ochi H (2000) Isolation and characterization of free radical scavenging activities peptides derived from casein. J Nutr Biochem 11:128–131
10. Dallas DC, Guerrero A, Khaldi N et al (2013) Extensive in vivo human milk peptidomics reveals specific proteolysis yielding protective antimicrobial peptides. J Proteome Res 12:2295–2304
11. Guerrero A, Dallas DC, Contreras S et al (2014) Mechanistic peptidomics: factors that dictate the specificity on the formation of endogenous peptides in human milk. Mol Cell Proteomics 13:3343–3351
12. Dallas DC, Guerrero A, Khaldi N et al (2014) A peptidomic analysis of human milk digestion in the infant stomach reveals protein-specific degradation patterns. J Nutr 144:815–820
13. Combes C, Paterson E, Amadò R (2002) Isolation and identification of low-molecular-weight peptides from Emmentaler cheese. J Food Sci 67:553–559
14. Toelstede S, Hofmann T (2008) Sensomics mapping and identification of the key bitter metabolites in Gouda cheese. J Agric Food Chem 56:2795–2804
15. Gupta A, Mann B, Kumar R, Sangwan RB (2010) Identification of antioxidant peptides in cheddar cheese made with adjunct culture Lactobacillus casei ssp. casei 300. Milchwissenschaft 65:396–399
16. Sforza S, Cavatorta V, Lambertini F et al (2012) Cheese peptidomics: a detailed study on the evolution of the oligopeptide fraction in Parmigiano-Reggiano cheese from curd to 24 months of aging. J Dairy Sci 95:3514–3526
17. Miclo L, Roux E, Genay M et al (2012) Variability of hydrolysis of β-, αs1-, and αs2-caseins by 10 strains of Streptococcus thermophilus and resulting bioactive peptides. J Agric Food Chem 60:554–565
18. Dallas DC, Citerne F, Tian T et al (2016) Peptidomic analysis reveals proteolytic activity of kefir microorganisms on bovine milk proteins. Food Chem 197:273–284
19. Rauh VM, Johansen LB, Ipsen R et al (2014) Plasmin activity in UHT milk: relationship between proteolysis, age gelation, and bitterness. J Agric Food Chem 62:6852–6860
20. Jensen S, Sidsel J, Therese J et al (2015) Storage-induced changes in the sensory characteristics and volatiles of conventional and lactose-hydrolyzed UHT processed milk. Eur Food Res Technol 240:1247–1257
21. Nielsen SD, Jansson T, Le TT et al (2017) Correlation between sensory properties and peptides derived from lactose-hydrolyzed UHT milk during storage. Int Dairy J 68:1–108
22. Dallas DC, Guerrero A, Parker EA et al (2013) Peptidomic profile of milk of Holstein cows at peak lactation. J Agric Food Chem 62:58–65
23. Guerrero A, Dallas DC, Contreras S et al (2014) Peptidomic analysis of healthy and subclinically mastitic bovine milk. Int Dairy J 46:46–52
24. Dallas DC, Guerrero A, Parker EA et al (2015) Current peptidomics: applications, purification, identification, quantification and functional analysis. Proteomics 15:1026–1038
25. Schilling B, Rardin MJ, MacLean BX et al (2012) Platform-independent and label-free quantitation of proteomic data using MS1 extracted ion chromatograms in Skyline

application to protein acetylation and phosphorylation. Mol Cell Proteomics 11:202–214
26. Vijayakumar V, Guerrero AN, Davey N et al (2012) EnzymePredictor: a tool for predicting and visualizing enzymatic cleavages of digested proteins. J Proteome Res 11:6056–6065
27. Waghu FH, Barai RS, Gurung P, Idicula-Thomas S (2016) CAMPR3: a database on sequences, structures and signatures of antimicrobial peptides. Nucleic Acids Res 44:D1094–D1097
28. Vizcaíno JA, Côté RG, Csordas A et al (2013) The Proteomics Identifications (PRIDE) database and associated tools: status in 2013. Nucleic Acids Res 41: D1063–D1069

# Chapter 16

# Neuropeptidomic Analysis of Zebrafish Brain

**Kristien A. Van Camp, Geert Baggerman, Ronny Blust, and Steven J. Husson**

## Abstract

A wide variety of bioactive peptides are present in all metazoan species where they govern diverse functions as small messenger molecules. In the last 15 years, mass spectrometry-based methods have identified endogenous peptides in diverse species. Mass spectrometry enables the precise peptide sequences to be determined, including the potential existence of truncated versions or the presence of post-translational modifications. Because small modifications can have a large effect on biological activity, knowledge of the actual peptide sequences paves the way for further functional studies such as analysis of neuropeptidergic signaling cascades. Zebrafish (*Danio rerio*) is an important animal model that is commonly used in a wide range of studies. Here we provide a detailed description of the peptide extraction procedure and peptidomics workflow for zebrafish.

**Key words** Zebrafish, *Danio rerio*, Neuropeptide, Peptide, Peptidomics, LC-MS, Mass Spectrometry

## 1 Introduction

The zebrafish (*Danio rerio*) is a commonly used model system for a wide variety of research disciplines, providing fundamental insights into genetics, developmental biology, neuroscience, and ecotoxicology. As a vertebrate model, zebrafish share substantial genetic and physiological homology with mammals, including underlying molecular players like small-molecule neurotransmitters, neuropeptides, hormones, receptors, and ion channels. In addition, zebrafish exhibit several behaviors that are functions of external cues or stress conditions (change in water temperature or oxygen levels, presence of food or predators, exposure to chemical compounds, etc.). From a neurobiological point of view, it is extremely interesting to correlate behavioral responses or physiological processes with underlying molecular pathways or signaling cascades, especially when modulatory effects on both the behavioral and molecular level can be studied upon exposure to external cues.

**Funding**
The authors highly appreciate funding from the University Research Fund (Bijzonder Onderzoeksfonds, BOF).

Michael Schrader and Lloyd Fricker (eds.), *Peptidomics: Methods and Strategies*, Methods in Molecular Biology, vol. 1719, https://doi.org/10.1007/978-1-4939-7537-2_16, 

Most neuropeptides are derived from inactive preproproteins or peptide precursors that harbor one or multiple peptide sequences. As such, the peptide precursors have to undergo extensive post-translational processing to produce the bioactive peptide entities. The resulting biologically active peptides typically interact with G-protein-coupled receptors to initiate an intracellular signaling pathway. This orchestrates a diverse array of physiological processes and behaviors such as feeding (e.g., melanocortin [1], pituitary adenylate cyclase-activating polypeptide [2], neuropeptide Y [3]), locomotion (corticotropin-releasing factor [4], hypocretin/orexin [5]), and reproduction; the latter is largely regulated by the hypothalamic-pituitary-gonadal axis in which gonadotropin-releasing hormone plays a central role. Therefore, disturbances in peptide homeostasis are often reflected on the level of overall physiology and behavior.

In analogy with proteomics, a peptidomics approach aims at the simultaneous identification and characterization of all endogenously present peptides in a tissue or organism by using liquid chromatography and mass spectrometry (LC-MS). Indeed, the emergence of high-throughput peptide sequencing via tandem mass spectrometry allows fast identification and characterization of peptides, including their post-translational modifications which cannot be predicted from nucleotide sequences in genomic databases. Doing so, we previously acquired experience with different organisms—ranging from nematodes to insects and crustaceans—in isolating peptides (*see* for example [6–8]). In these studies, peptides were specifically extracted from the tissue using an acidified methanolic solvent that is extremely efficient in avoiding the presence of proteins and their degradation products.

While equally important, no such high-throughput analyses have been performed on the freshwater teleost zebrafish (*Danio rerio*) until recently. Because a comprehensive overview of endogenously present peptides from the zebrafish brain was lacking, we set out to biochemically identify these peptides using a peptidomics approach [9]. We used LC-MS to yield a general inventory of endogenous brain peptides of the zebrafish. Because peptides are selected in a semi-stochastic way in data-dependent LC-MS analysis, repeated runs will increase the number of peptide identifications. In addition, the endogenous peptide content of a tissue or organism is highly dynamic, which also has to be taken into account when aiming to monitor a "general" peptide profile. It is important to mention that we encountered remarkable inter-individual variations, likely due to the spatial and temporal dynamics of the peptidome. Full details of the identified peptides are described and discussed in Van Camp et al. [9]. It is expected that this list of zebrafish peptides will pave the way for future (neuro)endocrine research by aiding further functional studies of defined peptidergic signaling pathways. In this chapter, we provide a detailed description of the materials and methods required for analysis of the zebrafish peptidome.

## 2 Materials

### 2.1 Sample Preparation

1. Adult wild-type zebrafish.
2. Vannas micro scissor.
3. Precision tweezers (such as Dumont tweezers).
4. Peptide extraction solvent: methanol, water, acetic acid (90:9:1; v:v:v %).
5. Sonicator (such as the Branson Sonifier SLPe cell disruptor)
6. Centrifuge (such as the Eppendorf microcentrifuge 5415R).
7. Vacuum centrifuge (such as the Eppendorf 5301 concentrator centrifugal evaporator).
8. $C_{18}$ spin columns (such as those from Pierce, Thermo Scientific).
9. Water.
10. Ethyl-acetate.
11. n-hexane.
12. Methanol.
13. Acetonitrile (ACN).
14. Formic acid (FA).
15. Column activation solution: 50% ACN in water.
16. Column equilibration solution: 5% ACN, 95% water containing 0.1% FA.
17. Column elution solution: 60% ACN, 40% water containing 0.1% FA.

### 2.2 Peptidomics Analysis by Liquid Chromatography and Mass Spectrometry

1. Nanoscale liquid chromatography system (nanoUPLC, such as the Eksigent nanoLC-Ultra System).
2. Reversed-phase (RP) C18 trapping column (such as the Pepmap C18 300 μm × 20 mm, Dionex).
3. RP C18 analytical column (such as the Pepmap C18 3 μm 75 μm × 150 mm, Dionex).
4. Electrospray Linear Trap Quadrupole Orbitrap Mass Spectrometer (such as the ESI-LTQ-Orbitrap Velos MS, Thermo Scientific, or comparable instrument).
5. Solvents: water, ACN, FA.
6. Peptide mass standard: Glu-1-fibrinopeptide B (Glu-Fib) or similar.
7. Mascot proteomics software (Matrix Science, London).
8. Proteome discoverer (1.3) software (Thermo Scientific).

## 3 Methods

### 3.1 Sample Preparation

1. Sacrifice the adult zebrafish by decapitation on ice. Dissect carefully the entire brain region of the zebrafish as quickly as possible and place the brain on the bottom of a 1.5 mL microfuge tube.
2. Immediately add 100 μL of ice cold extraction solvent containing methanol, water, acetic acid (90:9:1; v:v:v %) (*see* **Notes 1** and **2**) and place the sample on ice (*see* **Note 3**).
3. Next, sonicate the sample on ice three times for 10 s, with a 30 s interval between pulses.
4. Centrifuge for 15 min at 14,000 × *g* at 4 °C. Transfer the supernatant to a new tube and repeat the centrifugation step.
5. Transfer the supernatant to a new tube and evaporate the methanol using a vacuum centrifuge. This takes around 15 min in our case, but will depend of the volume and the vacuum centrifuge system. The goal of this step is to reduce the ~100 μL volume to approximately 10 μL, which is the aqueous portion that contains the endogenous peptides.
6. Delipidate the residue (~10 μL) by extraction with ethyl acetate and n-hexane. First, add 50 μL water containing 1% acetic acid. Then, add an equal volume (i.e., 60 μL) of ethyl acetate to the aqueous solution. Mix the solution by shaking the tube. Centrifuge for a short period at high speed (e.g., 1 min at 14,000 × *g*). Carefully remove the top organic layer and discard. Repeat this delipidation process 2 times with ethyl-acetate and one time with n-hexane.
7. Desalt the aqueous solution by solid phase extraction using $C_{18}$ spin columns according to the manufacturer's guidelines. In short, activate the resin using 50% of ACN and equilibrate resin using 5% ACN containing 0.1% FA. Add ACN and FA to the peptide sample so that the final concentration is around 5% ACN/0.1% FA. Then, bind the peptide sample to the resin. Next, wash the column using 5% ACN/0.1% FA and finally elute the peptides with 60% ACN containing 0.1% FA.
8. Dry the purified peptide sample using a vacuum centrifuge.
9. Reconstitute in 10 μL water containing 2% ACN and 0.1% FA.
10. Add a peptide mass standard to the sample (we made use of the Glu-Fib peptide mass standard as previously used [10]).

### 3.2 Peptidomics Analysis by Liquid Chromatography and Mass Spectrometry

A high-throughput LC-MS analysis is performed on a nanoUPLC system connected to an ESI-LTQ-Orbitrap mass spectrometer; similar platforms are also suited.

1. Load 5 μL of the peptide extract on the RP C18 trapping column with an isocratic flow of 2% ACN in water with 0.1% FA at a flow rate of 5 μL $min^{-1}$.
2. Next, place the trapping column online with the RP C18 analytical capillary column.
3. Separate the peptides using a linear gradient from 2% ACN in water/0.1% FA to 40% ACN in water/0.1% FA in 45 min at a flow rate of 350 nL $min^{-1}$.
4. The eluent of the nanoUPLC system is directly coupled to the electrospray interface (ESI) of the LTQ-Orbitrap MS. MS data are acquired by survey scans with a mass window of 350–5000 m/z and resolution 60,000. Next, a data-dependent acquisition method is selected where the 10 most abundant precursor ions (single-charged precursors are rejected) are subjected to fragmentation by collision-induced dissociation (CID) in the LTQ part of the instrument. The mass window of the MS2 fragmentation scans is 100–2000 m/z. The normalized collision energy is 35% in CID. Automated gain control (AGC) target value is set at $5 \times 10^4$ ions, maximum ion injection time is 100 ms, minimum signal threshold is 500 intensity units.
5. Perform database search against a proteome database (we use the NCBI database nr_20130601 filtered for taxonomy *Danio rerio*) using Mascot with Proteome discoverer software (*see* **Note 4**).

## 4 Notes

1. This extraction solvent is ideally suited to extract small endogenous peptides, while larger proteins precipitate together with the cellular debris. When interested in larger peptides (5–15 kDa) such as the insulin-like peptides, diluted acids might be a better extraction solvent.
2. All solvents should be "HPLC grade" and extremely pure. Lower quality solvents may contain contaminants that interfere with the detection of peptides during mass spectrometry.
3. All initial steps have to be performed on ice to reduce degradation of the peptides. Active peptidases that are present in the tissue might result in shortened and/or fragmented peptides. The goal of peptidomic analysis is to detect the forms of peptides present in the tissue, and post-extraction breakdown products are not of interest.
4. Mascot is a bioinformatics program that matches the fragmentation data from the peak list files against the selected proteome database. Following settings are recommended (which are also dependent on the hardware used): precursor mass tolerance and fragment mass tolerance are respectively 10 ppm

and 0.5 Da. “No enzyme” is specified as digesting enzyme. The variable modifications are pyroglutamic acid, carboxyterminal amidation, and methionine oxidation. Only medium confidence peptides (FDR < 5%), high confidence peptides (FDR < 1%), and first ranked peptides are included in the result file.

## References


1. Cerda-Reverter JM, Agulleiro MJ, RG R, Sanchez E, Ceinos R, Rotllant J (2011) Fish melanocortin system. Eur J Pharmacol 660:53–60
2. Sherwood NM, Adams BA, Isaac ER, Wu S, Fradinger EA (2007) Knocked down and out: PACAP in development, reproduction and feeding. Peptides 28:1680–1687
3. Matsuda K, Sakashita A, Yokobori E, Azuma M (2012) Neuroendocrine control of feeding behavior and psychomotor activity by neuropeptide Y in fish. Neuropeptides 46:275–283
4. Lowry CA, Moore FL (2006) Regulation of behavioral responses by corticotropin-releasing factor. Gen Comp Endocrinol 146:19–27
5. Panula P (2010) Hypocretin/orexin in fish physiology with emphasis on zebrafish. Acta Physiol (Oxf) 198:381–386
6. Husson SJ, Clynen E, Baggerman G, De Loof A, Schoofs L (2005) Discovering neuropeptides in *Caenorhabditis elegans* by two dimensional liquid chromatography and mass spectrometry. Biochem Biophys Res Commun 335:76–86
7. Husson SJ, Landuyt B, Nys T, Baggerman G, Boonen K, Clynen E, Lindemans M, Janssen T, Schoofs L (2009) Comparative peptidomics of *Caenorhabditis elegans* versus *C. briggsae* by LC—MALDI-TOF MS. Peptides 30:449–457
8. Baggerman G, Cerstiaens A, De Loof A, Schoofs L (2002) Peptidomics of the larval *Drosophila melanogaster* central nervous system. J Biol Chem 277:40368–40374
9. Van Camp KA, Baggerman G, Blust R, Husson SJ (2017) Peptidomics of the zebrafish *Danio rerio*: in search for neuropeptides. J Proteomics 150:290–296
10. Maes E, Valkenborg D, Baggerman G, Willems H, Landuyt B, Schoofs L, Mertens I (2015) Determination of variation parameters as a crucial step in designing TMT-based clinical proteomics experiments. PLoS One 10:1–15

# Chapter 17

# Identification, Quantitation, and Imaging of the Crustacean Peptidome

**Kellen DeLaney, Amanda Buchberger, and Lingjun Li**

## Abstract

Crustaceans serve as a useful, simplified model for studying peptides and neuromodulation, as they contain numerous neuropeptide homologs to mammals and enable electrophysiological studies at the single-cell and neural circuit levels. In particular, crustaceans contain well-defined neural networks, including the stomatogastric ganglion, esophageal ganglion, commissural ganglia, and several neuropeptide-rich organs, such as the brain, pericardial organs, and sinus glands. Due to the lack of a genomic database for crustacean peptides, an important step of crustacean peptidomics involves the discovery and identification of novel peptides and the construction of a database, more recently with the aid of mass spectrometry (MS). Herein, we present a general workflow and detailed methods for MS-based peptidomic analysis of crustacean tissue samples and circulating fluids. In conjunction with profiling, quantitation can also be performed with isotopic or isobaric labeling. Information regarding the localization patterns and changes of peptides can be studied via mass spectrometry imaging. Combining these sample preparation strategies and MS analytical techniques allows for a multifaceted approach to obtaining deep knowledge of crustacean peptidergic signaling pathways.

**Key words** Crustacean, Neuropeptides, Peptidome, Quantitation, MALDI mass spectrometry imaging, Microdialysis, Isotopic/isobaric labeling, De novo sequencing

## 1 Introduction

Crustacea is a diverse subphyla of arthropods including invertebrate animals that often live in aquatic environments. The most common members of this group include crabs, lobsters, crayfish, and shrimp. Crustaceans are a useful model for peptidomic studies because they possess a relatively simple nervous system with a rich repertoire of signaling peptides that modulate diverse functions of the animal. With numerous homologs to mammalian neuropeptides, crustaceans can provide a useful test-bed to determine how networks of neurons generate behavior under the actions of various peptide modulators and hormones. The crustacean stomatogastric nervous system includes several key tissues that are rich in neuropeptides and are the focal point of crustacean peptidomic

Michael Schrader and Lloyd Fricker (eds.), *Peptidomics: Methods and Strategies*, Methods in Molecular Biology, vol. 1719, https://doi.org/10.1007/978-1-4939-7537-2_17, 

studies. These tissues include the stomatogastric ganglion (STG), esophageal ganglion (OG), and a pair of commissural ganglia (CoG), which are all connected in series by motor nerves. The brain (e.g., the central nervous system) is connected to the OG by the inferior ventricular nerve (*ivn*). Other notable organs that play a crucial role in the neuronal signaling pathway are the sinus glands (SG), located in the eyestalks, and the pericardial organs (PO), located on either side of the heart. Figure 1 shows the location of important tissues in *Cancer borealis*, a commonly studied crustacean species. The neuroendocrine cells within these tissues release neuropeptide hormones into the circulating fluid (e.g., hemolymph) of the crustacean [1, 2]. By studying the chemical content of these tissues and the transporting hemolymph, a deep understanding of the simplified signaling pathways can be acquired, which can be later applied to more complex organisms.

While crustaceans serve as a useful model for understanding complex neuronal networks, they pose a challenge for peptidomic studies due to the lack of a genomic database. Extensive efforts have been put forth to generate in-house peptide databases based on experimental data and predicted sequences from transcriptomic data [3–5]. In recent years, extensive profiling has been performed on crustaceans, as well as in silico predictions based on transcriptomes [6], with novel peptides frequently being added to the growing database of crustacean peptides. Furthermore, BLAST homology searches are often performed to determine if identified peptide sequences are conserved across many species, allowing for possible functional information.

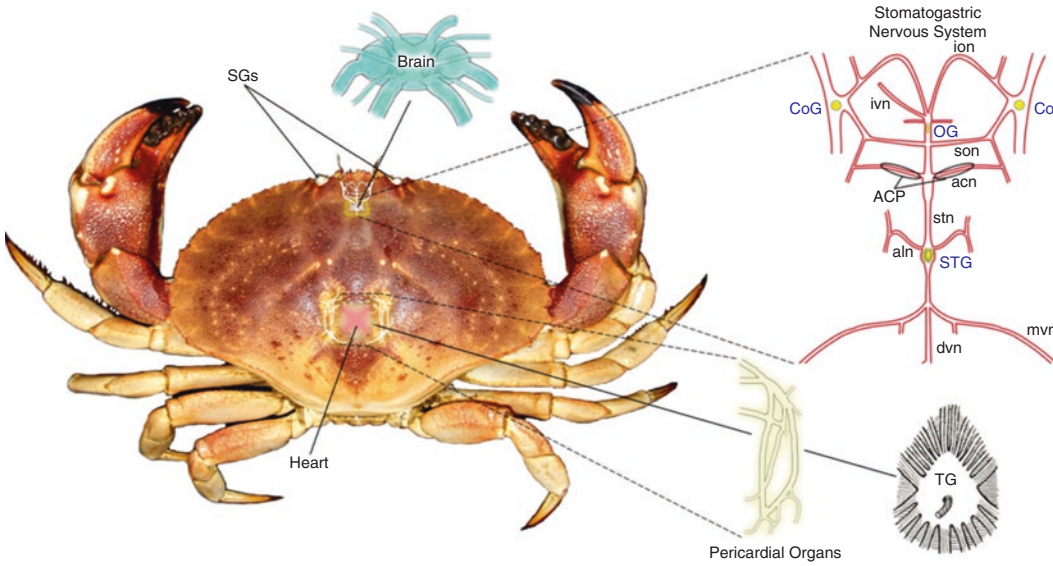



**Fig. 1** Schematic diagram of the location of tissues in *Cancer borealis* that are used for peptidomic studies, including the brain, CoGs, OG, STG, TG, POs, and SGs. The location of the heart is also indicated, which is important for probe placement in microdialysis studies

These efforts have been further driven forward by the development of mass spectrometry (MS)-based techniques for probing the crustacean peptidome. MS is an attractive choice for studying crustacean peptides because of its speed, selectivity, and sensitivity. High-resolution MS can detect minute mass differences of peptide isoforms, which can have varying functions within the crustacean nervous system. Furthermore, with MS, there is no need for prior knowledge about the analytes, making it ideal for discovery of novel peptides and building a peptide database.

In addition to identification methods, efforts have been directed toward studying the functions of crustacean neuropeptides. These species have been shown to be highly resilient to stress, making them ideal candidates for comparative studies involving environmental or biological stresses (e.g., temperature [7] and salinity changes [8]) as well as normal regulatory phenomena (e.g., feeding [9]). MS imaging has also been shown to be a useful strategy for mapping the location of neuropeptides within tissue and studying how these localization patterns differ in both two and three dimensions [10, 11].

Here we describe methods for profiling neuropeptides and quantifying dynamic changes of neuropeptidomes under physiological perturbations. These methods can be used to identify novel neuropeptides to expand a database, quantify changes in peptide expression as a result of biological activities, or explore the localization of neuropeptides via MS imaging techniques. Methods for compiling a database and detecting peptides will be discussed as well as quantitation methods, such as label-free, isotopic, and isobaric labeling strategies. Methods for mapping the location of peptides within tissue will also be discussed. A generic workflow is shown in Fig. 2. Details for preparation and handling of neuropeptides from collection to instrumental analysis and data analysis methods are described.

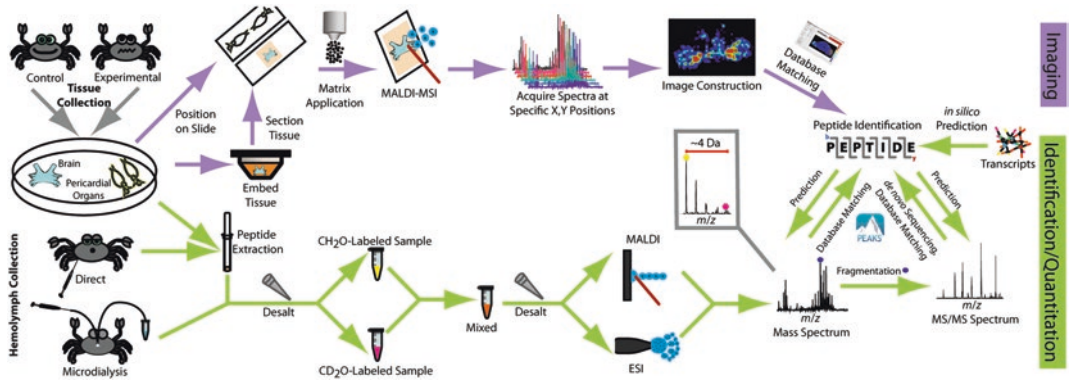



**Fig. 2** Workflow indicating the steps for sample collection, preparation, and data acquisition and analysis. The workflow contains steps for analysis of tissues, crude hemolymph, and microdialysis samples and describes the processes for both imaging and quantitation. For relative quantitation, duplex formaldehyde labeling is indicated, with $CH_2O$ indicating nonisotopic formaldehyde and $CD_2O$ indicating deuterated formaldehyde used

## 2 Materials

### 2.1 Chemicals and Equipment

1. Ultrapure water, used for all of the solutions listed below.
2. Methanol.
3. Acetonitrile.
4. Formic acid.
5. Acidified methanol (90:9:1 methanol:water:glacial acetic acid).
6. Isotopic formaldehyde solution (1% v/v).
7. Borane pyridine solution (30 mM).
8. Ammonium bicarbonate solution (100 mM).
9. Gelatin (100 mg/mL dissolved in water).
10. Matrix-assisted laser desorption/ionization (MALDI) matrix.
    (a) 2,5-dihydroxybenzoic acid (DHB) dissolved in 50:50 methanol:water with 0.1% formic acid (150 mg/mL for spots; 40 mg/mL for imaging).
    (b) α-cyano-4-hydroxycinnamic acid (CHCA) dissolved in 50:50 acetonitrile:water with 0.1% formic acid (10 mg/mL for spots; 5 mg/mL for imaging).
11. Glass manual homogenizer (e.g., Wheaten 1 mL Tissue Grinder, Tenbroeck).
12. Teflon pestle (e.g., Wheaten 1 mL Tissue Grinder, Tapered Pestle).
13. Bath sonicator.
14. Centrifuge capable of up to 16,000 × *g* and able to fit microfuge tubes and microfilters (e.g., Eppendorf 5424 R).
15. Vacuum centrifuge/concentrator, such as a SpeedVac.
16. 3 kDa or 10 kDa Molecular Weight Cutoff (MWCO) device (e.g., Amicon Ultra).
17. Reversed-phase C18-packed pipette tips (e.g., Omix 100 μL Tips or Millipore P10 ZipTips).
18. Sprayer for MALDI imaging (e.g., HTX Technologies).
19. Crab saline (440 mM NaCl, 11 mM KCl, 26 mM $MgCl_2$, 13 mM $CaCl_2$, 11 mM TRIS, 5 mM maleic acid, adjusted to pH 7.45 with HCl or NaOH. Stored at 4 °C).
20. Tweezers.
21. Plastic Cup (1 in. Height × 1 in. Diameter).
22. Razor blade.
23. Syringe Pump.
24. Syringe.
25. Automated sample collector with temperature control.

### 2.2 Instrumentation and Software

For quantitation and profiling, a Hybrid Quadrupole-Orbitrap Mass Spectrometer (we use the Thermo Q-Exactive) with a nano-electrospray ionization (ESI) source can be coupled to a UPLC system (we use the Waters nanoAQUITY) for separation of peptides with MS and MS/MS analysis for identification and quantitation. For imaging and complementary profiling, we use a hybrid MALDI-Ion Trap-Orbitrap Mass Spectrometer (Thermo MALDI-LTQ-Orbitrap XL). Other mass spectrometers with a MALDI source and an ESI source may also be used. The associated vendor software Xcalibur and ImageQuest can be used for data analysis and image processing. For MS/MS identification, de novo sequencing software can be used, such as PEAKS [12], MaxQuant [13], or PepNovo [14]. Additional image processing can be performed using freely available software, such as MSiReader [15] and Image J.

## 3 Methods

### 3.1 Identification and Quantitation

1. Sample Collection and Extraction.
   (a) Tissues.
   - Collection.
     - Collect the tissue of interest through dissection [16, 17].
     - Place tissues in a 0.6 mL microfuge tube with 10–20 μL of acidified methanol.
     - Store at −80 °C until ready to use (*see* **Note 1**).
   - Extraction.
     - Add 100 μL of acidified methanol per tissue to a glass manual homogenizer (*see* **Notes 2–5**).
     - Move the tissues to the homogenizer using tweezers. Transfer the storage solution as well.
     - Homogenize the tissue until no large particles are visible. Transfer the liquid to a clean 1.5 mL microfuge tube.
     - Add 100 μL of acidified methanol to the homogenizer, breaking up any residual particles. Transfer to same microfuge tube. Repeat once more.
     - Sonicate mixture for 10 min.
     - Centrifuge mixture for 20 min at >16,000 × *g*.
     - Transfer the supernatant to a new 1.5 mL tube.
     - Rinse the pellet with 100 μL of acidified methanol and break it up with a Teflon pestle.
     - Repeat steps 5–7. Discard the pellet.

- Dry down supernatant in a vacuum centrifuge on medium heat (*see* **Note 6**).

(b) Hemolymph.

- Collection.
  - Place the crab on ice for 5–10 min. Afterwards, place it in a metal dishpan on its back with its tail pointing toward you.
  - Prepare a 25 *g* needle by connecting it to a 1 mL plastic syringe.
  - Place the needle into an exposed leg joint of the crab.
  - Pull up on the syringe to create a vacuum. Hemolymph should come out immediately. If not, wiggle the needle at different angles until the liquid is withdrawn.
  - Add an equal amount of acidified methanol to the collected hemolymph in a 1.5 or 2 mL microfuge tube (*see* **Note 7**). A protein precipitate will be produced.
  - Store at −80 °C till ready to use (*see* **Note 1**).
- Extraction.
  - Using the Teflon pestle, break up the precipitate in the microfuge tube to produce a homogenous solution.
  - Sonicate tube for 10 min.
  - Centrifuge mixture for 20 min at >16,000 × *g*.
  - Transfer the supernatant to a new 2 mL microfuge tube.
  - Rinse the pellet with 500 μL of acidified methanol and break it up with a Teflon pestle.
  - Repeat steps 2–5. Discard the pellet.
  - Dry down supernatant in vacuum centrifuge on medium heat (*see* **Notes 6** and **8)**.
- Microfiltration.
  - Add 200 μL of 0.1 M NaOH solution to the 3 kDa or 10 kDa MWCO device (*see* **Note 9**). Centrifuge at >14,000 × *g* for 5 min.
  - Rinse MWCO device with 500 μL of 50:50 water:methanol. Centrifuge for 10 min at >14,000 × *g*.

- Dissolve sample in 500 μL of 30:70 water:methanol, vortex, and sonicate for 10 min.
- Add the sample to the MWCO device and run it through the membrane by centrifuging for 30 min at >14,000 × *g*. Save the flow-through from this step by placing it in a 1.5 mL microfuge tube (*see* **Note 10**).
- Rinse the membrane with 100 μL of 30:70 water:methanol. Collect the flow-through and add it to the tube from the previous step.
- Dry down the flow-through in vacuum centrifuge on medium heat.

(c) Microdialysis (*see* **Notes 11** and **12**).

- Collection.
  - Rinse probe with water and then crab saline using a 3 mL plastic syringe and a syringe pump set to 0.5 μL/min.
  - Surgically implant the probe into the animal directly above its heart [18]. Figure 1 indicates the location of the heart in *Cancer borealis*.
  - Allow the animal to recover for 24–48 h (*see* **Note 13**).
  - Collect samples at desired time points. Collection can be done manually with a 0.6 mL centrifuge tube on ice or using an automated sample collector set at 4 °C (*see* **Notes 14–16**).
  - Add enough formic acid to make the total concentration 0.1% (v/v).
  - Store samples at −80 °C if they are not being analyzed immediately. However, samples should be used as soon as possible to avoid degradation (*see* **Note 17**).

2. Desalting.

(a) Dissolve each sample in 0.1% formic acid (10 μL for tissues, 200 μL for hemolymph) by vortexing and sonicating for 10 min. (Microdialysis samples do not need to be redissolved.) If the pH is greater than 3, add small amounts of diluted formic acid (e.g., 10%) until the pH is below 3.

(b) Using the appropriately sized reversed-phase C18 pipette tip and respective volume (*see* **Note 18**) (10 μL for tissues and microdialysate, 100 μL for hemolymph), flush the packing material with pure acetonitrile at least 3 times.

(c) Flush the packing material with 0.1% formic acid 3 times.

(d) Bind the sample of interest to the packing material by flushing it with dissolved sample at least 15 times.

(e) Wash the packing material with 0.1% formic acid 3 times to wash away salts and other unbound contaminates. Add the first wash to its own tube or to the original sample vial as a precaution in the event that nothing binds to the packing material.

(f) Elute the sample into a 0.6 mL tube by flushing 10 μL of 50:50 acetonitrile:water up and down through the packing material 10 times.

(g) Dry down the eluate in a vacuum centrifuge on medium heat.

3. Quantitative Labeling.

(a) Dissolve each of the samples (e.g., of stressed and control) in 10 μL of water by vortexing and then sonicating for 10 min.

(b) Pipette 10 uL of the appropriate isotopic version of formaldehyde into its designated channel (*see* **Notes 19** and **20**).

(c) Pipette 10 uL of borane pyridine into each microfuge tube.

(d) Place the tubes in a 37 °C water bath for 15 min to allow complete labeling.

(e) Quench the reaction with 10 μL of ammonium bicarbonate solution.

(f) Mix equal amounts of each channel into a 0.6 mL tube. Do this for each set of channels.

(g) Dry down mixture in a vacuum centrifuge on medium heat.

4. Tissue Analysis and Data Acquisition.

(a) MALDI-MS.

- Matrix Application
  - Dissolve samples in 5 μL of 0.1% formic acid for each tissue used by vortexing and sonicating for 10 min (*see* **Note 21**).
  - In a separate 0.6 mL tube, mix equal amounts of matrix (*see* **Note 22**) and sample (*see* **Note 23**) solutions, vortexing them to ensure complete mixing.
  - Take 1 μL out of the tube using a pipette and spot it on the MALDI stainless steel plate. Rub the tip of the pipette along the edges of the spot circle to make sure the whole circle is filled.
  - Allow the matrix to crystalize fully before placing in the instrument.

- Analysis on hybrid MALDI-Ion Trap-Orbitrap-MS.
  - If necessary, attach a backing plate to the MALDI stainless steel spot plate (MALDI-LTQ-Orbitrap XL requires this), and insert the plate set into the instrument by placing the plate in the correct position and pressing the appropriate button on the vendor's tune page (e.g., the "Insert MALDI sample plate" button on the "MALDI source" window of the instrumental LTQ tune page).
  - Once the plate has been inserted, confirm the spot in which you placed your sample and shoot the spot with an appropriate laser energy to obtain a signal intensity (approximately 1E7 for MALDI-LTQ-Orbitrap XL, *see* **Note 24**).
  - Open Thermo Xcalibur Sequence Setup, and create a new sequence (or use alternative vendor's software with regard to its specifications).
  - Fill the sequence, including the title, path (for saving the data), instrumental method (*see* **Note 25**), and spot position (*see* **Note 26**).
  - Start automated or manual acquisition of every spot of the sequence (in Xcalibur: Select each line to be run and press the "Run Sequence" button).

(b) ESI-MS (*see* **Notes 27–29**).

- Sample Prep.
  - Dissolve samples (approximately 5 μg on average) in 5 μL of 0.1% formic acid per tissue used or per 0.25 mL hemolymph by vortexing and then sonicating for 10 min.
  - Transfer the volume into a LC vial with a screw top septum (we use Waters vials), making sure to get all the volume to the bottom with no bubbles.
- MS analysis (we use Q-Exactive, Thermo)
  - Place the LC vial into the chilled sample compartment of the LC system (we use the Waters NanoAquity). Remember the tray number and tray position in which the samples were placed.
  - Make sure the column system has been equilibrated to the starting conditions of the gradient you will be using.
  - On the instrument profile, set up the sequence as described above (*see* **Note 25**).
  - Select each sample in the sequence (*see* **Note 30**) to be run (in Xcalibur: press the "Run Sequence" button).

5. Data Analysis.
   (a) MALDI-MS.
      - Average all the MS spectra collected by left-clicking and dragging across the chromatogram.
      - Copy the raw data into your vendor's software using the exact masses (Xcalibur: by right-clicking on the spectrum, selecting "Export" and "Clipboard (exact mass))."
      - Paste data in blank spreadsheet file (we use Excel, Microsoft).
      - Compare the masses found to the in-house database (unlabeled or labeled) (*see* **Notes 31** and **32**). *See* Fig. 3 for an example spectrum.

   (b) ESI-MS
      - Load the data into a de novo sequencing software (we use PEAKS), creating a new project, and loading the raw data into the program (*see* **Note 33**).
      - Process the samples (using the "PEAKS search" button), indicating any enzyme used and any PTMs expected (*see* **Note 34**), and choosing the in-house database. De novo sequencing will be done with the same parameters (*see* **Notes 35**).
      - Once the search is complete, quantitative information can be mined (*see* **Note 36**), in PEAKS by using the "Quantification" icon.

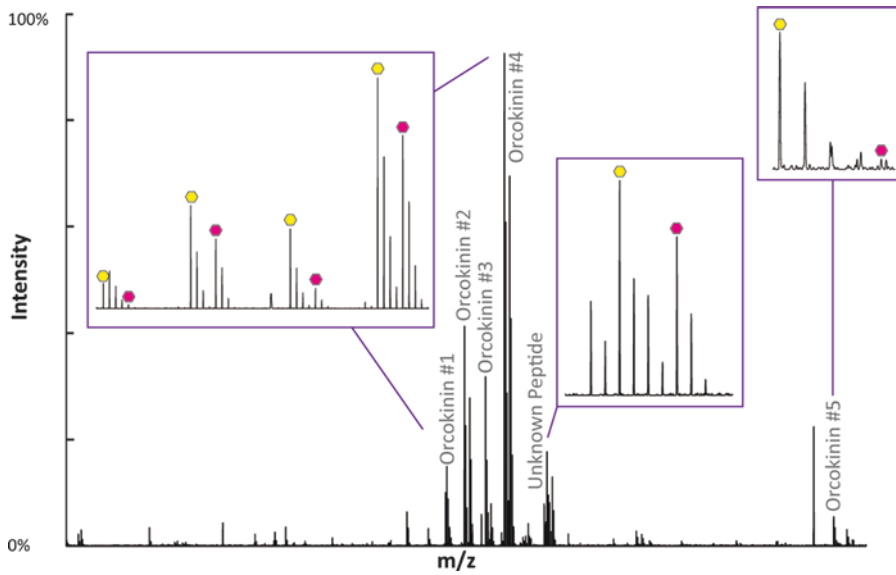



**Fig. 3** Example spectrum of results obtained from tissue analysis of blue crab, *Callinectes sapidus*, brain using a MALDI-LTQ-Orbitrap XL. Neuropeptides from the Orcokinin family are indicated, as well as an unknown peptide. Enlarged images show the mass shift between light and heavy-labeled peptides, which enables relative quantitation

### 3.2 Localization by Mass Spectrometric Imaging

1. Sample Preparation.
   (a) Small or thin tissue (e.g., PO, STNS, STG as in Fig. 1).
   - Collect the tissue of interest by means of dissection [17, 19], being mindful of tissue orientation (*see* **Note 37**).
   - Hold onto tissue with tweezers and briefly submerge it in a microfuge tube of water to desalt (*see* **Note 38**).
   - Place the clean tissue directly on a glass slide. Use a marker to label the orientation of the tissue (i.e., which area is closest to the head, tail, etc.). Stretch the tissue out on the slide to ensure that it is lying flat and easily visible.
   - Tissue should be analyzed immediately (*see* **Note 39**).

   (b) Large tissue (e.g., brain, TG).
   - Prepare gelatin solution, vortex, and place it in a 37 °C water bath until fully dissolved. Keep gelatin warm while not in use.
   - Pour enough gelatin into a plastic cup to cover the bottom of the cup. Allow the gelatin to solidify at room temperature.
   - Collect the tissue and desalt as described above for small tissue.
   - Place the tissue on top of the gelatin layer, making note of the tissue orientation. Fill the cup with warm gelatin. The tissue will float toward the top. Reorient the tissue if necessary to make sure it is centered in the cup.
   - Immediately place the cup with tissue in dry ice to flash freeze it (*see* **Note 39**).
   - Store tissue at −80 °C until use.
   - Cross-section tissue using a cryostat.
     - Remove tissue-embedded gelatin block from plastic cup and trim excess gelatin with a clean razor blade.
     - Place small droplet of water on cryostat chuck and place tissue block on top. Surround the rest of the block with water, making sure to not get any below the chuck.
     - Keep the block/chuck in the cryostat to allow the water to freeze, attaching the tissue block on the chuck (approximately 15 min).
     - Attach the chuck to cryostat and align tissue block so that even slices can be made through tissue.

- Obtain several 12–16 μm sections from throughout the tissue. Thaw-mount each section to a glass slide by warming the glass slide slightly and placing it directly above section. Section will adhere to slide (*see* **Note 40**).
- Store slides at −80 °C until ready for analysis (*see* **Note 41**).

2. Image Acquisition (MALDI-MS).

(a) Matrix Application via Sprayer (*see* **Note 42**).

- Turn on matrix sprayer (we use the TM-Sprayer, all subsequent experimental details refer to that) and solvent syringe pump, setting desired solvent flow rate.
- Set pressurized air to 10 psi.
- Open program and set temperature to 80 °C for DHB (*see* **Notes 43** and **22**).
- Set up the desired method (*see* **Note 44**).
- Switch the injection loop to "Load" and load matrix using a syringe. Switch the injection loop to "Spray" and ensure that matrix is being sprayed. It may take several minutes for matrix to reach the spray nozzle.
- Click "Start" on the Sprayer software (*see* **Notes 45** and **46**).
- When method has finished, flush the matrix loop with solvent three times and set the temperature of the sprayer to 30 °C. When the sprayer temperature has lowered to this temperature, turn off sprayer, syringe pump, pressurized air, and software.

(b) Analysis on MALDI-LTQ-Orbitrap

- Insert one or two slides into MALDI imaging plate.
- Place plate face-down in scanner, and scan image of entire plate. Ensure that the white crosses in the corners of the plate are clearly visible in the scanned image; otherwise, the image cannot be aligned to the plate in the instrument.
- Attach the backing plate to the MALDI imaging plate and insert the plate set into the instrument.
- Once the plate has been inserted, click on the "Tissue Imaging" tab of the MALDI Source page and check the box titled "Use Tissue Imaging Feature."
- Under "Position File," check "Import Image" and upload the scanned image of the MALDI plate (*see* **Note 47**).

- Enter the appropriate raster size (*see* **Note 48**) and ensure that the size "rectangular" is selected. (Rectangles are the preferred shape for methods.)
- Select "View Plate," and click on the square selection tool. Select the area of the tissue. Close the "View Plate" window and save the position file (*see* **Notes 49–51**).
- Click on the "Control" tab and shoot the laser at an area of matrix. Adjust the laser energy as needed in order to obtain sufficient signal (with our instrument, on the order of 1E7).
- Open the software setup (Thermo Xcalibur Sequence Setup), and create a new sequence as previously described. For the position, copy and paste the directory of the position file for each tissue into its respective row.
- Select each line to be run and press the "Run Sequence" button.

(c) Data Analysis.

- ImageQuest
  - Open raw data in ImageQuest.
  - Click the "New data set" icon and type in the mass of interest and the tolerance window with the "Base Peak" plot type. Under plot type, select the desired normalization, if any (*see* **Note 52**). Click "OK."
  - To save or export individual images, select the image and click "Copy" in the Edit tab. Paste the image in the desired location (*see* **Note 53**). *See* Fig. 4 for example MS images.
- MSiReader (*see* **Note 54**) [12].
  - Open the raw data in MSiReader using the appropriate open-access format (*see* **Note 55**).
  - Enter an *m/z* value from ImageQuest that displays a clear distribution in the tissue.
  - Enter the appropriate parameters for the *m/z* window, normalization (*see* **Note 52**), and color map.
  - Use the image overlay feature to upload a scanned image of the tissue and align the image to the MS image of the tissue. Set the transparency to around 50% to ensure both optical and MS images are clearly visible.
  - Generate an image of each neuropeptide signal of interest from a database using "generate an image for each peak in a list" button.

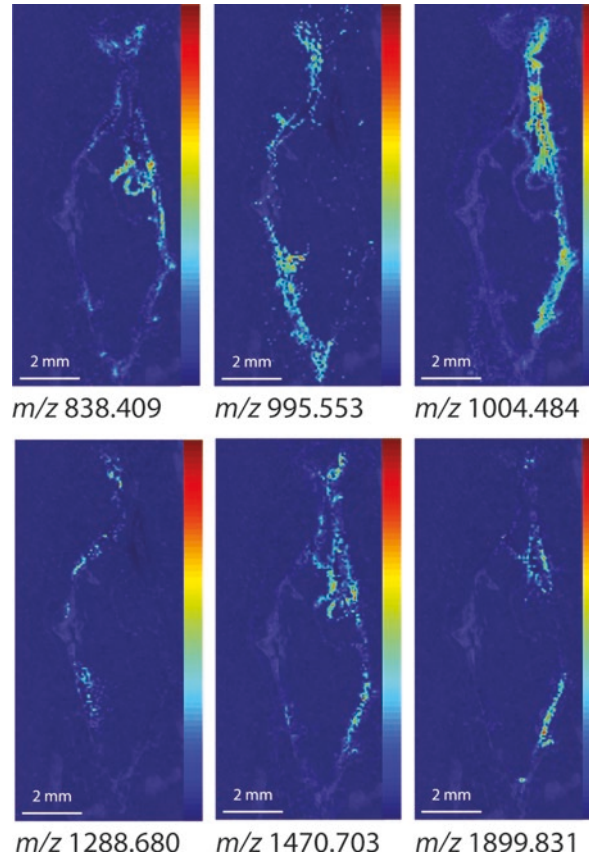



**Fig. 4** Example MS images of a blue crab, *Callinectes sapidus*, PO. Images were obtained on a MALDI-LTQ-Orbitrap XL and show the spatial distribution of six selected *m/z* values. All images are normalized to the TIC and overlaid on the optical image of the tissue

  - Manually examine images for detected neuropeptides and distinct spatial distributions.
  - Search for novel neuropeptides and *m/z* values outside of the database by using the peak finding tool.
- 3D image generation (Image J).
  - Open grayscale images of consecutive tissue slices with the same *m/z* value (saved from ImageQuest).
  - Align the images using the functions under "image."
  - Combine the images into a three-dimensional stack using the "images to stack" button.
  - View the three-dimensional image by clicking "image," "stacks," and then "3D project." Viewing parameters can be adjusted in the 3D projection window.

## 4 Notes

1. While storage at −80 °C minimizes postmortem degradation, tissues should only be stored for up to 6 months before use. There are several ways to decrease postmortem degradation for both tissues and hemolymph, which would extend their storage life. For tissues, the use of a Denator heat stabilizer system or boiling of the tissues for heat stabilization has been shown to be effective [20]. For hemolymph, the addition of protease inhibitor cocktail or EDTA can reduce protease activity.
2. Several different extraction solvents have been tested for their ability to extract neuropeptides from tissues, including acidified organic solvents [21]. A method termed "mixing on column" (MOC), where four different extraction solvents are incorporated sequentially, has been shown to be extremely effective for mammalian neuropeptides [22], but this technique's usefulness is still being tested for crustacean peptidomics.
3. While a manual homogenizer has been shown to be effective for the small crustacean tissues, an electronic sonicator may be needed to fully break down the tissue walls of large organs. These should be used with the tissues on ice to help with heat dissipation.
4. When deciding how many tissues to pool, it is important to balance biological variance with the amount of tissues required for the study. In most stress-related studies, at least three animals' tissues are pooled for each biological replicate.
5. Direct profiling is an alternative strategy to extraction. The tissue of interest can be placed on the stainless steel spotting plate, with matrix being spotted directly onto the tissue area of interest. This method has become phased out with the development of new imaging techniques (*see* "Localization Using Imaging" section in this book chapter).
6. Many crustacean peptides vary widely in their molecular masses. For example, crustacean neuropeptides can range from 0.5 to 9 kDa. With many Orbitrap-based instruments, it is difficult to analyze the larger peptides without taking a bottom-up approach. This means some sort of digestion prior to analysis may be required to get a full picture of a crustacean peptidome. For crustacean peptides, the digestion is usually done on the initial extract. After reducing all disulfide bonds with dithiothreitol and alkylating with iodoacetamide, an enzyme (e.g., trypsin) is added at a 25–50:1 peptide:trypsin ratio. Peptide content can be determined using the bicinchoninic acid assay or a similar assay.
7. Each crab has a limited amount of hemolymph stored in its body. For example, *Callinectes sapidus* has approximately 5 mL. After removal, hemolymph is restored over a period of

time. Thus, the amount of hemolymph removed depends on the species of crab and whether it is to survive afterwards. Replacing the volume of removed hemolymph with an equal volume of crab saline can increase the crab's chance at survival.

8. After drying down any hemolymph samples (initially to fully processed), they should be resuspended prior to storage in the −80 °C freezer to ensure full dissolution.
9. Depending on the size of the peptides of interest, two versions of MWCO filters are available: 3 K or 10 K Daltons. 10 K is normally used to purify a wide mass range of crustacean peptides (*see* **Note 5**). After digestion, 3 K MWCO filters can be used to simplify the sample and allow for a targeted analysis.
10. Many times, the MWCO filters become clogged with proteins that were not dissolved in the methanol:water mixture that was added to the extract, and which were not pelleted in the extraction step. In order to lessen the chance of this happening, (a) centrifuge the extract at >16,000 × *g* for 15–20 min before putting the supernatant through the filter, (b) separate the extract into two fractions with two MWCO filters, or (c) centrifuge the extract through the filter for a longer period of time (45+min).
11. There are several types of commercial probes that can be used for microdialysis. Alternatively, homemade probes can be implemented at a much lower cost. Whether purchasing or making probes, there are several factors worth considering, including membrane material, the molecular weight cutoff (the expected MWCO is approximately a third of the reported MWCO), area of membrane, and use of polymer coatings [23].
12. The microdialysis technique tends to have low recovery in vivo. The recovery can be enhanced through the use of affinity agents [24].
13. It is important to closely monitor the crab while it recovers from surgery, as the majority of post-surgery complications (i.e., crab dying, pulling out probe, clogging probe) typically occur within the first 24 h.
14. Collection windows are typically within the range of 30 min to 2 h. This is dependent on the necessary temporal resolution and recovery of sample. Shorter collection durations provide higher resolutions, but yield lower sample volumes. Collection times of several minutes have been shown feasible, often using online microdialysis collection coupled to ESI. However, this method also comes with drawbacks [25].
15. The optimum infusion rate for microdialysis is a trade-off between relative recovery and absolute recovery. Lower flow rates allow for more diffusion of sample through the membrane, but result in less sample volume, while higher flow rates enable the collec-

tion of more sample overall but run the risk of disturbing the crab. We have found the ideal flow rate to be 0.5 μL/min.

16. When collecting samples, it is important to account for the dead volume between the tip of the probe and the end of the tubing (and collection needle, if one is being used). The delay from sample diffusing into the probe to reaching the end of the tubing could be an hour or greater, depending on the length of tubing and infusion rate.
17. Microdialysis samples have been shown to degrade rapidly, even within several hours. Therefore, samples should be analyzed as soon as possible after collection. If longer storage time is necessary, there are several methods for improving the lifetime of the samples [25, 26].
18. Many different varieties of separations can be done to increase peptide purity, including strong cation exchange (SCX) and C18 pipette tips (e.g., ZipTips). Success has also been found in increasing peptide coverage in fractionation using C18 pipette tips or an a high-performance liquid chromatography (HPLC) system [27].
19. Formaldehyde labeling is commonly used to produce 2 or 3 different isotopic forms that differ in mass [28–30], although up to 5 different isotopic forms can be generated with different combinations of heavy and light formaldehyde and reducing agent (*see* Chapter 10). Other labeling methods exist, such as isotopic N,N-dimethyl leucine (iDiLeu), which boasts 5-plex labeling capabilities, which allows for relative or absolute quantitation by an in-solution calibration curve [31]. Absolute quantitation of a single peptide can also be done by adding a deuterated version of the peptide to the original extract of the sample.
20. For crustacean tissues, formaldehyde labeling has been shown to be extremely effective due to its quick and complete labeling of peptides. While MS-based quantitation strategies are simple, they have limitations, especially when analyzing more than five samples or samples with high spectral complexity. On the other hand, MS/MS-based quantitation strategies, such as iTRAQ, TMT, or N,N-dimethyl leucine (DiLeu), allow for higher multiplexing (up to 12 sample comparisons at once have been demonstrated, but up to 18 may be possible) with lower MS spectral complexity [32, 33]. Unfortunately, quantitative depth may suffer because, in order to become quantified, the peptide needs to be selected for MS/MS.
21. Prior to spotting with matrix, an offline separation can be done, for example with capillary electrophoresis or a HPLC. Success has been seen for separating tryptic peptides with both high-pH and strong-cation-exchange chromatography [34, 35].

22. Several matrices work well for crustacean peptide analysis, including DHB and CHCA. While DHB extracts peptides well, CHCA, known for being a "universal matrix," tends to provide a more homogenous layer.
23. Several spotting techniques exist, including the suggested premixing; alternatives are sandwiching, or individual mixing on plate prior to recrystallization.
24. Depending on the instrument, lifetime of the laser, instrumental setup, and the matrix of choice, the laser energy used will need to be optimized for each sample.
25. In order to acquire more confident identifications or perform de novo sequencing, tandem MS is necessary. For MALDI instruments, collision-induced dissociation and high-energy collisional dissociation are the main commercially available options. Unfortunately, MALDI ionization mainly produces singly charged ions, which lead to poor fragmentation. This means that most identifications by MALDI are performed through accurate mass matching. ESI provides much higher-quality fragmentation spectra, meaning that tandem MS can be used for high-confidence identification and discovery of novel peptides. As another fragmentation option, electron transfer dissociation, which is better for post-translational modification analysis, is now being more readily available. Even with high quality tandem MS spectra, de novo sequencing of peptides can be challenging. The use of chemical derivatization with a nonisotopic version of formaldehyde (*see* **Note 19**) has allowed for more complete coverage when de novo sequencing putative peptides [30]. Other methods, such as thiol reduction and alkylation, have also been used to achieve high-resolution sequencing of larger, disulfide-bonded crustacean peptides [36].
26. Xcalibur software only recognizes plates as having 96 wells, so it will only allow you pick spots that are within that plate size, even if you are using a 384-well plate.
27. Besides classic LC separation, several other complementary options exist. Capillary electrophoresis is compatible offline with both MALDI and ESI analysis for enhanced separation of crustacean neuropeptides [37, 38]. Ion mobility has also been shown to be effective at separating D/L-epimeric crustacean peptides [39].
28. The use of either a commercial or homemade column should be considered. Homemade packing lowers the cost of each column and allows for customization, which may improve peptide separation and thus coverage. However, low uniformity can cause inconsistency between columns in comparison to commercially available columns, although these columns can be much more expensive in comparison.

29. For peptides, the reverse-phase C18 column setup uses a gradient of water with 0.1% FA and acetonitrile with 0.1% FA to elute the sample. Depending on the complexity of the sample, the time of gradient, highest percent acetonitrile added, and instrumental parameters (e.g., dynamic exclusion) will need to be adjusted.
30. When beginning any LC-MS experiment, a quality control sample should be run to determine if the instrument is working optimally. Also, blanks should be run between each new sample type to decrease sample overlap.
31. For Orbitrap instruments, the standard mass tolerance is ±5 ppm.
32. This can be done manually or with a simple peak-picking program.
33. Several software packages are capable of performing de novo sequencing, database searching, and peak peaking for quantitation besides PEAKS (e.g., Proteome Discoverer).
34. Common post-translational modifications (PTMs) for crustacean neuropeptides include amidation and dehydration. If any labeling is done, it is important to include the tag or modification during the search (e.g., dimethylation).
35. PEAKS software package works by first de novo sequencing all of the raw data, which will then be matched to the provided database of precursor masses and proposed sequences. It is important to note that, while other databases come from sequenced genomes, the crustacean neuropeptide database has been developed in-house, as there is no genomic database for crustaceans. When looking at peptides with no digestion, PEAKS shows the peptide searched as a "protein" in the program. The "peptides" section is the individual peptides that make up all the crustacean peptides, which can include cleaved or degraded derivatives. Anything that doesn't match is placed in a "de novo only" tab, which can provide you with possible novel peptide groups. For neuropeptides, these "de novo only" peptides are compared to current family sequences, from which certain sequence themes can be found to identify novel neuropeptides. Confident identification can then be done by synthesizing standards to confirm MS/MS and LC retention patterns.
36. Both MS and MS/MS-based quantitation can be done, depending on the labeling used in the quantitation step. Label-free quantitation is also possible.
37. It may be helpful to label dissection dishes with the orientation of the tissues to avoid uncertainties after the tissue is dissected.

38. Delicate tissue such as the PO can be transferred to the water by either carefully folding the tissue into quarters or by holding the tissue at either end with separate sets of tweezers.
39. To prevent degradation, a Denator heat stabilizer can be implemented immediately after desalting. However, caution must be taken to avoid melting delicate tissue.
40. It is common for slices to fold, tear, or become distorted during the process of thaw mounting. Once the slice is on the slide, thoroughly check it to ensure the integrity of the tissue.
41. For best results, slides should be analyzed as soon as possible. During storage, slides can be wrapped in tin foil to protect them from damage. When removing the slides from the freezer, water will condense on the slides, which can cause diffusion, due to the major temperature change. Place them in a desiccator to minimize this effect.
42. There are several methods for applying matrix to a sample, including airbrush, automated sprayer, inkjet printer, and sublimation. The method described here utilizes an automated sprayer (e.g., TM-Sprayer) which has been found to be the most reproducible [40].
43. The sprayer temperature should be high enough to evaporate the matrix/solvent mixture. For the TM-Sprayer, turn up the temperature slowly until you hear a "puffing" noise. Reduce the temperature by 5 °C for the final method.
44. Parameters to consider when choosing a method include the number of times the sprayer passes over and coats the slide, the drying time between coats, the direction the sprayer moves in to coat the slide (e.g., horizontal or vertical), and the system flow rate. Our preferred method includes 12 passes with a 30-s dry time between passes, and the orientation alternating between horizontal and vertical with each pass to provide an even coating at a flow rate of 0.1 mL/min. Methods should be optimized to the peptide group of interest and matrix to minimize diffusion while increasing peptide extraction.
45. The matrix loop only has a finite volume it can hold, and so it may be necessary to reload it with matrix during the course of the method. If this is necessary, wait until the sprayer is done with its current pass, switch the injection loop to "Load," load matrix, and switch the injection loop back to "Spray" before it starts the next pass.
46. After clicking "Start," the system may say the temperature is unstable due to the sensitivity of the system. Since the temperature will never fully stabilize, click "Start Now" to start the method.
47. The MALDI-Orbitrap is also capable of scanning in the whole plate or each individual slide at varying image qualities.

Scanning the plate with an external system only takes a few minutes, while scanning one slide at "normal" resolution takes ~25 min. For other instruments, images need to be scanned by an external scanner.

48. Raster size determines the spatial resolution of the tissue being imaged. While small raster sizes provide better image resolution, they can drastically increase the analysis time of the instrument. The MALDI-Orbitrap allows for a raster size of 75 μm without oversampling, but other MALDI instruments boast spatial resolution down to a few microns.
49. After selecting the final area of interest, saving this image in the "View Plate" window allows for easier image overlay during data processing.
50. The software tends to shift the selected area unpredictably after closing the window. Therefore, it is recommended to close and reopen the "View Plate" window after a selection has been made to ensure that the area of interest is still within the selection box.
51. It is good practice to include a small amount of area outside of the tissue when selecting the area to be analyzed in order for it to be used as a means to distinguish signal from random noise.
52. Typically, normalization is performed in reference to the total ion current (TIC), in which each mass spectrum is divided by its TIC. This ensures that all spectra have the same integrated area under their curves. Other normalization strategies are emerging [41], but TIC continues to be the most widely used.
53. When comparing multiple images, it is often helpful to set them to the same intensity scale. This can be done by entering the Min and Max Plot Values in the "Scale" tab.
54. MSiReader is a freely available, open-access software that can be downloaded from http://www.msireader.com/ [15].
55. To convert to imzML in ImageQuest, click File, Export, imzML. Either all peaks or centroids only can be exported.

## Acknowledgments

This work was supported in part by a National Science Foundation grant (CHE-1413596, CHE-1710140) and the National Institutes of Health through grant 1R01DK071801. ARB acknowledges the National Institute of General Medical Sciences F31 Fellowship (1F31GM119365) for funding. KD acknowledges a predoctoral fellowship supported by the NIH, under Ruth L. Kirschstein National Research Service Award T32 HL 007936 from the National Heart Lung and Blood Institute to the University of Wisconsin-Madison Cardiovascular Research Center.

## References


1. OuYang CZ, Liang ZD, Li LJ (2015) Mass spectrometric analysis of spatio-temporal dynamics of crustacean neuropeptides. Biochimica Et Biophysica Acta-Proteins and Proteomics 1854(7):798–811
2. Christie AE, Stemmler EA, Dickinson PS (2010) Crustacean neuropeptides. Cell Mol Life Sci 67(24):4135–4169
3. Fricker LD, Lim J, Pan H, Che F-Y (2006) Peptidomics: identification and quantification of endogenous peptides in neuroendocrine tissues. Mass Spectrom Rev 25(2):327–344
4. Ye H et al (2015) Defining the Neuropeptidome of the spiny lobster Panulirus interruptus brain using a multidimensional mass spectrometry-based platform. J Proteome Res 14(11):4776–4791
5. Jia CX et al (2012) High-definition de novo sequencing of crustacean hyperglycemic hormone (CHH)-family neuropeptides. Mol Cell Proteomics 11(12):1951–1964
6. Taylor SW, Nikoulina SE, Andon NL, Lowe C (2013) Peptidomic profiling of secreted products from pancreatic islet culture results in a higher yield of full-length peptide hormones than found using cell lysis procedures. J Proteome Res 12(8):3610–3619
7. Bassett L et al (2014) Telemetry video-electroencephalography (EEG) in rats, dogs and non-human primates: methods in follow-up safety pharmacology seizure liability assessments. J Pharmacol Toxicol Methods 70(3):230–240
8. Zhang Y, Buchberger A, Muthuvel G, Li L (2015) Expression and distribution of neuropeptides in the nervous system of the crab Carcinus maenas and their roles in environmental stress. Proteomics 15(23–24):3969–3979
9. Casadonte R, Caprioli RM (2011) Proteomic analysis of formalin-fixed paraffin-embedded tissue by MALDI imaging mass spectrometry. Nat Protoc 6(11):1695–1709
10. Carlo D, Edd J, Irimia D, Tompkins R, Toner M (2008) Equilibrium separation and filtration of particles using differential inertial focusing. Anal Chem 80:2204–2211
11. Frechette J, Drazer G (2009) Directional locking and deterministic separation in periodic arrays. J Fluid Mech 627:379–401
12. Ma B et al (2003) PEAKS: powerful software for peptide de novo sequencing by tandem mass spectrometry. Rapid Commun Mass Spectrom 17(20):2337–2342
13. Cox J, Mann M (2008) MaxQuant enables high peptide identification rates, individualized p.P.B.-range mass accuracies and proteome-wide protein quantification. Nat Biotechnol 26(12):1367–1372
14. Frank A, Pevzner P (2005) PepNovo: de novo peptide sequencing via probabilistic network modeling. Anal Chem 77(4):964–973
15. Bell GW, Eggleston DB, Noga EJ (2010) Molecular keys unlock the mysteries of variable survival responses of blue crabs to hypoxia. Oecologia 163(1):57–68
16. Bergman N, Shevchenko D, Bergquist J (2014) Approaches for the analysis of low molecular weight compounds with laser desorption/ionization techniques and mass spectrometry. Anal Bioanal Chem 406(1):49–61
17. Talbi R, Klosen P, Laran-Chich MP, El Ouezzani S, Simonneaux V (2016) Coordinated seasonal regulation of metabolic and reproductive hypothalamic peptides in the desert jerboa. J Comp Neurol 524(18):3717–3728
18. Thorpy MJ (2015) Update on therapy for narcolepsy. Curr Treat Options Neurol 17(5):347
19. Taverna D, Norris JL, Caprioli RM (2015) *Cancer Borealis* Stomatogastric nervous system dissection histology-directed microwave assisted enzymatic protein digestion for MALDI MS analysis of mammalian tissue. Anal Chem 87(1):670–676
20. Sturm RM, Greer T, Woodards N, Gemperline E, Li LJ (2013) Mass spectrometric evaluation of Neuropeptidomic profiles upon heat stabilization treatment of neuroendocrine tissues in crustaceans. J Proteome Res 12(2):743–752
21. Wang GR, Yang F, Zhao W (2014) There can be turbulence in microfluidics at low Reynolds number. Lab Chip 14:1452–1458
22. Yin J, Mobarec JC, Kolb P, Rosenbaum DM (2014) Crystal structure of the human OX2 orexin receptor bound to the insomnia drug suvorexant. Nature 519:247–250
23. Thorpy MJ, Dauvilliers Y (2015) Clinical and practical considerations in the pharmacologic management of narcolepsy. Sleep Med 16(1):9–18
24. Lietz C, Yu Q, Li L (2014) Large-scale collision cross-section profiling on a travelling wave ion mobility mass spectrometer. J Am Soc Mass Spectrom 25(12):2009–2019
25. Tierney AJ, Kim T, Abrams R (2003) Dopamine in crayfish and other crustaceans: distribution in the central nervous system and physiological functions. Microsc Res Tech 60(3):325–335
26. Ajdari A, Bontoux N, Stone HA (2006) Hydrodynamic dispersion in shallow microchannels: the effect of cross-sectional shape. Anal Chem 78:387–392

27. Ma MM, Wang JH, Chen RB, Li LJ (2009) Expanding the crustacean Neuropeptidome using a multifaceted mass spectrometric approach. J Proteome Res 8(5):2426–2437
28. Boersema PJ, Aye TT, van Veen TAB, Heck AJR, Mohammed S (2008) Triplex protein quantification based on stable isotope labeling by peptide dimethylation applied to cell and tissue lysates. Proteomics 8(22):4624–4632
29. DeKeyser SS, Li LJ (2006) Matrix-assisted laser desorption/ionization Fourier transform mass spectrometry quantitation via in cell combination. Analyst 131(2):281–290
30. Bai L, Romanova EV, Sweedler JV (2011) Distinguishing endogenous D-amino acid-containing neuropeptides in individual neurons using tandem mass spectrometry. Anal Chem 83(7):2794–2800
31. Behrens HL, Li LJ (2010) Monitoring neuropeptides in vivo via microdialysis and mass spectrometry. In: Soloviev M (ed) Peptidomics: Methods and Protocols, Methods in molecular biology, vol Vol. 615. Humana Press Inc, Totowa, pp 57–73
32. Frost DC, Greer T, Li L (2015) High-resolution enabled 12-Plex DiLeu isobaric tags for quantitative proteomics. Anal Chem 87(3):1646–1654
33. McAlister GC et al (2012) Increasing the multiplexing capacity of TMT using reporter ion isotopologues with isobaric masses. Anal Chem 84(17):7469–7478
34. Ting L, Rad R, Gygi SP, Haas W (2011) MS3 eliminates ratio distortion in isobaric multiplexed quantitative proteomics. Nat Methods 8(11):937–940
35. Torfs P et al (2001) Pyrokinin neuropeptides in a crustacean—isolation and identification in the white shrimp Penaeus vannamei. Eur J Biochem 268(1):149–154
36. Christie AE (2014) Prediction of the peptidomes of Tigriopus californicus and Lepeophtheirus salmonis (Copepoda, Crustacea). Gen Comp Endocrinol 201:87–106
37. Hopkins PM (2012) The eyes have it: a brief history of crustacean neuroendocrinology. Gen Comp Endocrinol 175(3):357–366
38. Warkiani ME et al (2016) Ultra-fast, label-free isolation of circulating tumor cells from blood using spiral microfluidics. Nat Protoc 11:134–148
39. Zhang Z, Wang J, Hui L, Li L (2011) Membrane-assisted capillary isoelectric focusing coupling with matrix-assisted laser desorption/ionization-Fourier transform mass spectrometry for neuropeptide analysis. J Chromatogr 1218(31):5336–5343
40. Bergman P et al (2014) Narcolepsy patients have antibodies that stain distinct cell populations in rat brain and influence sleep patterns. Proc Natl Acad Sci U S A 111(35):E3735–E3744
41. Kallback P, Shariatgorji M, Nilsson A, Andren PE (2012) Novel mass spectrometry imaging software assisting labeled normalization and quantitation of drugs and neuropeptides directly in tissue sections. J Proteome 75(16):4941–4951

# Chapter 18

# Identification of Endogenous Neuropeptides in the Nematode *C. elegans* Using Mass Spectrometry

**Sven Van Bael, Samantha L. Edwards, Steven J. Husson, and Liesbet Temmerman**

## Abstract

The nematode *Caenorhabditis elegans* lends itself as an excellent model organism for peptidomics studies. Its ease of cultivation and quick generation time make it suitable for high-throughput studies. Adult hermaphrodites contain 959 somatic nuclei that are ordered in defined, differentiated tissues. The nervous system, with its 302 neurons, is probably the most known and studied endocrine tissue. Moreover, its neuropeptidergic signaling pathways display a large number of similarities with those observed in other metazoans. However, various other tissues have also been shown to express several neuropeptides. This includes the hypodermis, gonad, gut, and even muscle. Hence, whole mount peptidomics of *C. elegans* cultures provides an integral overview of peptidergic signaling between the different tissues of the entire organism. Here, we describe a peptidomics approach used for the identification of endogenous (neuro)peptides in *C. elegans*. Starting from a detailed peptide extraction procedure, we will outline the setup for an online liquid chromatography-mass spectrometry (LC-MS) analysis and describe subsequent data analysis approaches.

**Key words** *Caenorhabditis elegans*, Neuropeptide, Peptidomics, FMRFamide-like peptide, *flp*, Neuropeptide-like protein, *nlp*, Mass spectrometry, LC-MS

## 1 Introduction

### *1.1 Caenorhabditis elegans as a Model Organism*

*Caenorhabditis elegans* is a free-living nematode that can easily be found in anthropogenic habitats, such as compost heaps. It was first introduced into the lab in the 1960s and has since become a model organism of choice for thousands of researchers around the world. *C. elegans* can exist as self-fertilizing hermaphrodites, consisting of exactly 959 somatic nuclei that are ordered into fully differentiated tissues (Fig. 1), or males consisting of 1031 somatic nuclei. In favorable conditions, the abundance of males in the population is very low (±0.1%) but their numbers increase under conditions of stress. Also, male-mating will result in a 50% male progeny, which is useful when crossing hermaphrodites with males

Michael Schrader and Lloyd Fricker (eds.), *Peptidomics: Methods and Strategies*, Methods in Molecular Biology, vol. 1719, https://doi.org/10.1007/978-1-4939-7537-2_18, 

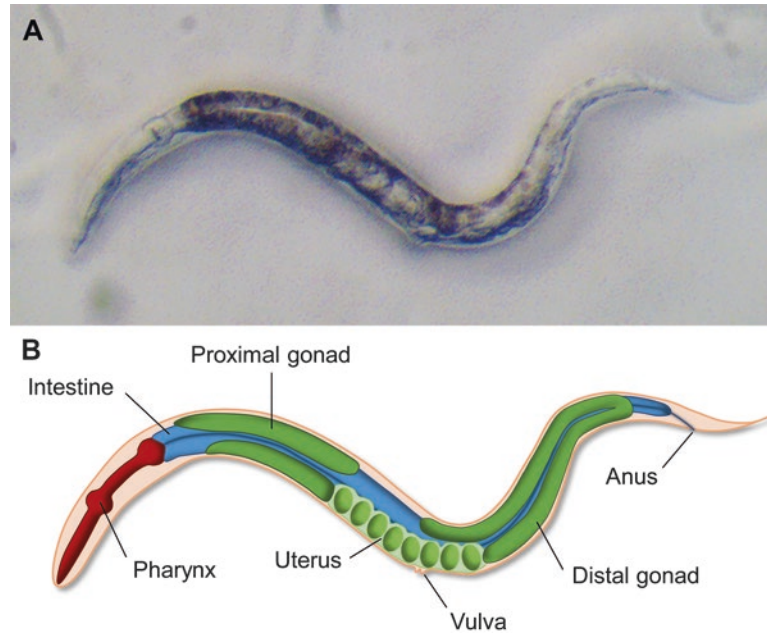



**Fig. 1** (**a**) Bright-field micrograph of an adult *C. elegans* hermaphrodite. (**b**) Schematic drawing of the most prominent anatomical structures

for genetic studies. In addition, *C. elegans* has a relatively short life cycle. It takes about 3 days to complete the cycle from egg to egg-laying adult, during which the animal goes through four larval stages (L1–L4). Unfavorable conditions during development can lead to resilient, thinner "dauer" larvae. These have a relatively impermeable cuticle, are nonfeeding, and can survive for months, in contrast to the average lifespan of around 2–3 weeks under standard conditions. For a more complete introduction on *C. elegans* as a model organism, readers are referred to Corsi et al. [1].

A comprehensive knowledge infrastructure has been developed within the worm community, with many resources, research methods, and protocols freely accessible online. "WormBase" is the community's centralized database, and next to its elaborate *C. elegans* data, it contains information on many other nematode species (http://www.wormbase.org). The "WormBook" (http://www.wormbook.org) is an open-access collection of peer-reviewed chapters that cover all kinds of different topics and protocols related to *C. elegans*. The anatomy of *C. elegans* has been studied in great depth and detailed microscopic images of these transparent animals are available on "WormAtlas" (http://www.wormatlas.org), together with a plethora of detailed schematic representations. Additionally, a 3D representation of the *C. elegans* connectome is available at OpenWorm (http://www.openworm.org), an open-source project that aspires to build a comprehensive computational model of the worm. *C. elegans* was also the first multicellular organism to have its genome fully sequenced [2]. The genome is about 100 Mb in size but contains a similar number of

genes (20,532; WormBase WS235 release) as the human genome (~20,500) [3]. Recent estimates relying on primary sequence information predict about 38% similarity between human and *C. elegans* genes [4]. *C. elegans* is an ideal organism to investigate processes that are relevant to human physiology such as aging. In addition, *C. elegans* is often used as a model system for parasitic nematodes.

### 1.2 Peptidomics of C. elegans

Neuropeptides are small messenger molecules that are derived from larger precursor proteins by the action of various processing enzymes. These biologically active peptides can be found in all metazoan species where they orchestrate a wide variety of physiological processes. The knowledge of the primary amino acid sequence of the neuropeptidergic signaling molecules is absolutely necessary to understand their function and interactions with G-protein coupled receptors. Three classes of neuropeptide-encoding genes have been inferred from the genomic data of *C. elegans*: these three groups are referred to as *flp* genes, *nlp* genes, and the insulin superfamily. Thirty-one FMRFamide-like peptide (*flp*) genes were predicted by searching genomic sequences and sequences of mRNA transcripts such as cDNA libraries and Expressed Sequence Tag (EST) data [5–10] (Table 1). Combining in silico searches of the *C. elegans* genome with biochemical methods such as peptidomics, 52 neuropeptide-like protein (*nlp*) genes have been identified (Table 2) [11–16]. These neuropeptide preproproteins all contain peptides without the RFamide motif, but with sequence homology to other invertebrate neuropeptides. This group includes *ntc-1*, *pdf-1*, and *snet-1* [17–19]. Additionally, bioinformatic analyses were able to identify 40 divergent members of the insulin superfamily. These putative insulin-like (*ins-1* to *ins-39*, and *daf-28*) genes are also catalogued as neuropeptides due to their expression in sensory neurons [20]. Based on sequence information alone, one cannot deduce whether all the predicted peptides are actually expressed and properly processed into mature, active peptides. Therefore, mature neuropeptides need to be identified in biological samples. Initial efforts in the field struggled especially with small-sized animals, such as *C. elegans*, and in the early days only 12 neuropeptides of *C. elegans* could be biochemically isolated and identified using Edman degradation analysis or gas-phase sequencing [21–26]. Our efforts of the last decade have helped to systematically search for and characterize neuropeptides of *C. elegans* using peptidomics, relying on liquid chromatography and mass spectrometry. We continuously aim to complete our understanding of which peptides are actually present in the nematode and to identify their post-translational modifications, which are often required for a peptide's bioactivity—all of this in the context of modeling the relevance of specific neuropeptides and their processing in fundamental processes regulating behavior and physiology [27–35]. In this chapter, we introduce the basic techniques and methods required to culture the nematodes and give a detailed protocol with which to perform the sample preparation. We then focus on online LC-MS analysis and describe subsequent data analysis approaches.

**Table 1**
**FMRFamide-like neuropeptides in *C. elegans***

| Gene | Peptide sequence[a,b] |
|---|---|
| *flp-1* | **SADPNFLRFamide** |
| | **SQPNFLRFamide** |
| | **ASGDPNFLRFamide** |
| | **SDPNFLRFamide** |
| | **AAADPNFLRFamide** |
| | (K)PNFLRFamide |
| | **AGSDPNFLRFamide** |
| | (K)PNFMRYamide |
| *flp-2* | SPREPIRFamide |
| | LRGEPIRFamide |
| *flp-3* | **SPLGTMRFamide** |
| | **TPLGTMRFamide** |
| | **EAEEPLGTMRFamide** |
| | **NPLGTMRFamide** |
| | **ASEDALFGTMRFamide** |
| | EDGNAPFGTMRFamide |
| | **EDGNAPFGTMKFamide** |
| | **SAEPFGTMRFamide** |
| | **SADDSAPFGTMRFamide** |
| | **NPENDTPFGTMRFamide** |
| *flp-4* | (GLRSSNGK)PTFIRFamide |
| | ASPSFIRFamide |
| *flp-5* | GAKFIRFamide |
| | AGAKFIRFamide |
| | APKPKFIRFamide |
| *flp-6* | **KSAYMRFamide** |
| | **pQQDSEVEREMM** |
| *flp-7* | **SPMQRSSMVRFamide** |
| | **TPMQRSSMVRFamide** |
| | SPMERSAMVRFamide |

(continued)

**Table 1**
**(continued)**

| Gene | Peptide sequence[a,b] |
|---|---|
| | SPMDRSKMVRFamide |
| | SSIDRASMVRLamide |
| *flp-8* | KNEFIRFamide |
| *flp-9* | **KPSFVRFamide** |
| *flp-10* | pQPKARSGYIRFamide |
| *flp-11* | **AMRNALVRFamide** |
| | **ASGGMRNALVRFamide** |
| | **NGAPQPFVRFamide** |
| | **SPLDEEDFAPESPLQamide** |
| *flp-12* | RNKFEFIRFamide |
| *flp-13* | **AMDSPLIRFamide** |
| | **AADGAPLIRFamide** |
| | **APEASPFIRFamide** |
| | **ASPSAPLIRFamide** |
| | **SPSAVPLIRFamide** |
| | **ASSAPLIRFamide** |
| | **SAAAPLIRFamide** |
| *flp-14* | KHEYLRFamide |
| *flp-15* | **GGPQGPLRFamide** |
| | **RGPSGPLRFamide** |
| *flp-16* | **AQTFVRFamide** |
| | **GQTFVRFamide** |
| *flp-17* | KSAFVRFamide |
| | KSQYIRFamide |
| *flp-18* | (DFD)GAMPGVLRFamide |
| | **EMPGVLRFamide** |
| | (K)**SVPGVLRFamide** |
| | **EIPGVLRFamide** |
| | **SEVPGVLRFamide** |
| | **DVPGVLRFamide** |
| | SYFDEKKSVPGVLRFamide |

(continued)

**Table 1**
**(continued)**

| Gene | Peptide sequence[a,b] |
|---|---|
| *flp-19* | **WANQVRFamide** |
| | **ASWASSVRFamide** |
| *flp-20* | AMMRFamide |
| | AVFRMamide |
| *flp-21* | GLGPRPLRFamide |
| *flp-22* | **SPSAKWMRFamide** |
| *flp-23* | VVGQQDFLRFamide |
| | TKFQDFLRFamide |
| | NDFLRFamide |
| *flp-24* | **VPSAGDMMVRFamide** |
| *flp-25* | **DYDFVRFamide** |
| | **ASYDYIRFamide** |
| *flp-26* | (E)FNADDLTLRFamide |
| | **GGAGEPLAFSPDMLSLRFamide** |
| | **NYYESKPY** |
| *flp-27* | **pQPIDEERPIFME** |
| | (EASAFGDIIGELKGK)GLGGRMRFamide |
| *flp-28* | **APNRVLMRFamide** |
| *flp-32* | AMRNSLVRFamide |
| *flp-33* | APLEGFEDMSGFLRTIDGIQKPRFamide |
| *flp-34* | ALNRDSLVASLNNAERLRFamide |
| | ADISTFASAINNAGRLRYamide |

[a]Sequences in bold have been confirmed by MS/MS of samples prepared according to the method described in this chapter

[b]Sequences between parentheses are possible alternative forms of the peptide (predicted)

**Table 2**
**Neuropeptide-like peptides in *C. elegans***

| Gene | Peptide sequence[a,b] |
|---|---|
| *nlp-1* | **MDANAFRMSFamide** |
| | **MDPNAFRMSFamide** |
| | **VNLDPNSFRMSFamide** |
| *nlp-2* | SIALGRSGFRPamide |
| | SMAMGRLGLRPamide |
| | SMAYGRQGFRPamide |
| *nlp-3* | **AINPFLDSMamide** |
| | **AVNPFLDSIamide** |
| | **YFDSLAGQSLamide** |
| *nlp-4* | SLILFVILLVAFAAARPVSEEVDRV |
| | DYDPRTEAPRRLPADDDEVDGEDRV |
| | DYDPRTDAPIRVPVDPEAEGEDRV |
| *nlp-5* | SVSQLNQYAGFDTLGGMGLamide |
| | ALSTFDSLGGMGLamide |
| | **ALQHFSSLDTLGGMGFamide** |
| *nlp-6* | (MA)**APKQMVFGFamide** |
| | **YKPRSFAMGFamide** |
| | **AAMRSFNMGFamide** |
| | LIMGLamide |
| *nlp-7* | pQADFDDPRMFTSSFamide |
| | **SMDDLDDPRLMTMSFamide** |
| | MILPSLADLHRYTMYD |
| | **LYLKQADFDDPRMFTSSFamide** |
| *nlp-8* | **AFDRFDNSGVFSFGA** |
| | **AFDRMDNSDFFGA** |
| | **SFDRMGGTEFGLM** |
| | **YPYLIFPASPSSGDSRRLV** |
| *nlp-9* | **GGARAFYGFYNAGNS** |
| | GGGRAFNHNANLFRFD |
| | GGGRAFAGSWSPYLE |

(continued)

**Table 2**
**(continued)**

| Gene | Peptide sequence[a,b] |
|---|---|
| | **TPIAEAQGAPEDVDDRRELE** |
| *nlp-10* | AIPFNGGMYamide |
| | STMPFSGGMYamide |
| | AAIPFSGGMYamide |
| | GAMPFSGGMYamide |
| *nlp-11* | HISPSYDVEIDAGNMRNLLDIamide |
| | **SAPMASDYGNQFQMYNRLIDAamide** |
| | **SPAISPAYQFENAFGLSEALERAamide** |
| *nlp-12* | **DYRPLQFamide** |
| | **DGYRPLQFamide** |
| *nlp-13* | **NDFSRDIMSFamide** |
| | **SGNTADLYDRRIMAFamide** |
| | **pQPSYDRDIMSFamide** |
| | **SAPSDFSRDIMSFamide** |
| | **SSSMYDRDIMSFamide** |
| | **SPVDYDRPIMAFamide** |
| | **AEDYERQIMAFamide** |
| *nlp-14* | **ALDGLDGSGFGFD** |
| | **ALNSLDGAGFGFE** |
| | **ALDGLDGAGFGFD** |
| | **ALNSLDGQGFGFE** |
| | **ALNSLDGNGFGFD** |
| *nlp-15* | **AFDSLAGSGFDNGFN** |
| | **AFDSLAGSGFGAFN** |
| | **AFDSLAGSGFSGFD** |
| | **AFDSLAGQGFTGFE** |
| | **AFDTVSTSGFDDFKL** |
| *nlp-16* | STEHHRV |
| | SEGHPHE |
| | ATHSPEGHIVAKDDHHGHE |
| | SSDSHHGHQ |

(continued)

**Table 2**
**(continued)**

| Gene | Peptide sequence[a,b] |
|---|---|
| | SVDEHHGHQ |
| | NAEDHHEHQ |
| | SEHVEHQAEMHEHQ |
| | STQEVSGHPEHHLV |
| *nlp-17* | **GSLSNMMRIamide** |
| | pQQEYVQFPNEGVVPCESCNLGTLMRIamide |
| *nlp-18* | **SPYRAFAFA** |
| | ARYGFA |
| | **SPYRTFAFA** |
| | ASPYGFAFA |
| | **SDEENLDFLE** |
| *nlp-19* | IGLRLPNFL |
| | IGLRLPNFLRF |
| | IGLRLPNML |
| | MGMRLPNIIFL |
| *nlp-20* | FAFAFA |
| | **SGPQAHEGAGMRFAFA** |
| | APKEFARFARASFA |
| *nlp-21* | GGARAMLH |
| | **GGARAFSADVGDDY** |
| | **GGARAFYDE** |
| | **GGARAFLTEM** |
| | **GGARVFQGFEDE** |
| | **GGARAFMMD** |
| | GGGRAFGDMM |
| | **GGARAFVENS** |
| | GGGRSFPVKPGRLDD |
| | pQYTSELEEDE |
| *nlp-22* | SIAIGRAGFRPamide |
| *nlp-23* | LYISRQGFRPA |

(continued)

**Table 2**
**(continued)**

| Gene | Peptide sequence[a,b] |
|---|---|
| | SMAIGRAGMRPamide |
| | AFAAGWNRamide |
| *nlp-24* | pQWGGGPYGGYGP |
| | GYGGGYGGamide |
| | YGGYGamide |
| | FTGPYGGYGamide |
| | GPYGYGamide |
| | GPYGGGGLVGALLamide |
| | pQWGGGPYGGYGPRGYGGGYGGamide |
| | YGGYGGRGPYGGYGGRGPYGYGamide |
| *nlp-25* | pQWGGGYGNPYGGYamide |
| | pQWGGGWNNGGGYGNPYGGYamide |
| | GGGYGGGYGGGFGAQQAYNVQNAA |
| | IGTEVAEGVLVAEEVSEAIamide |
| *nlp-26* | pQFGFGGQQSFGamide |
| | GGQFGGMQ |
| | GGFNGN |
| | GGFGQQSQFGamide |
| | GGNQFGamide |
| | GGSQFNamide |
| | GGFGFamide |
| | pQFGFGGQQSFGGRGGQFGGMQRGGFNGN |
| | GGSQFNGRGGNQFGamide |
| *nlp-27* | pQWGYGGMPYGGYGGMGGYGMGGYGMGY |
| | MWGSPYGGYGGYGGYGGWamide |
| *nlp-28* | GYGGYamide |
| | GYGGYGGYamide |
| | GMYGGWamide |
| | pQWGYGGYGRGYGGYGGYGRGMYGGYamide |
| | GMYGGYGRGMYGGWamide |

(continued)

**Table 2**
**(continued)**

| Gene | Peptide sequence[a,b] |
|---|---|
| *nlp-29* | pQWGYGGYamide |
| | GYGGYGGYamide |
| | GMYGGYamide |
| | GMYGGWamide |
| | pQWGYGGYGRGYGGYGGYGRGMYGGYamide |
| | GMYGGYGRGMYGGYGRGMYGGWamide |
| *nlp-30* | pQWGYGGYamide |
| | GYGGYGGYamide |
| | GYGGYamide |
| | GMWamide |
| | PYGGYGWamide |
| | pQWGYGGYGRGYGGYGGYGRGYGGYamide |
| | GYGGYGRGMWGRPYGGYGWamide |
| *nlp-31* | pQWGYGGYamide |
| | GYGGYGGYamide |
| | GYGGYamide |
| | GMYGGYamide |
| | PYGGYGWamide |
| | pQWGYGGYGRGYGGYGGYGRGYGGYGGYamide |
| | GYGGYGRGMYGGYGRPYGGYGWamide |
| *nlp-32* | YGGWGamide |
| | GGWamide |
| | GGamide |
| | GYGamide |
| | GGGWGamide |
| | GGGWamide |
| | GGGamide |
| | FGYGGamide |
| | GWamide |
| | YGGWGGRGGWGRGGGRGYGamide |

(continued)

**Table 2**
**(continued)**

| Gene | Peptide sequence[a,b] |
|---|---|
| | GGGWGGRGGGWGRGGGGRGFYGGamide |
| *nlp-33* | pQWGYGGPYGGYGGGYGGGPWGYGGGW |
| | HWGGYGGGPWGGYGGGPWGGYY |
| | RHWGGYGGGPWGGYGGGPWGGYY |
| *nlp-34* | PYGYGGYGGW |
| | PYGYGWamide |
| *nlp-35* | **AVVSGYDNIYQVLAPRF** |
| *nlp-36* | DDDVTALERWGY |
| | NIDMKLGPH |
| | SMVARQIPQTVVADH |
| *nlp-37* | NNAEVVNHILKNFGALDRLGDVamide |
| *nlp-38* | (ASDDR)VLGWNKAHGLWamide |
| | **TPQNWNKLNSLWamide** |
| | **SPAQWQRANGLWamide** |
| *nlp-39* | **EVPNFQADNVPEAGGRV** |
| *nlp-40* | **APSAPAGLEEKLR** |
| | pQPAADTFLGFVPQ |
| *nlp-41* | **APGLFELPSRSV** |
| *nlp-42* | **SALLQPENNPEWNQLGWAWamide** |
| | NPDWQDLGFAWamide |
| *nlp-43* | **KQFYAWAamide** |
| *nlp-44* | APHPSSALLVPYPRVamide |
| | LYMARVamide |
| | AFFYTPRIamide |
| *nlp-45* | RNLLVGRYGFRIamide |
| *nlp-46* | NIAIGRGDGLRPamide |
| *nlp-47* | pQMTFTDQWT |
| *nlp-48* | GVGDVPSMFFSPFRMMamide |
| *nlp-49* | **SPSMGLSLAEYMASPQGGDNFHFMPSamide** |
| *nlp-50* | **TEGLSRASANAYYRLamide** |

(continued)

**Table 2**
**(continued)**

| Gene | Peptide sequence[a,b] |
|---|---|
| *nlp-51* | **SQTQEANIQPFIRF** |
| *nlp-52* | **GDVKSVFFSPFRMVamide** |
| *pdf-1* | **SNAELINGLIGMDLGKLSAVamide** |
| | **SNAELINGLLSMNLNKLSGAamide** |
| | **SPLLYRAPQMYDDVQFV** |
| | **SPLLYRAPQYQMYDDVQFV** |
| *ntc-1*[c] | CFLNSCPYRRYamide |

[a]Sequences in bold have been confirmed by MS/MS of samples prepared according to the method described in this chapter
[b]Sequences between parentheses are possible alternative forms of the peptide (predicted)
[c]In the *ntc-1* peptide, a sulfur bridge is formed between $Cys_1$ and $Cys_6$

## 2 Materials

### 2.1 Culturing *C. elegans*

1. *C. elegans* strains can be ordered from the *Caenorhabditis* Genetics Center (CGC, http://www.cbs.umn.edu/CGC/), which is supported by the National Institutes of Health Office of Research Infrastructure Programs (P40 OD010440), or the National BioResearch Project (NBRP, Japan, http://shigen.nig.ac.jp/c.elegans). These centers collect, maintain, and distribute different *C. elegans* (mutant) strains. *C. elegans* N2 (Bristol) is referred to as the wild-type reference strain. Nematodes ordered at the stock centers are sent by regular post on small petri dishes.
2. *Escherichia coli* OP50 bacteria, available at the CGC.
3. Nematode Growth Medium (NGM): Dissolve 3 g NaCl, 17 g agar, and 7.5 g peptone in 1 L deionized $H_2O$. Sterilize by autoclaving and add 1 mL of 1 M $CaCl_2$, 1 mL of 5 mg/mL cholesterol in ethanol, 1 mL of 1 M $MgSO_4$, and 25 mL of a 1 M $K_2HPO_4/KH_2PO_4$ buffer (3.56%:10.83% w/v, pH 6.0). Pour NGM medium in petri dishes under sterile conditions (*see* **Note 1**).
4. Incubators (20 °C) (*see* **Note 2**).
5. Petri dishes, 90 mm diameter.

### 2.2 Sample Preparation

1. Microcentrifuge.
2. Centrifuge for 15–50 mL tubes.

3. Vacuum concentrator, here SpeedVac vacuum centrifuge (Savant).
4. Glass homogenizer, type Potter-Elvehjem.
5. Sonicator (here Branson Ultrasonic SLPe).
6. Small separatory funnel.
7. 60% sucrose solution. This solution can be stored at 4 °C for a couple of weeks.
8. S basal buffer: Dissolve 5.85 g NaCl, 1 g $K_2HPO_4$, 6 g $KH_2PO_4$ in 1 L deionized $H_2O$. Sterilize by autoclaving.
9. Extraction solvent: methanol, water, acetic acid (90:9:1, v/v/v), make fresh and keep cold on dry ice.
10. *n*-hexane.
11. 100 mM $CH_3COONH_4$ solution, at pH 7 (adjusted with 1 M NaOH).
12. Ultrapure $H_2O$ containing 0.1% formic acid (FA).
13. 50% acetonitrile containing 0.1% FA.
14. 70% acetonitrile containing 0.1% FA.
15. Acetonitrile containing 0.1% FA.
16. Solid-phase extraction cartridges (SPE) (such as Oasis HLB, Waters, Milford, MA, USA).
17. 0.45 μm filters (such as Millex LCR, Millipore, Bedford, MA, USA).
18. 10 kDa cut-off filters (such as Amicon Ultra-4, Millipore).
19. Sephadex size-exclusion columns (such as G-10 Sephadex, PD Miditrap, Waters).
20. $C_{18}$ pipette tips (such as ZipTip, Millipore).

### 2.3 Peptidomics Analysis

The analysis requires an ultra-high performance liquid chromatography (UHPLC) system for chromatographic separation, coupled online to a mass spectrometer for detection and quantification. The equipment we use is described in this section.

1. UHPLC system (Dionex UltiMate 3000, Thermo Scientific, Waltham, MA, USA).
2. $C_{18}$ pre-column (Acclaim PepMap100, $C_{18}$, 75 μm × 20 mm, 3 μm, 100 Å; Thermo Scientific).
3. Integrated $C_{18}$ analytical column/nano-electrospray ion source (EASY-Spray, PepMap RSLC, $C_{18}$, 75 μm × 500 mm, 2 μm, 100 Å; Thermo Scientific).
4. Quadrupole-Orbitrap mass spectrometer (Q Exactive Hybrid Quadrupole-Orbitrap, Thermo Scientific) (*see* **Note 3**).

5. Solvent A: ultrapure $H_2O$ containing 0.1% FA.
6. Solvent B: 80% acetonitrile containing 0.08% FA. All solvents need to be HPLC grade.

### 2.4 Data Processing and Peptide Identification

Several software solutions can be used to process MS data. We rely on:

1. *Mascot* search engine (www.matrixscience.com).
2. *PEAKS* software (www.bioinfor.com).
3. *ProteoWizard* software (http://proteowizard.sourceforge.net/).

## 3 Methods

### 3.1 Maintenance of C. elegans Cultures

This section provides a concise description of some basic methods used in culturing *C. elegans*.

1. *C. elegans* is normally grown on solid NGM medium containing the *E. coli* OP50 strain as a food source (*see* **Note 4**). Bacteria can be grown by inoculating LB broth and subsequent incubation at 37 °C.
2. Using sterile technique, apply 100 μL of an OP50 culture grown overnight on prepared NGM petri dishes. Spread the bacteria over the surface and let it dry. Grow overnight in a 37 °C incubator to form a visible OP50 lawn (*see* **Notes 5–8**).
3. Several methods exist to transfer worm cultures from an old plate to a fresh one. Using sterile technique, a small piece of agar can be cut from the old plate and transferred to a new one. Alternatively, it is also possible to pick individual animals using a "worm picker," which consists of a flat-ended platinum wire welded into a glass Pasteur pipette.
4. To maintain the strains, worms should be transferred to fresh plates 2–3 times a week (*see* **Note 9**).

### 3.2 Sample Preparation

1. Collect mixed-stage or age-synchronized (*see* **Note 10**) worms from 15 to 20 fully grown petri dishes by rinsing the plates with S basal buffer.
2. Living animals are separated from *E. coli* bacteria and dead animals by flotation on 30% sucrose. Add an equal volume of cold 60% sucrose solution (~4°C) to the S basal containing the worms and mix. Gently add 1 mL of S basal buffer, two phases will form due to the difference in density. Centrifuge for 4 min at 300 × *g*. Living animals will float on top of the sugar gradient. This step relies on density of the solvents and worms, which is temperature dependent. If the worms do not form a distinct layer or pellet, consider also keeping the S basal buffer on ice. Collect the animals by aspirating the top S basal layer (*see* **Note 11**), and wash by first pelleting the worms (centrifuge 3 min at 300 × *g*). Next, remove the supernatant and add

fresh S basal buffer. Repeat this procedure three times or until the supernatant is clear.

3. After the last washing step, remove as much supernatant as possible, keeping only the pelleted worms. From experience, we suggest that worm pellet volumes ranging from 300 to 600 μL are sufficient for successful extractions.
4. Transfer the nematodes to 10 mL of extraction solution, cooled on dry ice (*see* **Notes 12** and **13**), the worm pellet will flash freeze but the extract solution should remain liquid.
5. The worms are homogenized on dry ice using the glass homogenizer. Sonicate the homogenized solution three times for 20–30 s (*see* **Note 14**).
6. Centrifuge the sample for 12 min at 2100 × *g*. Discard the pellet and evaporate the methanol using a vacuum concentrator (*see* **Note 15**).
7. The remaining aqueous solution, which contains the peptides, is delipidated by re-extraction with *n*-hexane. In a small separatory funnel, add an equal volume of *n*-hexane to the aqueous sample and mix by vigorous inversion. Wait until the layers separate and carefully collect the bottom (aqueous) layer (*see* **Note 16**).
8. Some residual *n*-hexane may be in solution in the aqueous phase; evaporate this in the vacuum concentrator (±30 min.).
9. Rinse the Amicon Ultra-4 10 kDa filters with 4 mL of ultrapure $H_2O$ and centrifuge for 10 min at 4000 × *g*. Add the sample and centrifuge at 4000 × *g* until the majority of the sample has run through (20–40 min) (*see* **Note 17**).
10. Prepare the Sephadex size-exclusion columns by adding 16 mL of 100 mM $CH_3COONH_4$. Add the sample to the column (maximum 1 mL, *see* specifications of the column) and let it enter the packed bed. Add $CH_3COONH_4$ solution until the total volume of sample and buffer equals 1.7 mL. Discard the flow-through and elute the sample with 1.2 mL $CH_3COONH_4$ solution.
11. Further cleanup of the samples is done using solid-phase extraction cartridges. Activate the cartridge using acetonitrile containing 0.1% FA, then, rinse with water containing 0.1% FA. Add the sample and wash again with 0.1% FA in water. Elute the peptides with 50% acetonitrile containing 0.1% FA.
12. Samples are concentrated for LC-MS analysis using $C_{18}$ pipette tips. Activate the pipette tips by aspirating acetonitrile containing 0.1% FA (repeat three times). Rinse the column with water containing 0.1% FA (repeat three times). Aspirate the sample, and let it flow three times over the column bed,

repeat this for the whole sample. Rinse the column again three times with water containing 0.1% FA. Finally, elute with 50% acetonitrile containing 0.1% FA.

13. Samples can be stored at 4 °C prior to LC-MS analysis.

### 3.3 Liquid Chromatography-Mass Spectrometry

1. Since samples are stored in 50% acetonitrile containing 0.1% FA, excess acetonitrile is evaporated in the vacuum concentrator. Then, 2% acetonitrile containing 0.1% FA is added to a total volume of 15 µL.
2. The UHPLC system is set to draw 5 µL of sample and load it on the pre-column with a flow of 5 µL/min.
3. Peptides are separated by using a gradient ranging from 4 to 35% solvent B for 200 min, rising to 100% solvent B the following 25 min.
4. MS data is typically acquired using a data-dependent Top10 method choosing the most abundant precursor ions from a full MS survey scan for fragmentation by high-energy collisional dissociation (HCD). Full MS scans are acquired at a resolution of 70,000 at *m/z* 200, with a maximum injection time of 256 ms. The resolution for MS/MS scans after HCD fragmentation is set at 17,500 at *m/z* 200, with a maximum injection time of 64 ms. If one has an interest in specific peptides, rather than in overall profiling, targeted methods for MS/MS analysis can be applied.

### 3.4 Data Processing and Peptide Identification

1. MS/MS data can be transformed into *mgf* and *mzXML* files using freely available software such as ProteoWizard (available at http://proteowizard.sourceforge.net/).
2. The resulting *mgf* files can be submitted to a *Mascot* search, which matches the fragmentation data against any sequence database (*see* **Note 18**). As variable modifications select: amino-terminal pyroglutamic acid (from Glu and Gln), carboxy-terminal Gly-loss and amidation, and methionine oxidation. As enzyme, select "None" and leave the field with fixed modifications empty. Peptide mass tolerance and MS/MS fragment mass tolerance depend on the mass spectrometer used. In the case of the workflow described above, it would be set to 10 ppm and 15 mu respectively.
3. Unknown post-translational peptide modifications may complicate identification based on MS/MS data. This can be overcome by using commercial software like *PEAKS* (www.bioinfor.com), which combines an in silico spectrum database search with de novo sequencing of the raw MS/MS data. This combination will further increase the number of reliable peptide identifications.

## 4 Notes

1. Petri dishes that allow airflow under the lid are preferred for sufficient oxygen supply and to avoid condensation. Depending on the amount of plates needed, a peristaltic pump can be used to pour the NGM.
2. *C. elegans* is normally cultured at 20 °C. Lowering the temperature (e.g., to 15 °C) will slow down development, resulting in slower growth. This delay may be useful, depending on planning of experiments, but bear in mind it may also affect peptide profiles.
3. The Dionex UltiMate 3000 UHPLC system is coupled online to the Q Exactive mass spectrometer via the EASY-Spray analytical column and integrated nano-electrospray ion source.
4. The *E. coli* OP50 strain is uracil auxotroph, and has a limited growth on NGM plates. A thin bacterial lawn facilitates the observation of nematodes when using a microscope.
5. Damaging the surface of the NGM should be avoided since worms tend to crawl into the agar. When spreading the bacteria, try not to cover the edges of the surface. If the bacterial lawn reaches the edges of the petri dish, worms may crawl up the sides and die.
6. Depending on the experiment, bacteria can be grown for shorter or longer periods of time to produce a thinner or thicker lawn. The plates can also be left at room temperature for approximately 2 days, which will also result in a homogenous lawn.
7. NGM petri dishes with an OP50 lawn can be stored at 4 °C for several weeks, although it is better to use freshly made plates for each experiment.
8. When the plates are stored at 4 °C, allow them first to equilibrate at 20 °C for a certain amount of time before using them for culturing the nematodes.
9. The frequency of transferring worms to fresh plates depends on the size of the chunks, the dimension of the petri dish, and the growth temperature.
10. A synchronous *C. elegans* culture can be obtained by treatment of gravid hermaphrodites with a sodium hypochlorite/sodium hydroxide mixture to release the eggs. After a sucrose flotation step, which separates the eggs from the debris, eggs are transferred to S basal buffer and incubated at 20 °C for 24 h. Due to absence of food, the hatched eggs will arrest in the L1 stadium, which then can be transferred to NGM plates containing OP50.

11. Fifteen to twenty fully grown petri dishes yield a pellet of 300–500 μL of living nematodes.
12. The extraction solvent is designed to extract small endogenous peptides, while large proteins precipitate. Optionally, when interested in larger peptides (5–15 kDa) such as insulin-like peptides, one may achieve better results with diluted acids as an extraction solvent.
13. The homogenization and sonication steps are performed on dry ice to help inactivate the peptidases present in the sample. Active peptidases will degrade proteins and possibly shorten and/or fragment peptides, which are obviously not of interest.
14. Sonication of the sample will aid in disruption of the cell membrane, thereby releasing cellular contents into the extraction solvent. Since sonication heats up the sample, it is recommended to place the sample on dry ice and only sonicate for short intervals.
15. It is preferable not to dry the sample to completion in the vacuum concentrator, as it is often difficult to reconstitute the dry pellet. This could lead to considerable sample loss.
16. If no separatory funnel is available, delipidation can be carried out in a standard tube. When both phases have settled, aspirate the top organic layer with a pipette. Make sure no residual hexane remains in the sample by visually inspecting the tube.
17. Since most peptides of interest are within the 0.7–10 kDa range, we usually enrich this mass range by applying a Sephadex size-exclusion column followed by a 10 kDa molecular weight cut-off filter. This ensures the removal of molecules that are not of direct interest, and otherwise would interfere with the LC-MS analysis. Depending on the peptides of interest, the type of molecular weight cut-off filters can be adapted.
18. A local *Mascot* server has the advantage of custom-made sequence databases. We use in-house developed FASTA databases containing *C. elegans* neuropeptide precursors as well as a database with in silico cleaved neuropeptides. To check for potential false positive identifications, an additional database containing all proteins from *C. elegans* and *E. coli* along with a list of common contaminants (e.g., keratin) is used.

## Acknowledgments

SVB holds an IWT-Flanders fellowship; LT is an FWO-Flanders research fellow. The authors wish to thank KU Leuven (C14/15/049) and FWO-Flanders (G069713) for funding.

## References


1. Corsi AK (2015) A Transparent window into biology: a primer on *Caenorhabditis elegans*. WormBook:1–31. https://doi.org/10.1895/wormbook.1.177.1
2. C. elegans Sequencing Consortium (1998) Genome sequence of the nematode *C. elegans*: a platform for investigating biology. Science 282:2012–2018
3. Clamp M, Fry B, Kamal M et al (2007) Distinguishing protein-coding and noncoding genes in the human genome. Proc Natl Acad Sci 104:19428–19433. https://doi.org/10.1073/pnas.0709013104
4. Shaye DD, Greenwald I (2011) OrthoList: a compendium of C. Elegans genes with human orthologs. PLoS One 6:e20085. https://doi.org/10.1371/journal.pone.0020085
5. Kim K, Li C (2004) Expression and regulation of an FMRFamide-related neuropeptide gene family in *Caenorhabditis elegans*. J Comp Neurol 475:540–550. https://doi.org/10.1002/cne.20189
6. Li C, Nelson LS, Kim K et al (1999) Neuropeptide gene families in the nematode *Caenorhabditis elegans*. Ann N Y Acad Sci 897:239–252
7. Li C, Kim K, Nelson LS (1999) FMRFamide-related neuropeptide gene family in *Caenorhabditis elegans*. Brain Res 848:26–34. https://doi.org/10.1016/S0006-8993(99)01972-1
8. McVeigh P, Leech S, Mair GR et al (2005) Analysis of FMRFamide-like peptide (FLP) diversity in phylum Nematoda. Int J Parasitol 35:1043–1060. https://doi.org/10.1016/j.ijpara.2005.05.010
9. Li C, Kim K (2014) Family of FLP peptides in *Caenorhabditis elegans* and related nematodes. Front Endocrinol 5:150. https://doi.org/10.3389/fendo.2014.00150
10. Peymen K, Watteyne J, Frooninckx L et al (2014) The FMRFamide-like peptide family in nematodes. Front Endocrinol. https://doi.org/10.3389/fendo.2014.00090
11. Nathoo AN, Moeller RA, Westlund BA, Hart AC (2001) Identification of neuropeptide-like protein gene families in *Caenorhabditis elegans* and other species. Proc Natl Acad Sci U S A 98:14000–14005. https://doi.org/10.1073/pnas.241231298
12. Hobert O (2013) The neuronal genome of *Caenorhabditis elegans*. WormBook:1–106. https://doi.org/10.1895/wormbook.1.161.1
13. Husson SJ, Mertens I, Janssen T et al (2007) Neuropeptidergic signaling in the nematode *Caenorhabditis elegans*. Prog Neurobiol 82:33–55. https://doi.org/10.1016/j.pneurobio.2007.01.006
14. McVeigh P, Alexander-Bowman S, Veal E et al (2008) Neuropeptide-like protein diversity in phylum Nematoda. Int J Parasitol 38:1493–1503. https://doi.org/10.1016/j.ijpara.2008.05.006
15. Lindemans M, Liu F, Janssen T et al (2009) Adipokinetic hormone signaling through the gonadotropin-releasing hormone receptor modulates egg-laying in *Caenorhabditis elegans*. Proc Natl Acad Sci U S A 106:1642–1647. https://doi.org/10.1073/pnas.0809881106
16. Husson SJ, Reumer A, Temmerman L et al (2014) Worm peptidomics. EuPA Open Proteom 3:280–290. https://doi.org/10.1016/j.euprot.2014.04.005
17. Beets I, Janssen T, Meelkop E et al (2012) Vasopressin/oxytocin-related signaling regulates gustatory associative learning in *C. elegans*. Science 338(80):543–545. https://doi.org/10.1126/science.1226860
18. Janssen T, Husson SJ, Meelkop E et al (2009) Discovery and characterization of a conserved pigment dispersing factor-like neuropeptide pathway in *Caenorhabditis elegans*. J Neurochem 111:228–241. https://doi.org/10.1111/j.1471-4159.2009.06323.x
19. Yamada K, Hirotsu T, Matsuki M et al (2010) Olfactory plasticity is regulated by pheromonal signaling in *Caenorhabditis elegans*. Science 329(80):1647–1650. https://doi.org/10.1126/science.1192020
20. Pierce SB, Costa M, Wisotzkey R et al (2001) Regulation of DAF-2 receptor signaling by human insulin and ins-1, a member of the unusually large and diverse *C. elegans* insulin gene family. Genes Dev 15:672–686. https://doi.org/10.1101/gad.867301
21. Marks NJ, Shaw C, Maule AG et al (1995) Isolation of AF2 (KHEYLRFamide) from *Caenorhabditis elegans*: evidence for the presence of more than one FMRFamide-related peptide-encoding gene. Biochem Biophys Res Commun 217:845–851. https://doi.org/10.1006/bbrc.1995.2849
22. Marks NJ, Maule AG, Geary TG et al (1997) APEASPFIRFamide, a novel FMRFamide-related decapeptide from *Caenorhabditis elegans*: structure and myoactivity. Biochem Biophys Res Commun 231:591–595. https://doi.org/10.1006/bbrc.1997.6155
23. Marks NJ, Maule AG, Geary TG et al (1998) KSAYMRFamide (PF3/AF8) is present in the

free-living nematode, *Caenorhabditis elegans*. Biochem Biophys Res Commun 248:422–425. https://doi.org/10.1006/bbrc.1998.8982

24. Marks NJ, Maule AG, Li C et al (1999) Isolation, pharmacology and gene organization of KPSFVRFamide: a neuropeptide from *Caenorhabditis elegans*. Biochem Biophys Res Commun 254:222–230. https://doi.org/10.1006/bbrc.1998.9920
25. Marks NJ, Shaw C, Halton DW et al (2001) Isolation and preliminary biological assessment of AADGAPLIRFamide and SVPGVLRFamide from *Caenorhabditis elegans*. Biochem Biophys Res Commun 286:1170–1176. https://doi.org/10.1006/bbrc.2001.5524
26. Rosoff ML, Doble KE, Price DA, Li C (1993) The flp-1 propeptide is processed into multiple, highly similar FMRFamide-like peptides in *Caenorhabditis elegans*. Peptides 14:331–338. https://doi.org/10.1016/0196-9781(93)90049-M
27. Husson SJ, Clynen E, Baggerman G et al (2005) Discovering neuropeptides in *Caenorhabditis elegans* by two dimensional liquid chromatography and mass spectrometry. Biochem Biophys Res Commun 335:76–86. https://doi.org/10.1016/j.bbrc.2005.07.044
28. Husson SJ, Clynen E, Baggerman G et al (2005) Peptidomics of *Caenorhabditis elegans*: in search of neuropeptides. Commun Agric Appl Biol Sci 70:153–156
29. Husson SJ, Clynen E, Baggerman G et al (2006) Defective processing of neuropeptide precursors in *Caenorhabditis elegans* lacking proprotein convertase 2 (KPC-2/EGL-3): mutant analysis by mass spectrometry. J Neurochem 98:1999–2012. https://doi.org/10.1111/j.1471-4159.2006.04014.x
30. Husson SJ, Schoofs L (2006) Characterization of a key neuropeptide processing enzyme in *C. elegans* by mass spectrometry. Commun Agric Appl Biol Sci 71:171–174
31. Husson SJ, Janssen T, Baggerman G et al (2007) Impaired processing of FLP and NLP peptides in carboxypeptidase E (EGL-21)-deficient *Caenorhabditis elegans* as analyzed by mass spectrometry. J Neurochem 102:246–260. https://doi.org/10.1111/j.1471-4159.2007.04474.x
32. Husson SJ, Schoofs L (2007) Altered neuropeptide profile of *Caenorhabditis elegans* lacking the chaperone protein 7B2 as analyzed by mass spectrometry. FEBS Lett 581:4288–4292. https://doi.org/10.1016/j.febslet.2007.08.003
33. De Haes W, Van Sinay E, Detienne G et al (2015) Functional neuropeptidomics in invertebrates. Biochim Biophys Acta 1854:812–826. https://doi.org/10.1016/j.bbapap.2014.12.011
34. Husson SJ, Clynen E, Boonen K et al (2010) Approaches to identify endogenous peptides in the soil nematode *Caenorhabditis elegans*. Methods Mol Biol 615:29–47. https://doi.org/10.1007/978-1-60761-535-4_3
35. Husson SJ, Landuyt B, Nys T et al (2009) Comparative peptidomics of *Caenorhabditis elegans* versus *C. briggsae* by LC-MALDI-TOF MS. Peptides 30:449–457. https://doi.org/10.1016/j.peptides.2008.07.021

# Chapter 19

# EndoProteoFASP as a Tool to Unveil the Peptidome-Protease Profile: Application to Salivary Diagnostics

**Fábio Trindade, Inês Falcão-Pires, Adelino Leite-Moreira, Pedro S. Gomes, Julie Klein, Rita Ferreira, and Rui Vitorino**

## Abstract

In the quest to fully comprehend the proteolytic events leading to the generation of the salivary peptidome, we have developed a method for the sequential elution of salivary peptides throughout progressive endogenous proteolysis. By screening the time-dependent changes in the salivary peptidome we can predict the activity pattern of salivary proteases responsible for such peptide fingerprint and identify susceptible protein targets. Herein, we describe a step-by-step tutorial based on a filter-aided sample preparation (FASP) method, taking advantage of the endogenous salivary proteases armamentarium (endoProteoFASP), to produce new peptides from the salivary proteins, adding to those present in the sample at the time of collection. In this protocol, the different sets of peptides retrieved after sample elution are identified following a liquid chromatography-tandem mass spectrometry approach. The likelihood of a large set of endogenous proteases (collected from several public sources) to be responsible for the generation of such peptides can be predicted by the analysis of the cleavage site specificity by Proteasix (http://proteasix.cs.man.ac.uk/) algorithm. The attained peptidome-protease profile can be useful to elucidate the peptidome dynamics and the proteolytic events underpinning pathophysiological phenomena taking place locally within the oral cavity. This may help clinicians to diagnose oral pathologies and develop preventive therapeutic plans.

**Key words** EndoProteoFASP, FASP, Peptidomics, Peptidome, Protease, Saliva, Proteasix

## 1 Introduction

The disclosure of the proteolytic events architecting the human salivary peptidome remains an incomplete task. This is because the oral cavity displays a myriad of endo- and exoproteases that are responsible for the generation of thousands of small peptides, in a process that starts before secretion and extends beyond protein release and combination with the remaining salivary components [1, 2]. Several attempts have been made to understand this complexity following two main strategies. The first is based on peptidomic strategies, mainly liquid chromatography (LC) and/or mass spectrometry (MS) using matrix-assisted laser desorption ionization (MALDI) with tandem time-of-flight (TOF/TOF) followed by

Michael Schrader and Lloyd Fricker (eds.), *Peptidomics: Methods and Strategies*, Methods in Molecular Biology, vol. 1719, https://doi.org/10.1007/978-1-4939-7537-2_19, 

peptide cleavage sites analysis for protease prediction. The second resorts to zymography, followed by proteomics (LC-MS/MS), in an attempt to identify active proteases in zymograms [3–6]. Each strategy has its own strengths. In the former, direct separation and analysis of peptides through LC and MS reduces sample manipulation and eases protocol automation. In the latter, gel permeation allows the reduction of sample complexity and simultaneously provides robustness against sample contaminants that would hamper total protein digestion.

This chapter describes the advantages conferred by filter-aided sample preparation (FASP) techniques. FASP not only guarantees sample clean-up but also provides a reactor-like environment for protein degradation and a suitable platform for sequential elution of digests [7, 8]. This observation is the rationale behind the design of a new approach for the study of salivary degradomics, which is a bridge between salivary peptidome and salivary proteases. The endoProteo-FASP method described in this chapter relies on the endogenous salivary protease pool to produce new peptides, beyond those making part of the natural peptidome [9]. Then, following peptide elution and identification, cleavage site analysis by Proteasix allows the prediction of implicated proteases and the ability to relate specific peptidome-proteases profiles with different time-points of collection [10]. The versatility of endoProteoFASP is explained by the dual use of a 30 kDa cut-off filter as both a sample cleaner device and as a proteomic reactor. In this method saliva samples are first loaded onto spin filters and the natural peptides are eluted by centrifugation. Peptide elution results in concentrating high molecular weight material and consequently in protease concentration. Thus, the activity of salivary proteases is enhanced by concentration effect, as well as by incubation at 37 °C. When used this way, the spin filter acts as proteomic reactor, yielding new peptides that can be eluted sequentially. Eluted peptide fractions can then be separated and identified by high performance LC (HPLC) coupled to tandem MS. In the end, the time-dependent peptidome profile can be analyzed by Proteasix algorithm, in order to predict which proteases are active at the time of collection and those who remain active after incubation. Additionally, it is possible to use the sample retentate to perform zymography and directly assess the proteolytic activity in saliva. Hence, with endoProteoFASP we can associate specific peptidomic data with the predicted (Proteasix) or experimentally validated (zymography) activity of proteases. Also, this can be performed without the use of exogenous synthetic peptides (e.g., as in enzymatic assays) or proteases (e.g., as in trypsin-based shotgun proteomics) which provides a reliable platform for the characterization of the proteolytic phenomena taking place in the oral cavity. We have already demonstrated the utility of this technique to uncover potential markers for the diagnosis of chronic periodontitis [10], but it can be further useful to understand the proteolytic events underlying several oral conditions and to pinpoint new markers for salivary diagnostics.

## 2 Materials

Saliva samples should be immediately frozen at −80 °C, if they are not processed right after collection. Peptide fractions should be kept on ice, if they cannot be readily vacuum-dried after elution. All solutions should be prepared with ultrapure water (18.2 MΩ cm). Solvents used should be LC- and MS-compatible.

### 2.1 Saliva Collection and Processing

1. Sterile 50 mL tubes.
2. 1.5 mL or 2 mL tubes.
3. Refrigerated benchtop centrifuge.
4. Colorimetric assay for determination of protein concentration—e.g., DC kit (BioRad®).

### 2.2 Collection of Peptides

1. 30 kDa cut-off spin filters.
2. 2 mL microfuge tubes.
3. Benchtop centrifuge.
4. Centrifugal vacuum evaporator.
5. Equilibration solution: 50 mM ammonium bicarbonate (ABC) solution ($NH_4HCO_3$).
6. SDS solution: 1% (w/V) sodium dodecyl sulfate.
7. TFA solution: 0.1% (V/V) trifluoroacetic acid.

### 2.3 Peptide Separation and Identification by nanoHPLC-MALDI-TOF/TOF

1. Sonicator bath.
2. nanoHPLC apparatus equipped with a 150 mm × 75 μm capillary C18 column with 3 μm particle size and with a UV detector.
3. Automatic fraction collector robot (such as the Probot™ with μCarrier 2.0 software, Dionex Corporation, Amsterdam, The Netherlands).
4. MALDI plates.
5. MALDI-TOF/TOF spectrometer (4800 Proteomics Analyzer, Applied Biosystems, Foster City, CA, USA) and respective software (4000 Explorer software®, v3.5, Applied Biosystems, Foster City, CA).
6. Protein identification software (Mascot, Matrix Science Ltd., UK).
7. 50:49.7:0.3 acetonitrile (ACN)/water/TFA (V/V/V) solution.
8. TFA solution: 0.1% (V/V) trifluoroacetic acid.
9. HPLC mobile phase A: water (98%), 2% ACN, 0.1% TFA.
10. HPLC mobile phase B: water (5%), 95% ACN, 0.045% TFA.
11. MALDI matrix solution: 3 mg/mL of α-cyano-4-hydroxycinnamic acid in 70% ACN and 0.1% TFA, spiked with an internal standard Glu-Fib (15 fmol/uL) for internal calibration of the MS device.

## 3 Methods

The methodological approach described in this chapter is suitable to study the endogenous proteolytic activity in saliva, but it is likely to be useful in other biofluids. Without the need for addition of exogenous proteins or proteases, EndoProteoFASP makes the most of saliva's protease pool to generate peptides ex vivo from the peptide-depleted salivary matrix, being the own salivary proteins the targets of such enzymatic activity. The sequence of the peptides (either produced in vivo or artificially ex vivo) can be used for protease prediction. Furthermore, in this approach we can use affordable reagents and materials and extensive sample manipulation is not required. Time consumption will be defined only by the time of incubation (from some hours to several days) and the number of samples, which will require more runs in the LC-MS/MS apparatus.

Once peptides are identified, Proteasix analysis can be performed. This software relies on the alignment of the input peptides with the full-length sequence (from databases such as SwissProt) to determine the *N*- and *C*-termini cleavage sites (CS). Then, each CS is crossed with a CS database in order to collect all predicted CS-protease combinations [11]. The protease output will be divided into five sections: observed, predicted, and predicted with cleavage of one (input +1 AA), two (input +2 AA), and three amino acids (input +3 AA), catalyzed by exopeptidases. For predicted associations, the degree of confidence will be depicted from low to medium to high confidence, depending on specificity weight matrices of the MEROPS database (http://merops.sanger.ac.uk/, this and other useful websites and their main application are summarized in Table 1). Proteasix has an ontology program (Proteasix Ontology, PxO) which associates gene ontology location terms to retrieved proteases. Although it is not yet used in the prediction algorithm, PxO helps the user to attribute a biological meaning to the protease output [12]. The workflow to perform EndoProteoFASP + Proteasix analysis is depicted in Fig. 1.

### 3.1 Saliva Collection and Processing

1. Collect unstimulated whole saliva in 50 mL tubes by passive drooling, or use specially developed devices for saliva collection (e.g., Salivette®). Sampling should be conducted in early morning from subjects that have not drunk fluids or eaten for at least 60 min.
2. Centrifuge samples at 12,000 × *g* for 30 min (at 4 °C) in 1.5 mL or 2 mL microfuge tubes (*see* **Note 1**).
3. Collect the supernatant carefully and discard the *pellet*. Aliquot the supernatants in several fractions and store at −80 °C, if not immediately proceeding to the following steps (*see* **Note 2**).
4. Determine the protein concentration using the colorimetric assay (*see* **Note 3**).

**Table 1**
**Useful websites and webtools and their main application**

| Website/Webtool | Application |
|---|---|
| Mascot http://www.matrixscience.com/search_form_select.html | Search engine that uses MS data to identify proteins from primary sequences databases. It can also identify post-translational modifications. It is possible to perform automatic decoy search in order to calculate false discovery rates |
| Jvenn http://bioinfo.genotoul.fr/jvenn/example.html | A Venn diagram-maker tool that allows to process up to six peptide/protein datasets and retrieve the list of unique peptides for each set |
| MEROPS http://merops.sanger.ac.uk/ | Protease database, grouping these enzymes in families and clans based on sequence homology. It detains specific information about cleavage site specificity, physiological and pathological relevance. It also displays KEGG pathway annotations |
| CutDB http://cutdb.burnham.org/ | It is a proteolytic events database, either experimentally validated or predicted, that allows the user to search for specific events from several types of inputs: substrate, protease, disease, among others |
| BRENDA http://www.brenda-enzymes.org/ | Large repository of proteases with specific information regarding functional parameters, stability, structure, reaction characteristics, and implication in disease |
| Pfam http://pfam.xfam.org/ | Protein database with focus on functional domains. It provides detailed structural, functional, and evolutionary information of different protein families and clans |
| Prosite http://prosite.expasy.org/ | Protein domains, families, and functional sites database. It is possible to search for information about specific domains or to detect them from inputted protein sequences |
| STRING http://string-db.org/ | Web application that retrieves for one or a group of proteins (e.g., proteases) predicted and validated protein interactions as well as their intervention on biological processes, molecular functions, or KEGG pathways |
| DisGeNET http://www.disgenet.org/ | Platform gathering gene-disease associations across several databases and literature. It allows to search for each protein/protease already described associations to diseases and, thus, infer about its potential biomarker value |

### *3.2 EndoProteoFASP*

#### *3.2.1 Filter Equilibration and Pre-conditioning*

1. Add 200 µL of ABC solution to the filter.
2. Centrifuge at 14,000 × *g* for 10 min at room temperature (RT).
3. Discard filtrate (*see* **Note 4**).
4. Repeat steps 1–3 once.

#### *3.2.2 Collection of the Natural Peptidome*

1. Load filter with a sample volume corresponding to a particular amount of protein (e.g., 150 µg) (*see* **Note 5**).
2. Add ABC solution to dilute the sample up to 500 µL.

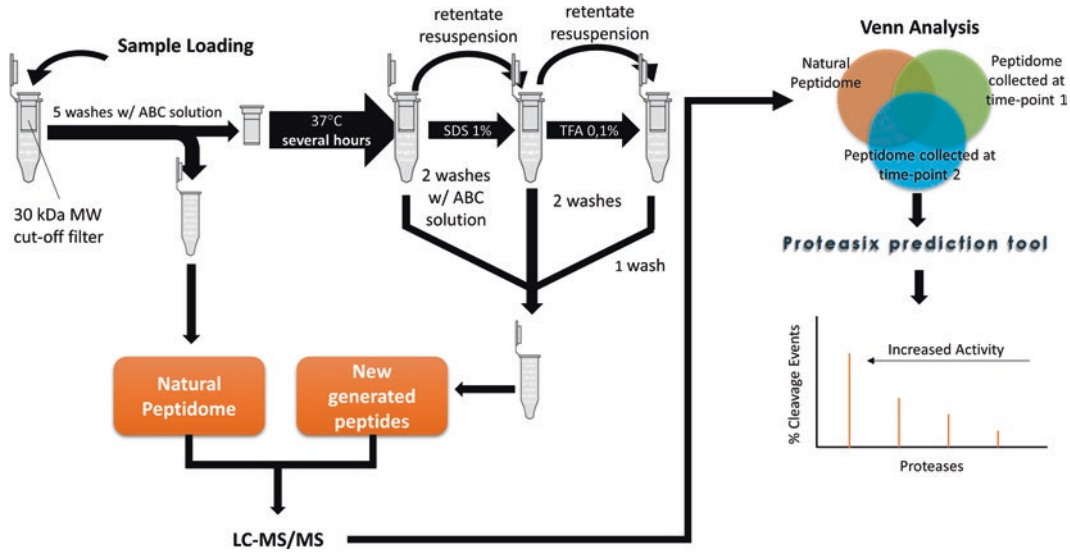



**Fig. 1** EndoProteoFASP + Proteasix analysis workflow. After sample loading, naturally occurring peptides are eluted from a 30 kDa cut-off filter. Then, samples are allowed to incubate for several hours or days at 37 °C to yield more peptides. The new peptidome is collected by sequential washings with ammonium bicarbonate (ABC) solution, sodium dodecyl sulfate (SDS), and trifluoroacetic acid (TFA). Peptides are then separated and identified by LC-MS/MS. Peptides specific to conditions determined a priori (e.g., time-points of collection, disease subtypes, different pathophysiological conditions) are selected by Venn analysis and the list of peptides (along with the information of starting and ending amino acids) is processed by Proteasix software to retrieve potentially active proteases, according to cleavage site specificity

3. Centrifuge at 14,000 × *g* for 10 min at RT to elute the naturally occurring peptides. Save the filtrate as well as the material retained in the filter.
4. Add 500 µL of ABC solution to the filter and repeat centrifugation (*see* **Note 6**). Save the filtrate (combine with filtrate from previous step).
5. Repeat step 4 three times more in order to wash away the majority of the peptides, saving the filtrate each time as well as the material retained in the filter.
6. Combine all 5 filtrates and evaporate to complete dryness using a centrifugal vacuum evaporator.

### 3.2.3 Saliva Autolysis

1. Add 500 µL of ABC solution to the filter to prevent sample drying (*see* **Note 7**).
2. Incubate samples for the desired period in a lab incubator at 37 °C (*see* **Notes 8** and **9**).

### 3.2.4 Collection of the Peptides Generated by Saliva Autolysis

1. Centrifuge at 14,000 × *g* for 10 min at RT to elute the newly formed peptides. Save the filtrate and the material retained in the filter.

2. Add 500 µL of ABC solution to the filter and repeat centrifugation (*see* **Note 6**). Save both filtrate and material in the filter.
3. Add 400 µL of SDS solution to the filter and gently resuspend retentate (*see* **Note 10**).
4. Centrifuge at 14,000 × *g* for 10 min at RT. Save both.
5. Add 400 µL of ABC solution to the filter and gently resuspend retentate.
6. Centrifuge at 14,000 × *g* for 10 min at RT. Save both.
7. Add 400 µL of TFA solution to the filter and gently resuspend retentate.
8. Perform a final centrifugation at 14,000 × *g* for 10 min at RT. Save the filtrate.
9. Combine all five filtrates and evaporate to complete dryness in a centrifugal vacuum evaporator.
10. Store peptide extracts at −80 °C until further analysis.
11. Sample retentate can be resuspended in 20 µL of ultrapure water and stored at −80 °C (*see* **Note 11**).

### 3.3 Peptide Separation and Identification by nanoHPLC-MALDI-TOF/TOF

#### 3.3.1 Sample Preparation

1. Resuspend peptide pellets (from **step 6** of Subheading 3.2.2 and from **step 9** of Subheading 3.2.4, separately) in 5 µL of ACN/water/TFA solution.
2. Sonicate for 3 min.
3. Add 25 µL of TFA solution.
4. Sonicate again for 3 min.
5. Use 10 µL of the peptide solution for downstream analysis and keep 20 µL (*see* **Note 12**).

#### 3.3.2 Peptide Separation by nanoHPLC

1. Program a user-defined injection program, for example with Chromeleon software (Dionex Corporation, Sunnyvale, CA) (*see* **Note 13**).
2. Prepare the Probot™ system using µCarrier 2.0 software and fill the syringe with MALDI matrix solution.
3. Prepare sample for injection by spinning down for 3 min in a benchtop mini-spin and safely transferring it into a sample vial in the autosampler (*see* **Note 14**).
4. Set as initial parameters:
   (a) Column oven temperature: 40 °C (*see* **Note 15**);
   (b) Loading pump flow: 300 nL/min;
   (c) MicroPump Mobile Phase B: 5%;
   (d) Fraction collection time: 20 s.

5. Set the chromatographic separation and collect the peptide fractions onto the MALDI plates, with the following gradient:
   (a) 10 min—5% HPLC mobile phase B
   (b) 50 min—60% HPLC mobile phase B
   (c) 55 min—100% HPLC mobile phase B
   (d) 65 min—100% HPLC mobile phase B

#### *3.3.3 MALDI Spectra Acquisition*

1. Follow the manufacturer instructions to prepare MALDI instrumentation.
2. Calibrate the mass spectrometer with Glu-Fib (monoisotopic *m/z* 1570.68) or similar peptide(s).
3. Acquire spectra in the positive ion reflector mode using 1000 laser shots in the 700–4100 Da *m/z* range.
4. Create a data-dependent acquisition method, selecting the 16 most intense peaks in each sample spot for subsequent tandem mass spectrometry (MS/MS) data acquisition (*see* **Note 16**).

#### *3.3.4 Protein Identification*

1. Select the desired spots, create and save the MS/MS generic data files (.mgf files).
2. Go to Mascot webpage (http://www.matrixscience.com/search_form_select.html) (*see* **Note 17**).
3. Perform search on "MS/MS Ion Search"
   (a) Select an appropriate database (e.g., SwissProt).
   (b) Choose the option "No Cleave" for enzyme (*see* **Note 18**).
   (c) Choose the option "*Homo sapiens*" in the taxonomy.
   (d) Set the precursor ion MS tolerance at 40 ppm and the fragment ions' tolerance (MS/MS tol.) at 0.4 Da.
   (e) Select "+1" as peptide charge.
   (f) Browse and select the MS/MS data file (.mgf file for Explorer software®).
   (g) Select the desired data format (e.g., Mascot generic).
   (h) Select "MALDI-TOF/TOF" as the instrument.
   (i) Check the "decoy" option.
   (j) Start the search.

### *3.4 Proteasix Analysis*

1. Prepare a spreadsheet from the list of identified peptides with four columns: peptide ID, protein accession number (SwissProt code), position of the start amino acid, and position of the stop amino acid (*see* **Notes 19** and **20**).
2. Go to Proteasix website (http://proteasix.cs.man.ac.uk/index.html) and select the "Prediction tool."

3. Copy and paste the tab-delimited peptide list to the proper field ("Input peptide list").
4. Select next step to "CS sequence" (*see* **Note 21**).
5. Once "The automatic reconstruction of *N*- and *C*-terminal Cleavage Sites is finished" select next step to "Observed proteases."
6. Once "The identification of observed Protease/Cleavage Site association is finished" select next step to "Predicted proteases" (*see* **Note 22**).
7. Once "The probability of Protease/Cleavage Site association has been calculated" select next step to "Detail results" (*see* **Note 23**).
8. Download results by selecting "Copy and paste detailed results" and then "Download detailed results."
9. Interpret results by comparing the peptidome profile with the set of predicted proteases, for each condition and/or time of collection (*see* **Notes 24**, **25**, and **26**).

Two case studies are presented below to explain the utility and the rationale behind the use of Proteasix in the study of the proteolytic events taking place in other biofluids beyond saliva and in different pathological settings.

Example 1
How to profile proteases potentially implicated in the generation of cerebrospinal fluid (CSF) peptidome?

To answer this question, we used peptide data reported by Hölttä et al. [13], which we processed according to the instructions provided in Subheading 3.4 (for graphical description see Fig. 2). In their study, CSF samples were collected and pooled from three neurological disorder-free patients. Peptide extracts were analyzed by offline LC-MALDI-TOF/TOF, resulting in 730 identified peptides from 104 different proteins [13]. From the list of identified peptides, a new spreadsheet can be created assembling a user-defined peptide ID, the respective protein AC/ID code, in addition to the start and end amino acid positions. This new matrix can now be copied and pasted to Proteasix prediction tool and, after its analysis, a further matrix is created depicting the peptide-protease associations. Given the fact that Proteasix software can only process as many as 350 peptides at a time, we had to perform more than one analysis. After compilation of all data, it was possible to pinpoint potential proteases generating CSF peptidome by counting the number of cleavage events for each particular protease (this can be rapidly done by using Excel's COUNTIF function) and attribute the respective percentage. It is useful to define a percentage threshold (in Fig. 2 we set 0.5% as threshold) from which the activity of a protease is considered. Not surprisingly, as you may see in Fig. 2,

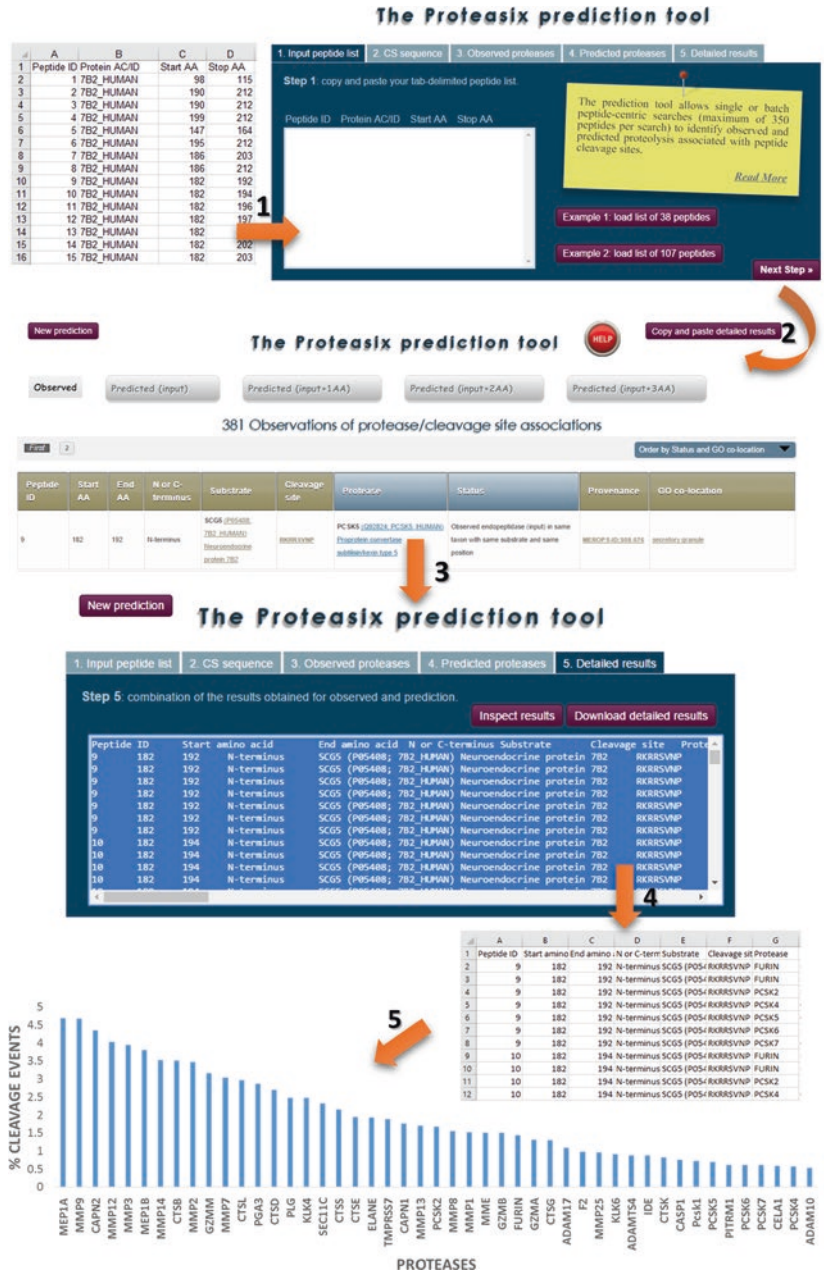



**Fig. 2** Main steps needed to perform Proteasix analysis. From the list of identified peptides, prepare a new spreadsheet with four columns (peptide ID, protein accession number/UniProt ID, position of the start amino acid, and position of the end amino acid). (1) Go to Proteasix website (http://proteasix.cs.man.ac.uk/) and copy/paste the input list in the proper field; (2) Run analysis from **step 1** to **step 5**, until detailed results are exhibited; (3) Select the option "Copy and paste detailed results"; (4) Select the whole output list and paste to a .txt file and then to an .xls file (using the special text import wizard—with the option tab-separated text); (5) Count the number of cleavage events with the Excel function COUNTIF, determine the percentage, and display it in a bar chart. From data observation, set an adequate percentage threshold to select predicted active proteases. In this case, bar chart represents the profile of predicted proteases to be active in cerebrospinal fluid, from the data collected by Hölttä et al. [13] using 0.5 as the percentage threshold

**Table 2**
**Contingency table depicting the frequency of amino acid residues in the flanking positions of the cleavage sites that generated endogenous cerebrospinal fluid peptides collected and identified by Hölttä et al. [13]**

| Amino acid residues | Amino acid frequency in cleavage site-flanking positions | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | P4 | P3 | P2 | P1 | P1' | P2' | P3' | P4' |
| A | 41 | 61 | 51 | 57 | 47 | 41 | 45 | 39 |
| C | 0 | 1 | 0 | 2 | 2 | 0 | 2 | 0 |
| D | 8 | 5 | 9 | 11 | 15 | 12 | 17 | 24 |
| E | 29 | 23 | 26 | 35 | 34 | 37 | 35 | 47 |
| F | 10 | 10 | 11 | 13 | 11 | 9 | 5 | 11 |
| G | 47 | 37 | 32 | 36 | 37 | 38 | 45 | 39 |
| H | 14 | 6 | 7 | 9 | 8 | 11 | 10 | 6 |
| I | 7 | 9 | 14 | 9 | 14 | 15 | 10 | 9 |
| K | 18 | 15 | 25 | 27 | 18 | 16 | 15 | 16 |
| L | 32 | 44 | 47 | 41 | 61 | 50 | 39 | 35 |
| M | 6 | 3 | 5 | 5 | 7 | 2 | 3 | 5 |
| N | 3 | 8 | 3 | 7 | 6 | 3 | 5 | 8 |
| P | 48 | 41 | 30 | 19 | 13 | 26 | 38 | 41 |
| Q | 19 | 22 | 20 | 16 | 21 | 21 | 20 | 25 |
| R | 29 | 22 | 28 | 56 | 21 | 26 | 29 | 16 |
| S | 37 | 38 | 38 | 27 | 40 | 34 | 33 | 31 |
| T | 15 | 9 | 16 | 8 | 11 | 16 | 10 | 8 |
| V | 34 | 43 | 33 | 19 | 33 | 36 | 33 | 33 |
| W | 1 | 1 | 1 | 0 | 0 | 3 | 2 | 3 |
| Y | 3 | 3 | 5 | 4 | 2 | 5 | 5 | 5 |

Most frequent amino acid residues (>30 times) in each position (P4 to P4') were highlighted in a gray shade

(matrix) metalloproteases are one group of proteases potentially implicated in the generation of CSF peptides, but calpain 2 (CAPN2) and some cathepsins (e.g., CTSB, CTSL, CTSD) also seem to participate in the proteolysis of CSF proteins.

The prediction of such proteases by Proteasix relies on the recognition of specific motifs in the inputted peptides. Those motifs are summarized in MEROPS' specificity matrix, in which a specific score is associated with each amino acid residue possible. Table 2 depicts the contingency board: the frequency of each amino acid in the P4-to-P4' positions of the cleavage site. If one takes a closer look at this table, we can understand the rationale behind protease prediction. For instance, the recognition motifs of MMP-9, the matrix metalloprotease predicted to be the most active, include Pro (P) in P3, Gly (G) in P1, Leu (L) in P1', and Gly (G) in P3', which are very frequent in the input peptide

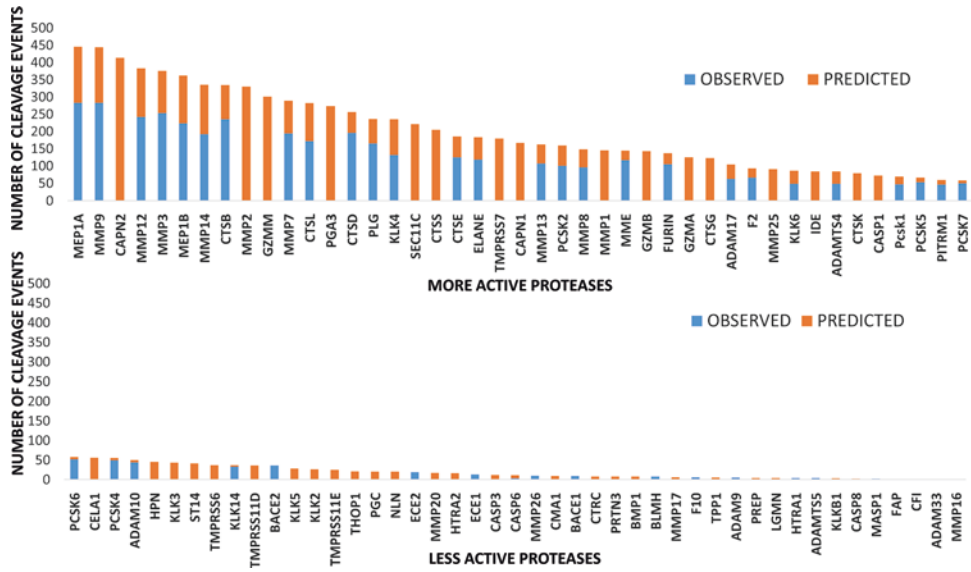



**Fig. 3** Distribution of the predicted active proteases in cerebrospinal fluid and the relative contribution of observed data (in blue) and prediction data (in orange) to less active (left) and more active (right) proteases

cleavage sites. Also, the recognition motifs of calpain 2, the third most active predicted protease, are characterized by the presence of Leu (L) in P2, either Ala (A) or Ser (S) in P1' and Pro (P) in P3' which were also very common in the cleavage sites. A final example is cathepsin predicted to be the most active, specifically cathepsin B. The cathepsin B recognition motifs include Leu (L) in P3, Ala (A) or Val (V) in P2, Gly (G) or Ala (A) in P1, and Gly (G) in P3' which are all well represented in the contingency table.

Despite the useful information one can elucidate from EndoProteoFASP + Proteasix analysis, caution should be taken with data interpretation. For instance, one should acknowledge that some highlighted proteases may arise only from predictions. For instance, the activity of calpain 2, matrix metalloprotease 2 (MMP2), and granzyme M (GZMM), which account for several cleavage events according to Proteasix's algorithm, in CSF, has only prediction-based support (see orange bars in Fig. 3). In these cases, the validation of these proteases' activities is even more important (*see* **Note 25**).

Example 2

How to identify proteases with potential value for ovarian cancer diagnosis?

In order to take a snapshot at the set of proteases involved in the generation of peptides in ascites fluid in women with ovarian cancer, we used the dataset reported by Bery et al. [14]. This study enrolled 3 subjects with ovarian cancer and 3 subjects with liver cirrhosis (benign ascites), from whom ascites fluid was collected and

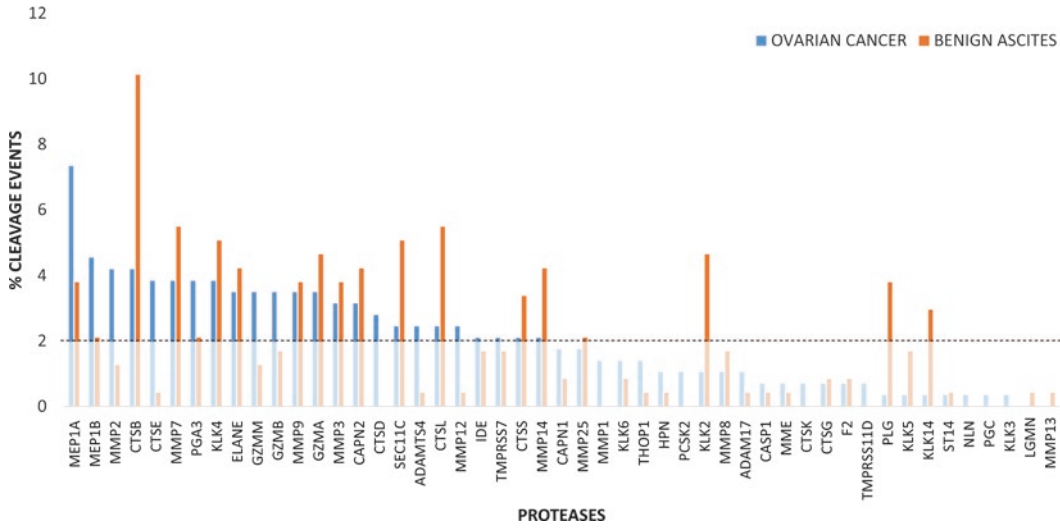



**Fig. 4** Comparison of the predicted protease profile in ascites fluid from patients with ovarian cancer (blue bars) and from controls (orange bars—benign ascites derived from liver cirrhosis). The distribution of proteases is based on the percentage of cleavage events. A threshold has been set at 2%, so that only proteases associated with higher percentages of cleavage events could be considered

peptides identified by nanoLC-MS/MS (LTQ-Orbitrap XL). Overall, more than 2000 peptides could be identified, originating from 259 different proteins. Despite the fact that we did not have a large enough number of individuals to draw conclusions with regard to the diagnostic value of highlighted proteases, we used this dataset to demonstrate how performing Proteasix analysis can virtually help in such a task. In order to do that, two analyses were performed with Jvenn, an online Venn diagram tool (http://bioinfo.genotoul.fr/jvenn/example.html), to, first, retrieve peptides that are unique in benign ascites (1054 peptides) and in ovarian cancer (777 peptides) and, second, to identify those that are common to all subjects with benign ascites (29 peptides) and those common to all patients with cancer (35 peptides). Thus, by subjecting these last two subsets of peptides (29 peptides for benign ascites and 35 peptides for ovarian cancer) to Proteasix analysis we could predict the most active proteases in both conditions, assuring that the attained profiles were not a reflex of one individual. We followed the same analysis protocol as described in the previous example but, in this case, we set 2% as the threshold to consider for the activity of a given protease. As shown in Fig. 4 and also in Fig. 5, it is possible to detect some proteases with remarkable predicted over-activity in ovarian cancer and in liver cirrhosis (benign ascites). For instance, meprin 1A subunit alpha (MEP1A) and beta (MEP1B), matrix metalloprotease 2 (MMP2), and cathepsin E (CTSE) seem to be overactive in ascites fluid of women with ovarian cancer, while in the opposite cathepsin B (CTSB), L (CTSL), signal peptidase

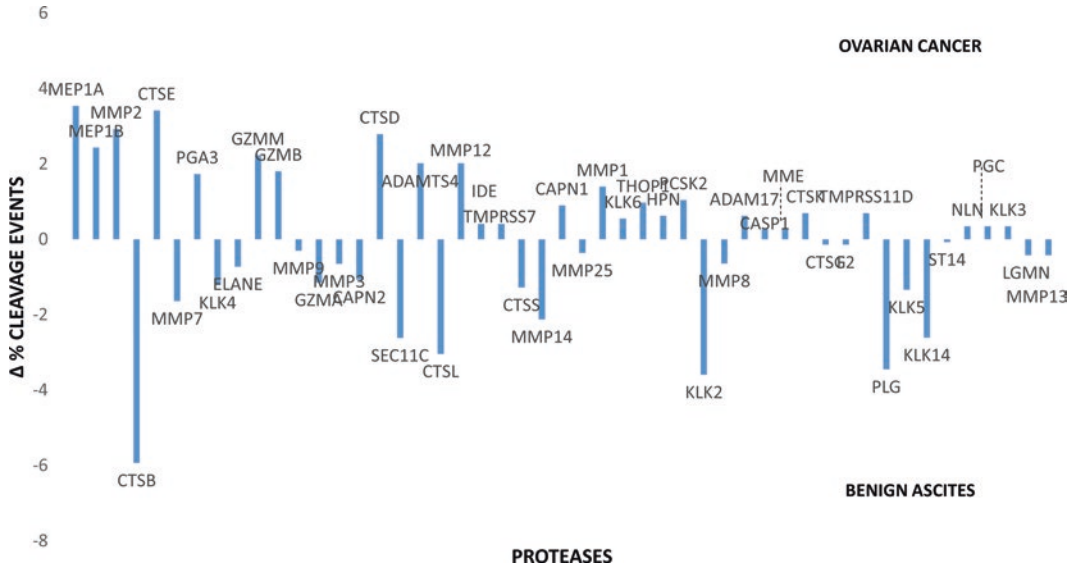



**Fig. 5** Bar chart illustrative of the difference of cleavage events attributed to each protease in ovarian cancer (upwards) and in benign ascites (downwards), showing the set of proteases more active in both conditions

complex catalytic subunit SEC11C, kallikrein 2 (KLK2), and plasminogen (PLG) seem to be overactive in benign ascites. Thus, the attained protease profile can be of diagnosis relevance, but again, validation is always warranted (*see* **Note 25**).

Hence, the endoProteoFASP approach, together with Proteasix, can be useful to take a first look at the profile of active proteases in several biological samples, across different pathological settings. Additionally, by comparing the peptidome and protease profile in a time course, one can understand in more detail the proteolytic events taking place in a biological medium. Final validation can be used with the same platform by adding specific protease inhibitors or an excess of synthetic protease substrates to the filter unit or, alternatively, one can perform zymography studies with sample retentate and identify proteases by zymography-LC-MS/MS strategies.

## 4 Notes

1. Use fresh samples, if possible, and avoid multiple freezing/thawing cycles to prevent protein precipitation and the decline of protease activity.
2. Do not aspirate the saliva supernatant close to the pellet. The pellet contains food remnants, cell debris, bacteria, and other insoluble materials, and these can contaminate sample. It is preferable to waste part of the sample, as saliva can be noninvasively and safely collected in considerably large amounts (>10 mL).

3. It is unnecessary to use the complete RC (reducing agent compatible)-DC (detergent compatible) kit, because samples have not been treated with any reducing agent.
4. If the filter retains the equilibration solution, it should be discarded. Membrane filters might be defective from the manufacturing process or the product might be expired. In this case, use filters from a new product batch.
5. The amount of sample to load will depend on the number of peptide elutions one wants to perform as well as the time of sample incubation. Accordingly, the higher the time of incubation and the greater the number of elutions, the higher will be the amount of protein to be loaded. The starting protein mass may require optimization.
6. After each centrifugation cycle check for potential filter blockade. Whenever this is the case, gently suspend sample retentate avoiding filter damage and/or perforation and perform an extra centrifugation cycle.
7. Pay attention to the tubes' lids. Tubes should be properly sealed; otherwise there is the risk of water evaporation and sample drying. Loss of proteolytic activity can occur under these circumstances. Furthermore, proper tube sealing can minimize the risk of filter drying and blockade.
8. We have designed this protocol to study the activity of salivary proteases under physiological conditions (37 °C). Nevertheless, the same method can be applied for biotechnological purposes, such as to test the activity of salivary proteases at lower or higher temperatures. When that is the case, acknowledge the risk of protein aggregation (at lower temperatures) or denaturation (at higher temperatures).
9. We have successfully collected new peptides after 18 hours and after 115 h [9, 10], but longer incubation periods can be tested.
10. After adding SDS, the volume of the remaining solutions to resuspend the retentate should not exceed 400 μL owing to foam formation, the spilling risk, and consequent sample contamination.
11. Sample retentates, which are enriched with proteins larger than 30 kDa, can be subjected to downstream SDS-PAGE or zymography analysis.
12. Samples can be dried again and kept at −80 °C for repetition of analysis, if deemed necessary.
13. For details on how to use this program the reader is referred to Sect. 3.5.1 of Cova et al. [15].
14. Avoid the formation of bubbles, as these will compromise regular chromatographic separation.

15. The column oven temperature can be changed in order to optimize elution of peptides. Generally, increasing the temperature will reduce peptide retention and prevent peptide co-elution.
16. Collision energy should be of 2 kV using air as collision gas (pressure should be $2 \times 10^{-7}$ Torr).
17. Protein identification can also be performed with alternative algorithms such as SCIEX's ProteinPilot™ (http://sciex.com/products/software/proteinpilot-software).
18. Recall that we are exploring the salivary endogenous proteases. This way, we have not added any exogenous protease. Still, if the protein composition of the retentate is to be analyzed, then trypsin could be used.
19. The peptide identification can be a number, a name, or any identification system the user is comfortable with.
20. Only peptides from proteins identified with statistical significance (p-value <0.05) should be selected.
21. Proteasix makes use of the SwissProt knowledgebase to gather the amino acid sequence for each peptide listed.
22. Proteasix retrieves information from MEROPS (the Peptidase Database, release 10.0, http://merops.sanger.ac.uk/) in order to predict the likelihood of a given protease to generate the input list of peptides. This database presents, for each protease, its substrate specificity using the P4, P3, P2, P1, P1', P2', P3', and P4' nomenclature system. Furthermore, this database can be useful to explore the type of proteases potentially generating the peptides analyzed by Proteasix, as it displays specific information regarding their physiological role, pharmaceutical relevance as well as their participation in KEGG pathways.
23. The Proteasix output displays, for each peptide, detailed information of the predicted protease, namely the type of cleavage (*N*- or *C*-termini), the predicted cell location (based on gene ontology), the source of information (e.g., MEROPS or CutDB, BRENDA, PubMED), and additional information concerning the existence of additional exopeptidases generating the peptide.
24. From the list of predicted proteases, it is useful to analyze the percentage of cleavages associated with each protease and define a threshold to define a set of proteases potentially active in each experimental condition (before and after autolysis, in health and disease, before and after addition of an inhibitor, among others).
25. Validation of specific proteases can be ensued by adding specific protease inhibitors or synthetic substrates to the filter unit in order to see the changes in the protease profile and in the

peptidome profile, respectively. Furthermore, it is possible to perform zymography with sample retentate (high molecular weight material) in order to directly assess protease activity. Besides it is possible to cut the bands with higher optical density, perform tryptic digestion, and identify proteases by LC-MS/MS. Final validation can be done by using antibody-based methods, namely western blot.

26. Additional information about proteases, particularly related to functional domains, can be found in Pfam website (http://pfam.xfam.org/), a database that collects a large number of protein families, with detailed structural and functional information. Pfam groups proteins into families and clans, based on sequence similarities, related structure and function as well as profile-profile comparisons. Thus, it is possible to study the evolutionary relationships between proteases using this database [16, 17].

## Acknowledgments

This work was supported by the Portuguese Foundation for Science and Technology (FCT) through UnIC, iBiMED, QOPNA research units (UID/IC/00051/2013, UID/BIM/04501/2013, PEst-C/QUI/UI0062/2013) and by project DOCnet (NORTE-01-0145-FEDER-000003), supported by Norte Portugal Regional Operational Programme (NORTE 2020), under the PORTUGAL 2020 Partnership Agreement, through the European Regional Development Fund (ERDF). Rui Vitorino and Fábio Trindade are supported by individual fellowship grants (IF/00286/2015 and SFRH/BD/111633/2015, respectively).

### References

1. Cabras T, Iavarone F, Manconi B, Olianas A, Sanna MT, Castagnola M, Messana I (2014) Top-down analytical platforms for the characterization of the human salivary proteome. Bioanalysis 6:563–581. https://doi.org/10.4155/bio.13.349
2. Amado F, Lobo MJC, Domingues P, Duarte JA, Vitorino R (2010) Salivary peptidomics. Expert Rev Proteomics 7:709–721. https://doi.org/10.1586/epr.10.48
3. Thomadaki K, Helmerhorst EJ, Tian N, Sun X, Siqueira WL, Walt DR, Oppenheim FG (2011) Whole-saliva proteolysis and its impact on salivary diagnostics. J Dent Res 90:1325–1330
4. Vitorino R, Barros A, Caseiro A, Domingues P, Duarte J, Amado F (2009) Towards defining the whole salivary peptidome. Proteomics Clin Appl 3:528–540. https://doi.org/10.1002/prca.200800183
5. Lucchi G, Chambon C, Truntzer C, Pecqueur D, Ducoroy P, Schwartz C, Nicklaus S, Morzel M (2009) Mass-spectrometry based characterisation of infant whole saliva peptidome. Int J Pept Res Ther 15:177–185. https://doi.org/10.1007/s10989-009-9167-2
6. Sun X, Salih E, Oppenheim FG, Helmerhorst EJ (2009) Activity-based mass spectrometric characterization of proteases and inhibitors in human saliva. Proteomics Clin Appl 3:810–820. https://doi.org/10.1002/prca.200800242
7. Wiśniewski JR, Zougman A, Nagaraj N, Mann M (2009) Universal sample preparation

method for proteome analysis. Nat Methods 6:359–362

8. Manza LL, Stamer SL, Ham A-JL, Codreanu SG, Liebler DC (2005) Sample preparation and digestion for proteomic analyses using spin filters. Proteomics 5:1742–1745. https://doi.org/10.1002/pmic.200401063
9. Trindade F, Amado F, Gomes PS, Vitorino R (2015) endoProteoFASP: a novel FASP approach to profile salivary peptidome and disclose salivary proteases. Talanta 132:486–493. https://doi.org/10.1016/j.talanta.2014.09.026
10. Trindade F, Amado F, Oliveira-Silva RP, Daniel-da-Silva AL, Ferreira R, Klein J, Faria-Almeida R, Gomes PS, Vitorino R (2015) Toward the definition of a peptidome signature and protease profile in chronic periodontitis. Proteomics Clin Appl 9:917–927. https://doi.org/10.1002/prca.201400191
11. Klein J, Eales J, Zürbig P, Vlahou A, Mischak H, Stevens R (2013) Proteasix: a tool for automated and large-scale prediction of proteases involved in naturally occurring peptide generation. Proteomics 13:1077–1082. https://doi.org/10.1002/pmic.201200493
12. Arguello Casteleiro M, Klein J, Stevens R (2016) The proteasix ontology. J Biomed Semantics 7:1–7. https://doi.org/10.1186/s13326-016-0078-9
13. Hölttä M, Zetterberg H, Mirgorodskaya E, Mattsson N, Blennow K, Gobom J (2012) Peptidome analysis of cerebrospinal fluid by LC-MALDI MS. PLoS One 7:e42555. https://doi.org/10.1371/journal.pone.0042555
14. Bery A, Leung F, Smith CR, Diamandis EP, Kulasingam V (2014) Deciphering the ovarian cancer ascites fluid peptidome. Clin Proteomics 11:1–9. https://doi.org/10.1186/1559-0275-11-13
15. Cova M, Oliveira-Silva R, Ferreira JA, Ferreira R, Amado F, Daniel-da-Silva AL, Vitorino R (2015) Glycoprotein enrichment method using a selective magnetic nano-probe platform (mnp) functionalized with lectins BT - clinical proteomics: methods and protocols. In: Vlahou A, Makridakis M (eds) . Springer, New York, pp 83–100
16. Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, Potter SC, Punta M, Qureshi M, Sangrador-Vegas A, Salazar GA, Tate J, Bateman A (2016) The Pfam protein families database: towards a more sustainable future. Nucleic Acids Res 44:D279–D285. https://doi.org/10.1093/nar/gkv1344
17. Finn RD, Mistry J, Schuster-Böckler B, Griffiths-Jones S, Hollich V, Lassmann T, Moxon S, Marshall M, Khanna A, Durbin R, Eddy SR, Sonnhammer ELL, Bateman A (2006) Pfam: clans, web tools and services. Nucleic Acids Res 34:D247–D251. https://doi.org/10.1093/nar/gkj149

# Chapter 20

# Methodology for Urine Peptidome Analysis Based on Nano-HPLC Coupled to Fourier Transform Ion Cyclotron Resonance Mass Spectrometry

**Alexey S. Kononikhin, Victoria A. Sergeeva, Anna E. Bugrova, Maria I. Indeykina, Natalia L. Starodubtseva, Vitaliy V. Chagovets, Igor A. Popov, Vladimir E. Frankevich, Patrick Pedrioli, Gennady T. Sukhikh, and Eugene N. Nikolaev**

## Abstract

Urine is a sample of choice for noninvasive biomarkers search because it is easily available in large amounts and its molecular composition provides information on processes in the organism. The high potential of urine peptidomics has been demonstrated for clinical purpose. Several mass spectrometry based approaches have been successfully applied for urine peptidome analysis and potential biomarkers search. Summarizing literature data and our own experience we developed a protocol for comprehensive urine peptidome analysis. The technology includes several stages and consists of urine sample preparation by size exclusion chromatography and identification of featured peptides by nano-HPLC coupled to Fourier transform ion cyclotron resonance mass spectrometry, semiquantitative and statistical data analysis.

**Key words** Urine peptidome, Size exclusion chromatography, FTICR MS, LC-MS/MS

## 1 Introduction

The main task for peptidomics is characterization of native peptides in a biological sample [1, 2]. Many peptides have specific functions as hormones, neurotransmitters, cytokines, or growth factors [3]. Some peptides being the products of enzymatic degradation proteins in the body often serve as indicators of normal or pathological processes that can be used for the detection of new markers of early stages of the disease or mediators of pathological processes. Peptidomics uses the full arsenal of proteomic methods and approaches to address methodological challenges and in fact is part of modern proteomics. Similar to top-down proteomics, no trypsin or other proteases are applied for peptidomic studies to conserve

Michael Schrader and Lloyd Fricker (eds.), *Peptidomics: Methods and Strategies*, Methods in Molecular Biology, vol. 1719, https://doi.org/10.1007/978-1-4939-7537-2_20, 

the endogenous information on the peptides from a biological sample, including posttranslational modifications (PTMs) and proteolytic products revealing the natural proteases participating in the proteolytic processes [4]. Endogenous peptides may vary significantly in size (from 2 to 100 amino acids) and properties and complementary methodologies of preparation/separation, MS analysis and data processing are recommended to analyze the entire range of peptides [5, 6]. Each step is equally important and can have a different implementation based on capabilities, tasks, volume, and characteristics of the biomaterial.

Analysis of endogenous peptides in body fluids is successfully used for biomarker discovery [7–11]. Proteomic/peptidomics biomarker discovery relies on the identification of features (proteins and peptides) that are significantly differentially distributed in a specific "disease" group compared to a matched "control" group, and are therefore characteristic of a pathological condition. Urine is a desirable material for diseases study and diagnostics. The high potential of a urine peptidomic approach was demonstrated by several research groups who suggested potential biomarkers for diagnosis and prognosis [12–14]. Several methodological platforms were created for urine peptidome analysis. I.A. Buhimschi and coauthors developed an approach for high throughput raw urine peptidome analysis based on surface-enhanced laser desorption/ionization mass spectrometry (SELDI-MS) with H4 and H50 arrays without any prior urine treatment. Tandem mass spectrometry followed by de novo sequencing of potential biomarker peptides has been reported [12]. D.M. Carty and coauthors performed SEC extraction and urine peptidome analysis by capillary electrophoresis online coupled to micro-time-of-flight mass spectrometry (CE-MS) and support vector machine-based software with further LC-MS/MS identification of featured peptides. Later, the same authors demonstrated prevalence of LC-MS/MS compared to CE-MS for maximum coverage of peptides in urine [15].

Here, we describe the methodology which includes the most promising features from previous studies and provides the optimal urine peptidome coverage with reasonable analytical efforts and robustness. The protocol includes several stages and consists of urine sampling, sample preparation, SEC extraction and nano-HPLC coupled to Fourier transform ion cyclotron resonance mass spectrometry (FTMS), identification of peptides, semiquantitative and statistical analysis.

## 2 Materials

### 2.1 Urine Collection

Urine sampling is based on a standard protocol for proteomic analysis established by Human kidney and Urine Proteomics Project (HKUPP http://www.hkupp.org/) and European

Kidney and Urine Proteomics (EuroKUP http://www.eurokup.org/):

1. The second morning urine (midstream) is collected (*see* **Note 1**).
2. The sample is centrifuged for 10 min at 2000 × *g*, 4 °C (*see* **Note 2**) within 20 min of collection.
3. The supernatant is stored at −80 °C until the analysis.

### 2.2 Sample Preparation

1. Polypropylene tubes of different volume (*see* **Note 3**).
2. Concentrator spin columns with molecular weight cutoff of 10 kDa, which can accommodate 4 mL sample volume.
3. Sample solution: 4 M urea, 20 mM $NH_4OH$, 0.2% SDS, deionized water ($dH_2O$) (*see* **Notes 4** and **5**).
4. Size-exclusion columns (we use Sephadex (G-25), *see* **Note 6**).
5. Equilibration solution: 0.01% $NH_4OH$, $dH_2O$.
6. The bicinchoninic acid assay (BCA assay).
7. Lyophilizer/freeze dryer (we use Alpha 2-4 LSC, Martin Christ).

### 2.3 Liquid Chromatography/Mass Spectrometry

1. Nano-HPLC system (we use Agilent 1100).
2. Self-packed reversed-phase C18 nano-column (o.d. 360 μm, i.d. 75 μm, tip diameter 15 μm, length 12 cm, we use PicoTip emitters, packed with Reprosil-Pur Basic C18, 3 μm, 100 Å pores, *see* **Note 7**).
3. Solution A: 0.1% formic acid in water v/v.
4. Solution B: 0.1% formic acid in acetonitrile v/v.
5. 2 mL glass autosampler vial with fused 350 μL insert.
6. Mass spectrometry software (we use Xcalibur™ software, Thermo Fisher Scientific Inc., USA).
7. Mass spectrometer equipped with adapter for nanospray (we use 7-Tesla LTQ-FT Ultra mass spectrometer and Ion Max ion source with self-made adapter).

### 2.4 Data Processing

1. Up to date version of UniProtKB database.
2. Quantification software package for proteomics data (we use MaxQuant, Max Planck Institute of Biochemistry, Computational Systems Biochemistry, Germany, version 1.5.3.30).
3. Software package for shotgun proteomics data analysis (we use Perseus software, Max Planck Institute of Biochemistry, Computational Systems Biochemistry, Germany, version 1.5.3.2).
4. Software for univariate and further statistical analyses (we use ropls package [16]).

## 3 Methods

### 3.1 Extraction of Peptides by Ultrafiltration and Purification by Size-Exclusion Chromatography

1. Mix 1.5 mL of urine and 1.5 mL of sample solution (*see* **Note 8**).
2. Rinse the spin concentrator columns with 4 mL of water and centrifuge at 4000 × *g* for 5 min. Discard the water (*see* **Note 9**).
3. Put 3 mL of mixture on the concentrator columns. Centrifuge at 4000 × *g* for 20 min (*see* **Note 10**). Collect the filtrate that has passed through the filter.
4. Equilibrate size exclusion columns with 25 mL (at least) of Equilibration solution (*see* **Note 11**).
5. Load 2.5 mL of filtrate onto the column. Let the solution soak in.
6. Replace waste tank with 2 mL tubes per each column. Elute with 2 mL of Equilibration solution.
7. Seal open tubes with parafilm, freeze for 2 h at a temperature of −20 °C, and make holes in the film with a clean needle.
8. Lyophilize.
9. Dissolve in 100 μL of water.
10. Quality of sample preparation is controlled by estimation of peptides concentration in 10–20 μL of sample from **step 9** using BCA assay (*see* **Note 12**).

### 3.2 Liquid Chromatography/Mass Spectrometry

1. Dilute 10–20 μL sample from **step 9** of Subheading 3.1 with solution A to a final concentration of 1 μg/μL according to concentration assessment obtained on **step 10** of Subheading 3.1 (*see* **Note 12**).
2. Place the sample volume (10–20 μL), sufficient for a minimum of three replicate injections, in a glass autosampler vial, close cap tightly and label the vial (*see* **Note 13**).
3. Place vial in a cooled (4 °C) autosampler rack (*see* **Note 14**).
4. Place vial with water-acetonitrile mixture (50/50, v/v) as a blank sample (*see* **Note 15**).
5. Chromatographic conditions for a nano-HPLC:
   (a) 75 min analysis time at flow rate of 0.3 μL/min.
   (b) 0–15 min: 3% solution B, 97% solution A (loading of the sample on the column).
   (c) 15–45 min: linear gradient 3–50% of solution B.
   (d) 45–50 min: linear gradient 50–90% of solution B.
   (e) 50–60 min: 90% of solution B.
   (f) 60–65 min: linear gradient 90–3% of solution B.
   (g) 65–75 min: reequilibration of the column in 3% solution B.

6. MS/MS analysis is performed in data-dependent mode with the following parameters:
   (a) The precursor ion MS spectra (m/z 300–1600) are acquired in the ICR trap with resolution of $R$ = 50,000 at m/z 400.
   (b) Five most intense ions are isolated and fragmented in the LTQ, dynamic exclusion is used with 30 s exclusion duration.

### 3.3 Data Processing

Data analysis is performed by MaxQuant (*see* **Note 16**) using the following parameters:

#### 3.3.1 Peptides Identification. Semiquantitative Label-Free Analysis

1. Identification parameters are
   (a) Unspecific degradation or no-enzyme options.
   (b) Mass accuracy for the precursor ion is 5 ppm.
   (c) Mass accuracy for MS/MS fragments is 0.50 Da.
   (d) Possible variable modifications are oxidation of methionine, lysine, and proline residues (*see* **Note 17**).
2. Peptides with minimum of five amino acids are considered.
3. Cutoff value for false discovery rates of 0.01 (1% FDR).
4. At least two peptide identifications per protein.
5. At least one peptide must be unique for the protein group.
6. Second peptides and dependent peptides options are turned on.
7. Label free quantitation is turned on (default parameters give a good starting result—*see* **Note 18**).
8. Match between the runs (*see* **Note 19**).
9. MS/MS is required for LFQ comparison.

#### 3.3.2 Statistical Data Analysis

1. In case of pairwise analysis for few peptides, statistical data analysis must be performed using Mann-Whitney test with Bonferroni correction [17, 18].
2. In case of pairwise comparison of several hundred peptides, linear correlations between dependent variables and predictor variables must be calculated using partial/orthogonal partial least squares discriminant analysis (PLS-DA/OPLS-DA) methods (*see* **Note 20**) by the ropls package in R [16].
3. Statistical differences between more than two groups must be searched using the Kruskal–Wallis test [19]. Significantly different peptides ($p$-value <0.05) can be used for the Gene Ontology (GO) term enrichment (*see* **Note 21**).

## 4 Notes

1. Pregnant women are treated with a catheter. The catheter gives pure samples of urine. Standard collected urine can be investigated by this protocol as well.

2. This stage is commonly used to get rid of cellular fragments and big particles which can occur because of inflammation associated with pathological condition.
3. Use 5 mL tubes to mix urine with sample buffer and 2 mL tubes for final peptide elution. Make sure that total volume of your tubes is a little bit bigger than 2 mL (like those from Eppendorf).
4. Do not add SDS before urea is completely dissolved. Shake the solution to accelerate the process. After adding SDS beware of bubbling.
5. The solution must be freshly prepared every time because it develops a significant concentration of reactive cyanate ions on standing.
6. One can fill columns with Sephadex G-25 Medium (particle size 85–260 μm) or use commercial prefilled column. We purchase PD-10 Desalting Columns from GE Healthcare.
7. Commercially available C18 nano-columns can be used.
8. You should not shake up the mixture because the SDS will form bubbles. Pipette it gently and leave while rinsing the concentrator columns.
9. This stage is recommended by Sartorius to remove traces of glycerine and sodium azide.
10. It may require more time for some samples. Centrifuge until you get at least 2.5 mL of filtrate.
11. This stage takes about 10 min, so you can start equilibration before the centrifugation is completed.
12. Assessment of peptides' concentration by BCA gives approximate value because there are a lot of short fragments (for example, poly-proline chains) undetectable by standard assays.
13. Final sample volume in a vial must be 1–2 μL per injection plus an extra 10–15 μL to prevent air injection into the LC system. Each sample should undergo at least 3–4 LC MS/MS runs to obtain best results, especially if label free quantitation analysis is to be carried out.
14. If a thermostatic rack is unavailable, samples should be loaded to the autosampler queue in small groups, so that no sample is required to stay at room temperature for more than 24 h. Long periods at room temperature may lead to sample degradation.
15. Blank injection of a water-acetonitrile mixture (50/50, v/v) must be run between samples to eliminate sample carry over.
16. MaxQuant is a freely available software program that is used in the first step of the data processing. It allows a lot of active data

analysis, but requires high capacity data storage and consumes much computer time and power. The output is a txt-file which can be easily uploaded to any program and used for consecutive bioinformatic analysis. Estimate computer space: 1 GB of LC data will be converted into 10 GB of intermediate files and a 1 MB txt-file of results. If one is finished with data analysis and does not plan on rerunning it, delete everything except the txt-folder.

17. An important factor in peptide identification is to specify variable PTMs. The number of PTMs is considerably higher for endogenous peptides than for tryptic peptides of urine proteins. PEAKS de novo sequencing method (Bioinformatics Solutions, Waterloo, Canada) can be used in order to identify all possible modifications.
18. Be very careful organizing your data as fractions, groups, and experiments, since MaxQuant will match all fractions irrespective of which experiment they come from, but only with the same or adjacent fraction numbers.
19. MaxQuant can transfer peptide identifications from an LC-MS run in which the peptide has been identified by MS/MS to another LC-MS run, where either no MS/MS spectrum has been acquired for that MS peptide feature or the peptide could not be identified.
20. PLS is a supervised method. It finds the linear correlations between dependent variables and predictor variables.
21. Gene Ontology (GO) term enrichment can be done by Perseus program which performs bioinformatic analyses of the output of MaxQuant, and thus completes the proteomics analysis pipeline.

## Acknowledgments

The study was supported by RFBR grants no. 16-54-21011_SNF_a, 17-08-01537 A and SNF grant no. SNF IZLRZ3_163911.

### References

1. Filip S, Pontillo C, Peter Schanstra J et al (2014) Urinary proteomics and molecular determinants of chronic kidney disease: possible link to proteases. Expert Rev Proteomics 11:535–548. https://doi.org/10.1586/14789450.2014.926224
2. Schrader M, Schulz-Knappe P (2001) Peptidomics technologies for human body fluids. Trends Biotechnol 19:S55–S60. https://doi.org/10.1016/S0167-7799(01)01800-5
3. Ivanov VT, Yatskin ON (2005) Peptidomics: a logical sequel to proteomics. Expert Rev Proteomics 2:463–473. https://doi.org/10.1586/14789450.2.4.463
4. Schrader M, Schulz-Knappe P, Fricker LD (2014) Historical perspective of peptidomics. EuPA Open Proteom 3:171–182. https://doi.org/10.1016/j.euprot.2014.02.014
5. Machtejevas E, Marko-Varga G, Lindberg C et al (2009) Profiling of endogenous peptides by mul-

tidimensional liquid chromatography: on-line automated sample cleanup for biomarker discovery in human urine. J Sep Sci 32:2223–2232. https://doi.org/10.1002/jssc.200900058

6. Yang X, Hu L, Ye M et al (2014) Analysis of the human urine endogenous peptides by nanoparticle extraction and mass spectrometry identification. Anal Chim Acta 829:40–47. https://doi.org/10.1016/j.aca.2014.04.040
7. Broker ME, Lalmahomed ZS, Roest HP et al (2013) Collagen peptides in urine: a new promising biomarker for the detection of colorectal liver metastases. PLoS One 8:e70918. https://doi.org/10.1371/journal.pone.0070918
8. Cutillas PR, Norden AG, Cramer R et al (2003) Detection and analysis of urinary peptides by on-line liquid chromatography and mass spectrometry: application to patients with renal Fanconi syndrome. Clin Sci 104:483–490. https://doi.org/10.1042/CS20020342
9. Ling XB, Lau K, Deshpande C et al (2010) Urine peptidomic and targeted plasma protein analyses in the diagnosis and monitoring of systemic juvenile idiopathic arthritis. Clin Proteomics 6:175–193. https://doi.org/10.1007/s12014-010-9058-8
10. Ling XB, Lau K, Kanegaye JT et al (2011) A diagnostic algorithm combining clinical and molecular data distinguishes Kawasaki disease from other febrile illnesses. BMC Med 9:130. https://doi.org/10.1186/1741-7015-9-130
11. Nkuipou-Kenfack E, Zurbig P, Mischak H (2016) The long path towards implementation of clinical proteomics: exemplified based on CKD273. Proteomics Clin Appl 11:1600104. https://doi.org/10.1002/prca.201600104
12. Buhimschi IA, Zhao G, Funai EF et al (2008) Proteomic profiling of urine identifies specific fragments of SERPINA1 and albumin as biomarkers of preeclampsia. Am J Obstet Gynecol 199:551.e1–551.16. https://doi.org/10.1016/j.ajog.2008.07.006
13. Carty DM, Siwy J, Brennand JE et al (2011) Urinary proteomics for prediction of preeclampsia. Hypertension 57:561–569. https://doi.org/10.1161/HYPERTENSIONAHA.110.164285
14. Kononikhin AS, Starodubtseva NL, Bugrova AE et al (2016) An untargeted approach for the analysis of the urine peptidome of women with preeclampsia. J Proteome 149:38–43. https://doi.org/10.1016/j.jprot.2016.04.024
15. Klein J, Papadopoulos T, Mischak H et al (2014) Comparison of CE-MS/MS and LC-MS/MS sequencing demonstrates significant complementarity in natural peptide identification in human urine. Electrophoresis 35:1060–1064. https://doi.org/10.1002/elps.201300327
16. Thevenot EA, Roux A, Xu Y et al (2015) Analysis of the human adult urinary metabolome variations with age, body mass index, and gender by implementing a comprehensive workflow for univariate and OPLS statistical analyses. J Proteome Res 14:3322–3335. https://doi.org/10.1021/acs.jproteome.5b00354
17. Dunn OJ (1961) Multiple comparisons among means. J Am Stat Assoc 56:52–64. https://doi.org/10.1080/01621459.1961.10482090
18. Mann HB, Whitney DR (1947) On a test of whether one of two random variables is stochastically larger than the other. Ann Math Stat 18:50–60. https://doi.org/10.1214/aoms/1177730491
19. Kruskal WH, Wallis WA (1952) Use of ranks in one-criterion variance analysis. J Am Stat Assoc 47:583–621. https://doi.org/10.2307/2280779

# Chapter 21

# Identification of Components in Frog Skin Secretions with Therapeutic Potential as Antidiabetic Agents

J. Michael Conlon, R. Charlotte Moffett, Jérôme Leprince, and Peter R. Flatt

## Abstract

Several peptides that were first identified on the basis of their antimicrobial or immunomodulatory properties have subsequently shown potential for development into agents for the treatment of patients with Type 2 diabetes. A strategy is presented for the isolation and characterization of such peptides in norepinephrine-stimulated skin secretions from a range of frog species. The methodology involves fractionation of the secretions by reversed-phase HPLC, identification of fractions containing components that stimulate the rate of release of insulin from BRIN-BD11 clonal β-cells without simultaneously stimulating the release of lactate dehydrogenase, identification of active peptides in the mass range 1–6 kDa by MALDI-TOF mass spectrometry, purification of the peptides to near homogeneity by further HPLC, and structural characterization by automated Edman degradation. The effect of synthetic replicates of the active peptides on glucose homeostasis in vivo may be evaluated in mice fed a high fat diet to produce obesity, glucose intolerance, and insulin resistance.

**Key words** Frog skin secretions, Antidiabetic peptide, BRIN-BD11 cells, Insulin, Incretin

## 1 Introduction

The dramatic increase in the global incidence of Type 2 diabetes mellitus has necessitated a search for new types of therapeutic agent. Several peptide-based drugs that mimic the enteroinsular axis by stimulating the release of insulin (known as incretins) have already been adopted in clinical practice [1]. A number of frog skin peptides that were first identified on the basis of their antimicrobial and/or immunomodulatory activities have subsequently been shown to possess the ability to release insulin from BRIN-BD11 cells at low concentrations that are not cytotoxic to the cells. The glucose-responsive BRIN-BD11 clonal β-cell line, established after electrofusion of rat insulinonoma-derived RINm5F cells with New England Deaconess Hospital rat pancreatic islet β-cells [2], is a well-established and convenient model to study insulin release in response to a range of nutrients, neurotransmitters, and pharmacological agents [3]. Insulin-releasing peptides have been identified in

Michael Schrader and Lloyd Fricker (eds.), *Peptidomics: Methods and Strategies*, Methods in Molecular Biology, vol. 1719, https://doi.org/10.1007/978-1-4939-7537-2_21, 

skin secretions from frogs belonging to the families Alytidae, Dicroglossidae, Leptodactylidae, Hylidae, Pipidae, and Ranidae (reviewed in Ref. 4). Such peptides represent lead compounds for development of drugs for the treatment of patients with Type 2 diabetes mellitus.

While several frog skin incretin peptides are cytotoxic both to prokaryotic and to eukaryotic cells, generally at concentrations appreciably higher than those producing a significant effect on insulin release, others such as tigerinin-1R from *Hoplobatrachus rugulosus* [5] and frenatin-2D from *Discoglossus sardus* [6] are nontoxic at concentrations up to 500 μM. There is no single mechanism mediating the insulin-releasing action of the frog skin peptides but patch-clamp studies have shown that the analog [Arg$^4$]tigerinin-1R blocks $K_{ATP}$ channels in BRIN-BD11 cells and the resulting depolarization indirectly increases the activity of the L-type Ca channels leading to increase $Ca^{2+}$ influx and consequent increase in the rate of insulin secretion [7]. Major obstacles to the use of frog skin peptides as useful antidiabetic drugs are their toxicities and their short half-lives in the circulation. However, analogs of naturally occurring frog skin peptides have been designed that show reduced cytotoxic activity against mammalian cells while maintaining the ability to stimulate insulin release both in vitro [8] and in vivo [9]. Similarly, long-acting analogs of incretin peptides have been designed that incorporate D-amino acids to increase stability to peptidases and contain a fatty acid moiety that facilitates binding to albumin [10].

This article presents a general strategy for the identification and characterization of peptides in frog skin secretions with therapeutic potential as antidiabetic agents. Secretions are stimulated either by intradermal injection of norepinephrine or, in the case of sub-adult or very small frogs, by immersion of the animal in a solution of norepinephrine. After partial purification on Sep-Pak C-18 cartridges, the secretions are fractionated by reversed-phase HPLC on a preparative or semi-preparative column. Fractions are incubated with BRIN-BD11 cells and effects on insulin release are measured by radioimmunoassay or ELISA and cytotoxicity by measurement of the rate of release of lactate dehydrogenase. Bioactive peptides in the molecular mass range 1–6 kDa are identified by MALDI-TOF mass spectrometry and these components are purified to near homogeneity (>98% purity) by further reversed-phase HPLC using columns containing a range of different packing materials. The primary structures of the purified peptides are determined by automated Edman degradation.

Once peptides with insulin-releasing activity in vitro have been identified and characterized, the potential as antidiabetic agents of synthetic replicates may be evaluated using animal models of Type 2 diabetes. Mice fed a high fat diet to produce obesity, insulin resistance, and impaired glucose tolerance represent an appropriate model [11]. The effects of frog skin peptides on glucose homeostasis have been recently described for acute [12] and long-term studies [13–16].

## 2 Materials

The reagents and equipment required for stimulation of skin secretions from the frog and partial purification of peptides in the secretions using Sep-Pak C-18 cartridges have been described in previous chapters in this series [17, 18].

### 2.1 Preparative HPLC

1. Acetonitrile: (gradient grade for HPLC). This reagent is toxic and should preferably be used in a fume hood.
2. Ultrapure water: (18.2 MΩ-cm at 25 °C).
3. Trifluoroacetic acid: (redistilled suitable for protein sequencing). This reagent is extremely corrosive and hand and eye protection should be worn and solutions must be prepared in a fume hood.
4. An HPLC system capable of generating a binary gradient using pumps operating in the flow rate range 0.1–10 mL/min, a Rheodyne 7125 injection system equipped with a 2 mL loop, a detector capable of simultaneously monitoring at two wavelengths (typically 214 and 280 nm) (*see* **Note 1**).
5. Vydac (2.2-cm × 25-cm) 218TP1022 C-18 preparative reversed-phase HPLC column (Grace) (*see* **Notes 2** and **3**).
6. Helium cylinder for degassing of solvents (*see* **Note 4**).
7. Gastight 2.5 mL injection syringe.
8. 15-mm × 100-mm polypropylene tubes (*see* **Note 5**).
9. 1.5 mL clear polypropylene microfuge tubes.
10. Fraction collector with capacity to accommodate 15-mm × 100-mm tubes.
11. Speed-Vac vacuum centrifuge concentrator.

### 2.2 Cell Culture

1. Incubator maintained at 37 °C in an atmosphere of 5% $CO_2$-air.
2. Hemocytometer.
3. Inverted microscope.
4. Sterile 150 $cm^2$ vented tissue culture flasks (Corning).
5. Cell line: BRIN-BD11 rat clonal β-cells are supplied by the European Collection of Authenticated Cell Cultures (ECACC). The cells have been mycoplasma eradicated at ECACC and the stocks available for supply have undergone a further ten passages without detection of mycoplasma (*see* **Note 6**).
6. Media for cell culture: prepare all solutions in ultrapure water (18.2 MΩ-cm at 25 °C) under aseptic conditions.

   RPMI-1640 (Gibco) supplemented with 10% (v/v) fetal calf serum (Gibco, mycoplasma screened, virus screened), 11.1 mM glucose, and 5000 IU/L penicillin-streptomycin (*see* **Note 7**). Media is stored at 4 °C until use.

Hanks Buffered Saline Solution (HBSS) without indicator. Add 50 mL of concentrated (×10) HBSS to 450 mL autoclaved ultrapure water. Label with date of preparation and store at room temperature.

7. Reagents for insulin-release assay: prepare 100 mL Krebs Ringer bicarbonate buffer (KRBB) with the following composition: 115 mM NaCl, 4.7 mM KCl, 1.28 mM $CaCl_2$, 1.2 mM $KH_2PO_4$, 1.2 mM $MgSO_4$, 24 mM $NaHCO_3$ and supplement with 20 mM HEPES, 5.6 mM glucose and 0.1% (w/v) bovine serum albumin. Adjust the pH to 7.4 with 1 M NaOH and store 10 mL aliquots at −20 °C until use (*see* **Note 8**).
8. Trypsin: add 50 mL of concentrated (×10) trypsin-EDTA solution to 500 mL HBSS. Aliquot mixture into 20 mL sterile polypropylene tubes and store at −20 °C until use.
9. Trypan blue solution (0.4% w/v).
10. 24-well sterile polystyrene tissue culture plates.
11. Sterile pipette tips (200 µL and 1 mL).
12. 50 mL sterile polypropylene tubes.
13. 11-mm × 64-mm LP3 tubes for radioimmunoassay or ELISA.
14. Reagents for cytotoxicity assay: Cytotox 96 nonradioactive cytotoxicity assay kit (Promega).

### 2.3 MALDI-ToF Mass Spectrometry

1. A matrix-assisted laser desorption/ionization-time-of-flight mass spectrometer with reflector capability providing high-resolution mass spectra and the highest mass accuracy (*see* **Note 9**).
2. Sample diluent: 50% (v/v) acetonitrile-water containing 0.1% (v/v) trifluoroacetic acid.
3. MALDI matrix: α-cyano-4-hydroxycinnamic acid (α-CHCA) recrystallized and cation-depleted.
4. Peptide calibration samples. *Mixture 1*: [des-Arg$^1$]bradykinin ($[M + H]^+$ = 904.468) 2.3 µg, angiotensin-I ($[M + H]^+$ = 1296.685) 4.2 µg, [Glu$^1$]fibrinopeptide B ($[M + H]^+$ = 1570.6775) 1 µg, neurotensin ($[M + H]^+$ = 1672.917) 0.2 µg. *Mixture 2*: angiotensin-I ($[M + H]^+$ = 1296.685) 6.5 µg, $ACTH_{(1–17)}$ ($[M + H]^+$ = 2093.087) 10.5 µg, $ACTH_{(18–39)}$ ($[M+H]^+$ = 2465.199) 9.3 µg, $ACTH_{(7–38)}$ ($[M+H]^+$ = 3657.929) 27.5 µg, bovine insulin ($[M + H]^+$ = 5730.609 and $[M + 2H]^{2+}$ = 2865.808) 50.2 µg. These calibration mixtures are supplied in lyophilized form by Applera.
5. RBS35 detergent for cleaning sample plate.
6. Ultrasonic bath.
7. Laboratory oven.

## 3 Methods

### 3.1 HPLC Fractionation

1. Prepare solvent A by adding 1.2 mL of trifluoroacetic acid to 1000 mL ultrapure water and solvent B by adding 1.0 mL of trifluoroacetic acid to 700 mL acetonitrile +300 mL ultrapure water (*see* **Note 10**).
2. Degas solvent A and solvent B with helium for 1 min. This time should not be exceeded.
3. Column preparation. Before injecting the sample, it is necessary to "condition" whatever column is to be used in order to improve resolution by following these steps: irrigate the column at a flow rate of 6 mL/min with solvent B for 10 min; decrease the concentration of solvent B to 0% over 10 min using a linear gradient; increase the concentration of solvent B to 100% over 10 min; decrease the concentration of solvent B to 0% over 10 min; equilibrate the column with HPLC solvent A for 20 min.
4. Centrifuge the sample for 5 min at 13,000 × *g* in a 1.5 mL polypropylene microfuge tube to ensure clarity of solution (*see* **Note 11**).
5. Program the HPLC system to perform chromatography under the following conditions using linear gradients for elution: (a) increase concentration of solvent B from 0% to 30% over 10 min; (b) increase concentration of solvent B from 30% to 90% over 60 min; (c) increase concentration of solvent B from 90% to 100% over 1 min and hold at 100% until UV-absorbance returns to baseline value (*see* **Note 12**).
6. Inject the sample onto the column. Up to 1.5 mL may be injected into a 2 mL loop.
7. Increase the concentration of solvent B according to the elution program listed in Subheading 3.1, **step 5**. Collect fractions (1 min) into (15 mm × 100 mm) polypropylene tubes using a fraction collector (*see* **Note 13**).
8. Take aliquots (50–200 μL depending on the size of the UV-absorbing peak) of each fraction for determination of insulin-releasing activity and dry under reduced pressure in 1.5 mL microfuge tubes using a Speed-Vac concentrator.

### 3.2 BRIN-BD11 Cell Culture

Carry out all procedures under aseptic conditions preferably in a laminar flow cabinet. Prior to experimentation, the cells are grown to confluence in cell culture medium (Subheading 2.2, **item 6**) in a 150 $cm^2$ tissue culture flask.

1. Remove media from the tissue culture flask.
2. Wash cells with 10 mL HBSS and remove supernatant.

3. Add 3 mL of trypsin-EDTA solution (Subheading 2.2, **item 8**), swirl over cells and remove supernatant.
4. Wait until cells have detached.
5. Add 8 mL media, rock the tissue culture flask, and examine under an inverted microscope (20× magnification) to ensure a population of single cells.
6. Resuspend cells by aspirating into a 1 mL pipette tip.
7. Place cell suspension in a 50 mL polypropylene tube and centrifuge (900 rpm, 5 min).
8. Decant supernatant and add 10 mL media to a 50 mL polypropylene tube and resuspend cells.
9. Transfer 100 μL cell suspension to a 1.5 mL microfuge tube for determination of cell concentration.
10. Add 100 μL 0.4% trypan blue solution.
11. Mix thoroughly by aspirating into the pipette tip.
12. Apply cell suspension to both ends of the counting chamber of the hemocytometer.
13. Count the number of unstained cells in four large squares (each contains 16 small squares). Only preparations with cell viability >98% are used.
14. Using sterile cut pipette tips, add calculated volume of cell suspension to each well of a 24-well tissue culture plate. For overnight attachment, $2 \times 10^5$ cells/well should be used. If cells are to remain in wells for a longer period, it is advisable to reduce the numbers added to each well to allow for further growth.
15. Add 1 mL media per well.
16. Leave the plate in an incubator overnight and perform acute test the following morning.

### 3.3 Insulin-Release Assay

1. Reconstitute HPLC fractions (Subheading 3.1, **step 7**) by adding 60 μL 0.12% trifluoroacetic acid-water, vortex for 30 s, and allow to stand 30 min. Add 1 mL supplemented KRBB buffer (Subheading 2.2, **item 7**) containing 5.6 mM glucose.
2. Prepare 1 mL pipette tips by cutting off a small portion of the end of the tips.
3. Warm all solutions to 37 °C in a water bath.
4. Pour off media from the plate gently and blot any drops of fluid on the plate gently with a tissue.
5. Tip plate slightly by setting it on top of its lid.
6. Add 1 mL HBSS buffer gently to each well at an angle using a cut pipette tip applying at the same point in each well.

7. Pour off the excess HBSS buffer and blot gently with a tissue.
8. Repeat washing procedure three times and return to incubator for 40 min after final wash with cells in contact with 1 mL HBSS.
9. Pour off HBSS buffer and blot gently.
10. Start clock.
11. At time 0 s, add 1 mL of reconstituted HPLC fraction #1 (Subheading 3.3, **step 1**) to well #1.
12. At time 10 s, add 1 mL of HPLC fraction #2 to well #2.
13. Repeat at 10 s intervals until all fractions are added to all wells, changing tips for each test solution.
14. Replace the 24-well plate in the incubator.
15. After 20 min, remove 975 μL from well #1 and transfer to a numbered LP3 tube.
16. Repeat at 10 s intervals until all test samples are transferred.
17. Immediately after removing samples from wells centrifuge (900 rpm for 5 min, at 4 °C).
18. Transfer 900 μL of the supernatant into a LP3 tube carefully avoiding removal of cellular debris.
19. Store all samples at −20 °C until time of analysis by radioimmunoassay or ELISA (*see* **Note 14**).

### 3.4 Cytotoxicity Assay

Cytotoxicity is assessed by determining the ability of the frog skin peptides to compromise the integrity of the plasma membrane of BRIN-BD11 cells leading to an increase in the rate of release of the cytosolic enzyme lactate dehydrogenase (LDH).

1. Seed $2 \times 10^5$ BRIN-BD11 cells into 24 well plates and allow to attach during overnight culture at 37 °C (Subheading 3.2).
2. Remove cell supernatants and replace by 1 mL supplemented KRBB buffer (Subheading 2.2, **item 7**) containing 1.1 mM glucose.
3. Preincubate monolayers of cells for 40 min at 37 °C.
4. Remove cell supernatants.
5. Incubate cells with reconstituted HPLC fractions (Subheading 3.3, **step 1**) for 20 min at 37 °C using supplemented KRBB buffer containing 5.6 mM glucose.
6. Remove 975 μL of the cell supernatant to numbered LP3 tubes.
7. Measure LDH concentrations in the cell supernatants using a CytoTox96 nonradioactive cytotoxicity assay kit following the manufacturer's recommended protocol. Full details of the experimental procedure are available online at www.promega.com/tbs.

### 3.5 MALDI Matrix Solution

1. Dissolve α-cyano-4-hydroxycinnamic acid in sample diluent to give a final concentration of 10 mg/mL.
2. Vortex solution for 15 s at low speed and place in an ultrasonic bath for 1 min.
3. Centrifuge solution (5000 × *g* for 1 min) to remove any undissolved matrix and allow the supernatant to stand for 10 min (*see* **Note 15**).

### 3.6 Analyte and Peptide Calibration Mixtures

1. Reconstitute each dried HPLC fraction in 10 μL of sample diluent.
2. Vortex samples for 15 s at low speed and centrifuge (5000 × *g* for 1 min).
3. Reconstitute the peptide calibration mixtures in 100 μL of sample diluent (*see* **Note 16**). Working solutions are prepared by dilution of 1 μL of the stock solution in 24 μL of matrix solution.
4. Preparation of the sample plate: scrub the sample plate clean with RBS35 detergent and rinse extensively with (a) ethanol, (b) 50% (v/v) ethanol–water, and (c) ultrapure water. Dry the plate in a laboratory oven at 80 °C for 15 min and allow to return to room temperature.
5. Load 1 μL of each reconstituted HPLC fraction onto the sample plate in a defined position changing tips for each sample.
6. Load 1 μL of matrix solution onto each sample drop and mix by aspiration into the pipette tip (*see* **Note 17**).
7. Allow the analyte/matrix mixture to dry in air at room temperature.
8. Load the sample plate into the mass spectrometer.

### 3.7 Mass Range Determination

1. Obtain a preliminary spectrum for each sample using the following parameters: linear mode (*see* **Note 18**), positive polarity (detection of $[M + H]^+$ ions), wide mass range (0.5–15 kDa), and default instrument settings (accelerating voltage 20,000 V; grid 95%; guide wire 0.05%; delay time 450 ns; 500 shots/spectrum) (*see* **Note 9**).
2. Adjust the laser intensity manually to improve signal-to-noise ratio (approximately 50:1). If the laser intensity is too high, the signal may be saturated.
3. Obtain a second spectrum of each sample using the following parameters: reflector mode (*see* **Note 18**), positive polarity, narrower mass range (0.5–6 kDa), and default instrument settings (accelerating voltage 20,000 V; grid 76%; guide wire 0.002%; delay time 255 ns; 500 shots/spectrum). The laser intensity is again adjusted manually to improve signal-to-noise ratio.
4. The sample plate is ejected from the instrument.

### 3.8 Accurate Mass Determination

The first round of spectral acquisitions has allowed determination of the mass range of the constituents in each fraction and the best mode of analysis for each sample.

1. Select a peptide calibration mixture appropriate to the mass range of each sample.
2. Apply 0.5–1 μL of working solution of the appropriate peptide calibration mixture to the sample plate as close as possible to the sample.
3. Allow the mixture to dry in air at room temperature.
4. Load the sample plate into the instrument.
5. In general, peptides from amphibian skin secretions do not exceed 6 kDa (50 amino acid residues) and are readily detectable in reflector mode and positive polarity. Consequently, two methods of spectrum acquisition covering two mass ranges are used. The first method covers the mass range from 0.5 to 2.5 kDa and employs peptide calibration mixture 1. The default instrument settings are as follows: accelerating voltage 20,000 V; grid 76%; guide wire 0.002%; delay time 100 ns; 500 shots/spectrum. The second method covers the mass range from 0.5 to 6 kDa and uses peptide calibration mixture 2. The default instrument settings are as follows: accelerating voltage 20,000 V; grid 76%; guide wire 0.002%; delay time 150 ns; 500 shots/spectrum.
6. Select the calibration peptide spot assigned to the first sample.
7. Analyze this spot using a default calibration and the appropriate method. Adjust the laser intensity manually to improve signal-to-noise ratio (approximately 50:1). Optimize resolution by setting the grid and guide wire voltages and delay time. Acceptable resolution is determined by the mass range. In the range 0.5–2.5 kDa, 6000 or greater is acceptable; for the range 0.5–6 kDa, at least 7000 is required.
8. Save the best spectrum and adjust the values of the observed masses of the peptides to correspond to the reference masses. Calibration standards appropriate to the mass of the frog skin peptide that will be analyzed are selected. The calibrated spectrum of the standard peptide mixture is generated using the software provided by the manufacturer of the instrument and saved.
9. Select the sample spot adjacent to the calibration peptide spot.
10. Analyze using the calibrated spectrum of the standard peptide mixture as external calibration file. Adjust the laser intensity and instrument settings to obtain a spectrum with acceptable signal-to-noise ratio and resolution. The spectrum provides

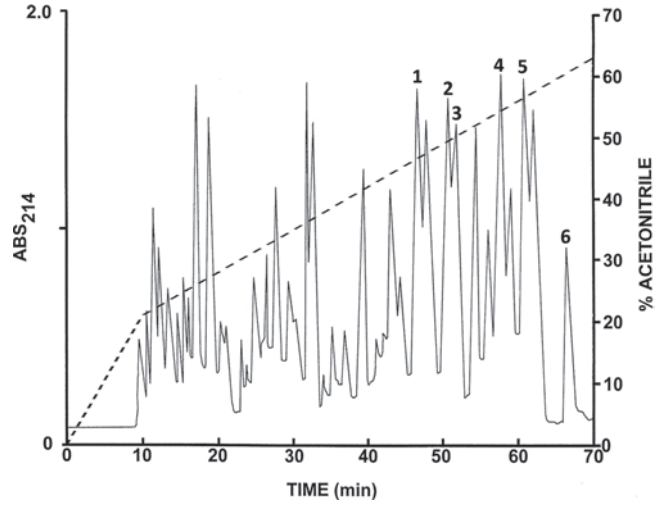



**Fig. 1** Elution profile (UV detection) on a preparative C18 RP-HPLC column of skin secretions from *Lithobates catesbeianus* after partial purification on Sep-Pak C-18 cartridges. Under the conditions of assay, the peaks designated 1–6 contained peptides that stimulated the rate of release of insulin from BRIN-BD11 clonal β-cells. Only the peptide present in peak 6 stimulated the rate of release of lactate dehydrogenase. The dashed line shows the concentration of acetonitrile in the eluting solvent. The primary structures, molecular masses and activities of the peptides are shown in Table 1

an accurate mass of each peptide component with a precision of 0.002% according to the specifications of the instrument.

11. Repeat the procedure for each sample/peptide calibration mixture pair (*see* **Note 19**).

In the example shown in Fig. 1, norepinephrine-stimulated skin secretions obtained from the bullfrog *Lithobates catesbeianus*, after partial purification on Sep-Pak cartridges, were chromatographed on a (2.2-cm × 25-cm) Vydac 218TP510 preparative $C_{18}$ column. Aliquots of the major peaks in the chromatogram were subjected to analysis for insulin-releasing activity (Subheading 3.3) and cytotoxic activity (Subheading 3.4). Under the conditions of assay, the peaks designated 1–6 were shown to stimulate the release of insulin from BRIN-BD11 cells but only the aliquot from peak 6 stimulated the release of lactate dehydrogenase. The monoisotopic masses of the peptides present in peaks 1–6 (Fig. 1), determined by MALDI-ToF mass spectrometry, are shown in Table 1. Determination of their amino acid sequences by automated Edman degradation demonstrated that the peptides belonged to the brevinin-1, palustrin-2, ranatuerin-1, ranatuerin-2, and temporin families previously identified on the basis of their antimicrobial activity.

**Table 1**
**Primary structure, molecular masses and effects on the rate of release of insulin and lactate dehydrogenase from BRIN-BD11 cells of peptides isolated from norepinephrine-stimulated skin secretions of the bullfrog *Lithobates catesbeianus***

| Peak no. | Peptide | Primary structure | $[M + H]^+$ observed (calc.) | Insulin release (ng/$10^6$ cells/min) | LDH release % of control |
|---|---|---|---|---|---|
| | Basal | | | 1.01 ± 0.06 | 100 |
| 1 | Temporin-CBf | FLPIASMLGKYL.$NH_2$ | 1352.8 (1352.7) | 2.21 ± 0.14* | 102 ± 1 |
| 2 | Temporin-CBa | FLPIASLLGKYL.$NH_2$ | 1334.7 (1344.7) | 2.03 ± 0.08* | 104 ± 1 |
| 3 | Ranatuerin-1CBa | SMLSVLKNLGKVGLGFVACKINKQC | 2649.8 (2650.3) | 2.50 ± 0.08* | 103 ± 2 |
| 4 | Ranatuerin-2CBc | GFLDIIKNLGKTFAGHMLDKIKCTIGTCPPSP | 2820.0 (2820.5) | 2.72 ± 0.12* | 103 ± 1 |
| 4 | Palustrin-2CBa | GFLDIIKDTGKEFAVKILNNLKCKLAGGCPP | 3302.5 (3303.0) | 2.36 ± 0.02* | 103 ± 1 |
| 5 | Ranatuerin-2CBd | GFLDIIKNLGKTFAGHMLDKIRCTIGTCPPSP | 3443.5 (3444.2) | 2.39 ± 0.12* | 102 ± 1 |
| 6 | Brevinin-1CBb | FLPFIARLAAKVFPSIICSVTKKC | 2651.9 (2652.4) | 2.88 ± 0.15* | 139 ± 6* |

Peptides were tested at a concentration of 3 μM
*Denotes a significant increase ($P < 0.001$)
The values in parentheses show the calculated monoisotopic ion masses

## 4 Notes

1. Although most HPLC systems incorporate an on-screen computer recording system, simultaneous monitoring with a dual pen flatbed chart recorder is useful especially when there is a significantly delay in the response of the absorbance detector and the appearance of the peak on the computer screen.
2. High quality preparative reversed-phase HPLC columns suitable for peptide purification are also manufactured by Agilent, Phenomenex, Supelco, and Waters.
3. When skin secretions are collected from very small or sub-adult frogs, the sample may contain only low amounts of peptide material. In this case, chromatography should be carried out on a Vydac (1.0-cm × 25-cm) 218TP510 semi-preparative C-18 reversed-phase column (Grace) operated at a flow rate of 2 mL/min.
4. Degassing of solvents may be achieved although less efficiently by using an ultrasonic bath but degassing by reducing the pressure is not recommended as it may lead to relative loss of volatile components.
5. Polypropylene tubes should be used throughout, not glass or polystyrene, in order to minimize irreversible binding of peptides to the tubes.
6. While studies with the BRIN-BD11 cell line are described here, there are several other readily available rodent-derived (HIT-T15, INS-1, MIN6, RINm5F) and human-derived (1.1B4, EndoC-βH1) insulin secreting cell lines that can be used [19].
7. The penicillin-streptomycin solution is aliquoted into sterile 5 mL vials and stored at −20 °C until use. With continued use, RPMI-1640 media will turn from an orange color to a pink color. Media should be discarded if it has become too pink/red.
8. For 100 mL of KRBB media, the following amounts (mg) are used: NaCl 672.1, KCl 35.04, $CaCl_2{\cdot}6H_2O$ 28.04, $KH_2PO_4$ 16.33, $MgSO_4{\cdot}7H_2O$ 29.58, $NaHCO_3$ 210.0.
9. The procedure described refers to the use of a Voyager DE-PRO mass spectrometer equipped with delayed extraction reflector (Applied Biosystems) but is readily adaptable to other instruments.
10. A slightly greater concentration of trifluoroacetic acid in HPLC solvent A than in solvent B produces a flat baseline under HPLC gradient elution conditions.
11. Filtering the sample is not recommended unless absolutely necessary as it can lead to appreciable loss of peptide by irreversible binding to the filter material.

12. On completion of the chromatography, wash the column with acetonitrile (100 mL) and store in this solvent.
13. Fractions may be stored in stoppered tubes at −20°C for up to several months. However, peptides containing methionine and tryptophan residues may oxidize on prolonged storage, even at low temperature.
14. In the authors' laboratory, insulin radioimmunoassay is performed according to the procedure described in Ref. 20 using an antiserum raised in-house. However, a radioimmunoassay kit for measurement of rat insulin concentrations is supplied by Merck-Millipore (RI-13K). Alternatively, rat insulin may be determined by ELISA using kits supplied by Thermo Fisher Scientific (ERINS), Abnova (KA3811), and Alpco (80-INSRT-E01).
15. The matrix solution is stored at 4°C and may be used for up to 1 week.
16. The peptide calibration mixtures are stored in single use aliquots (1 μL) at −20 °C.
17. It is important not to touch the surface of the plate with the pipette tip to avoid uneven crystallization.
18. The linear mode of operation is the most sensitive due to shorter flight path whereas reflector mode provides higher resolution and greater mass accuracy due to longer flight path and focusing action at the detector.
19. Peptide purity is assessed by a symmetrical peak shape measured at two wavelengths (214 and 280 nm) and by mass spectrometry. When mass spectrometry reveals that the fraction contains multiple components, peptides may be separated by further chromatography on a (1.0-cm × 25-cm) semi-preparative butylsilylsilica (C-4) column and a (1.0-cm × 25-cm) semi-preparative phenyldimethylsilyl silica column. The use of these columns and the general strategy for purification of peptides to near homogeneity by reversed-phase HPLC are discussed in detail in Refs. 17, 21, and. At this stage of the purification procedure, peak collection by hand is preferable to use of a fraction collector and generally results in peptide fractions that are sufficiently pure for amino acid sequence analysis.

## 5 Conclusions

At the present time (October 2016), 6640 species of frogs have been described [22] but only a small proportion of them have been investigated for the presence of bioactive peptides in their secretions. The focus of this article has been the identification of peptides with insulin-releasing activity with a view to their development

into agents for the treatment of patients with Type 2 diabetes. However, by employing different appropriate bioassays, the general strategy described may be applied to identification of frog skin peptides with antibacterial, antifungal, and antiviral activity for treatment of infections produced by multidrug-resistant microorganisms [4, 23] as well as immunomodulatory peptides with complex effect on the production of anti-inflammatory and pro-inflammatory cytokines for treatment of patients with sepsis and with a possible role in anticancer therapy [24]. More speculatively, frog skin secretions contain peptides with potential as agents to promote wound healing [25], peptides with spermicidal activities for development into contraceptives [26], and opioid peptides for development into analgesics [27].

## References


1. Østergaard L, Frandsen CS, Madsbad S (2016) Treatment potential of the GLP-1 receptor agonists in type 2 diabetes mellitus: a review. Expert Rev Clin Pharmacol 8:1–25
2. McClenaghan NH, Barnett CR, Ah-Sing E, Abdel-Wahab YHA, O'Harte FP, Yoon TW et al (1996) Characterization of a novel glucose-responsive insulin-secreting cell line, BRIN-BD11, produced by electrofusion. Diabetes 45:1132–1140
3. McClenaghan NH, Flatt PR (1999) Physiological and pharmacological regulation of insulin release: insights offered through exploitation of insulin-secreting cell lines. Diabetes Obes Metab 1:137–150
4. Conlon JM, Mechkarska M, Lukic ML, Flatt PR (2014) Potential therapeutic applications of multifunctional host-defense peptides from frog skin as anti-cancer, anti-viral, immunomodulatory, and anti-diabetic agents. Peptides 57:67–77
5. Srinivasan D, Ojo OO, Abdel-Wahab YHA, Flatt PR, Guilhaudis L, Conlon JM (2014) Insulin-releasing and cytotoxic properties of the frog skin peptide, tigerinin-1R: a structure-activity study. Peptides 55:23–31
6. Conlon JM, Mechkarska M, Pantic JM, Lukic ML, Coquet L, Leprince J et al (2013) An immunomodulatory peptide related to frenatin 2 from skin secretions of the Tyrrhenian painted frog *Discoglossus sardus* (Alytidae). Peptides 40:65–71
7. Ojo OO, Srinivasan DK, Owolabi BO, McGahon MK, Moffett RC, Curtis TM et al (2016) Molecular mechanisms mediating the beneficial metabolic effects of [Arg$^4$]tigerinin-1R in mice with diet-induced obesity and insulin resistance. Biol Chem 397:753–764
8. Manzo G, Scorciapino MA, Srinivasan D, Attoub S, Mangoni ML, Rinaldi AC et al (2015) Conformational analysis of the host-defense peptides pseudhymenochirin-1Pb and -2Pa and design of analogues with insulin-releasing activities and reduced toxicities. J Nat Prod 78:3041–3048
9. Owolabi BO, Ojo OO, Srinivasan DK, Conlon JM, Flatt PR, Abdel-Wahab YH (2016) *In vitro* and *in vivo* insulinotropic properties of the multifunctional frog skin peptide hymenochirin-1B: a structure-activity study. Amino Acids 48:535–547
10. Uccellatore A, Genovese S, Dicembrini I, Mannucci E, Ceriello A (2015) Comparison review of short-acting and long-acting glucagon-like peptide-1 receptor agonists. Diabetes Ther 6:239–256
11. Heydemann A (2016) An overview of murine high fat diet as a model for Type 2 diabetes mellitus. J Diabetes Res 2016:2902351
12. Abdel-Wahab YH, Patterson S, Flatt PR, Conlon JM (2010) Brevinin-2-related peptide and its [D4K] analogue stimulate insulin release in vitro and improve glucose tolerance in mice fed a high fat diet. Horm Metab Res 42:652–626
13. Srinivasan D, Ojo OO, Owolabi BO, Conlon JM, Flatt PR, Abdel-Wahab YH (2015) The frog skin host-defense peptide CPF-SE1 improves glucose tolerance, insulin sensitivity and islet function and decreases plasma lipids in high-fat fed mice. Eur J Pharmacol 764:38–47
14. Ojo OO, Srinivasan DK, Owolabi BO, Conlon JM, Flatt PR, Abdel-Wahab YH (2015) Magainin-AM2 improves glucose homeostasis and beta cell function in high-fat fed mice. Biochim Biophys Acta 1850:80–87
15. Ojo OO, Srinivasan DK, Owolabi BO, Vasu S, Conlon JM, Flatt PR et al (2015) Esculentin-2CHa-related peptides modulate islet cell

function and improve glucose tolerance in mice with diet-induced obesity and insulin resistance. PLoS One 10:e0141549

16. Owolabi BO, Ojo OO, Srinivasan DK, Conlon JM, Flatt PR, Abdel-Wahab YH (2016) Glucoregulatory, endocrine and morphological effects of [P5K]hymenochirin-1B in mice with diet-induced glucose intolerance and insulin resistance. Naunyn Schmiedeberg's Arch Pharmacol 389:769–781
17. Conlon JM, Sonnevend A (2010) Antimicrobial peptides in frog skin secretions. Methods Mol Biol 618:3–14
18. Conlon JM, Leprince J (2010) Identification and analysis of bioactive peptides in amphibian skin secretions. Methods Mol Biol 615: 145–157
19. McCluskey JT, Hamid M, Guo-Parke H, McClenaghan NH, Gomis R, Flatt PR (2011) Development and functional characterization of insulin-releasing human pancreatic beta cell lines produced by electrofusion. J Biol Chem 286:21982–219992
20. Flatt PR, Bailey CJ (1981) Abnormal plasma glucose and insulin responses in heterozygous lean (ob/+) mice. Diabetologia 20: 573–577
21. Conlon JM (2007) Purification of naturally occurring peptides by reversed-phase HPLC. Nat Protoc 2:191–197
22. Frost DR (2016) Amphibian species of the world: an online reference. Version 6.0 Electronic Database. American Museum of Natural History, New York. http://research.amnh.org/herpetology/amphibia/index.html. Accessed 21 Oct 2016
23. Xu X, Lai R (2015) The chemistry and biological activities of peptides from amphibian secretions. Chem Rev 115:1760–1846
24. Pantic JM, Radosavljevic GD, Jovanovic IP, Arsenijevic NN, Conlon JM, Lukic ML (2015) *In vivo* administration of the frog skin peptide frenatin 2.1S induces immunostimulatory phenotypes of mouse mononuclear cells. Peptides 71:269–275
25. Mangoni ML, McDermott AM, Zasloff M (2016) Antimicrobial peptides and wound healing: biological and therapeutic considerations. Exp Dermatol 25:167–173
26. Zairi A, Tangy F, Bouassida K, Hani K (2009) Dermaseptins and magainins: antimicrobial peptides from frogs' skin-new sources for a promising spermicides microbicides-a mini review. J Biomed Biotechnol 2009:452567
27. Vardy E, Sassano MF, Rennekamp AJ, Kroeze WK, Mosier PD, Westkaemper RB, Stevens CW, Katritch V, Stevens RC, Peterson RT, Roth BL (2015) Single amino acid variation underlies species-specific sensitivity to amphibian skin-derived opioid-like peptides. Chem Biol 22:764–775

# Chapter 22

# High-Accuracy Mass Spectrometry Based Screening Method for the Discovery of Cysteine Containing Peptides in Animal Venoms and Toxins

**Luuk N. van Oosten, Martijn W.H. Pinkse, Mervin Pieterse, Pierre Escoubas, and Peter D.E.M. Verhaert**

## Abstract

Venom and toxin samples derived from animal origins are a rich source of bioactive peptides. A high proportion of bioactive peptides that have been identified in venom contain one or more disulfide bridges, which are thought to stabilize tertiary structure, and therefore influence the peptides' specificity and activity. In this chapter, we describe a label-free mass spectrometry-based screening workflow specifically to detect peptides that contain inter- and intramolecular disulfide bonds, followed by elucidation of their primary structure. This method is based on the determination of the normalized isotope shift (NIS) and the normalized mass defect (NMD) of peptides, two parameters which are heavily influenced by the presence of sulfur in a peptide, where cysteines are the main contributing residues. Using ant defensive secretions as an example, we describe the initial fractionation of the venom on strong cation exchange followed by nanoflow HPLC and mass spectrometry. High resolution zoom scan spectra of high-abundance peptides are acquired, allowing an accurate determination of both monoisotopic and average mass, which are essential for calculation of NMD and NIS. Candidate peptides exhibiting relative low NMD and high NIS values are selected for targeted de novo sequencing. By fine-tuning the collision energy for optimal fragmentation of each selected precursor ions, the full sequence of several novel inter- and intramolecular disulfide bond containing ant defensive peptides can be established.

**Key words** Mass defect, Isotope shift, Orbitrap, Cysteine, Disulfide bridge, Venom, Toxin, Peptide, De novo sequencing

## 1 Introduction

Venomous and toxic animal peptides are a great source of biologically active peptides. So far, over 6000 proteins and peptides from animal toxin origin have been deposited in the ToxProt database (a subset of UniProt) [1]. As such, this database already represents a wealthy source of biomolecules with potentially interesting biological activities. The pharmacological action of most of the peptides in ToxProt appears directed toward ion

Michael Schrader and Lloyd Fricker (eds.), *Peptidomics: Methods and Strategies*, Methods in Molecular Biology, vol. 1719, https://doi.org/10.1007/978-1-4939-7537-2_22, 

channels, G-protein coupled receptors, and membrane transporter proteins, which are all well-known drug targets [2]. Therefore, this specific group of peptides is a favorite study object with respect to their potential as drug leads [3–5]. In the ToxProt database the venoms of insects—and in particular peptides from the Hymenoptera class (wasps, bees, bumblebees, and ants), represent a rich and diverse resource with relatively few hymenopteran species studied so far (reviewed in Ref. 6).

The identification of individual bioactive peptide toxins from complex animal secretions is a laborious and time-consuming task, which typically requires extensive purification steps. Moreover, de novo sequencing of unidentified "natural" or "endogenous" (non-tryptic) peptides is further complicated by the high number and variety of posttranslational modifications (PTMs) present on them [7]. One of the most frequently occurring PTM is the disulfide bond formed by the condensation of two cysteine residues. It is known that thus formed cystines are involved in peptide secondary and tertiary structure formation (folding) and stabilization [8]. Therefore, peptides containing disulfide bridges are assumed to possess pharmacologically interesting bioactivity.

Our earlier review of ToxProt's compositional analysis [9] illustrated the high prevalence of cysteines (6.91% of all residues, approximately 6000 entries) compared to the complete UniProt database (1.37% of all residues, approximately 550,000 entries, Release 2015_1; the percentage of cysteine residues is unchanged in Release 2017_1).

In this chapter, we describe a technique that uses the presence of cysteine residues in peptides as a selection criterion for further de novo sequencing efforts. The basic principle and proof of concept of the method were previously published, using defensive skin secretory peptides of different amphibian species as a model [9]. The method makes use of the relatively large negative mass defect (difference between isotopic and nominal mass) and the positive isotopic shift (difference between average and monoisotopic mass) of the element sulfur (for a comprehensive review *see* Ref. 10). This allows for discrimination of sulfur containing (primarily cysteines) and sulfur lacking peptides. Normalization of these two mass-related shifts, which are calculated from high resolution mass spectra, results in two values which represent nonadditive and independent peptide characteristics [11]. Plotting the so-called normalized nominal mass defect (NMD) and normalized isotopic shift (NIS) against the mass of a peptide yields 3D mass maps in which cysteine residue containing peptides tend to cluster together due to their relatively low NMD, and increasing NIS compared to non-sulfur containing peptides [11]. Whereas this approach was originally shown to work for sets of known Cys encompassing peptides by Artemenko and coworkers [11], we demonstrated that the

high mass accuracy of the typical orbitrap analyzer allows for 3D mass mapping of high resolution Fourier transform MS data recorded from convoluted mixtures of uncharacterized peptide amalgams [9]. As such 3D mass maps elegantly assist in the detection of cysteine- and cystine-containing peptides out of complex peptide mixtures.

Prior to sequence analysis of peptides, disulfide bonds may be reduced to improve the quality of $MS^2$ spectra [12]. We previously discussed some shortcomings of other available methods [9] which rely on derivatization of free –SH groups on Cys residues with isotopic tags prior to differential analysis [13–17]. Most of these methods are relatively expensive and involve multiple reaction steps, and hence require relatively large amounts of pure starting material. Even with these premises fulfilled, these methods do not allow for easy detection (and subsequent elucidation) of hetero- and homo-multimeric peptides. The present method does not suffer from these drawbacks.

The comparative analysis of the mass shift of peptides before and after sample reduction, followed by alkylation, represents an easy way to establish the number of disulfide bridges in unlabeled samples. Drawbacks of this approach are the requirement of two separate high resolution mass analyses and the difficulty to identify dimeric peptides in complex mixtures. The method described in this chapter is label-free, allowing the detection of peptides from complex samples in their unreduced form, hence maintaining any secondary structure information.

## 2 Materials

All reagents were analytical grade solutions prepared using HPLC grade water (resistivity >18 MΩ·cm at 25 °C). Solutions required for SCX-HPLC and LC-MS were filtered and degassed prior to use.

### 2.1 Animals and Venom Sampling

1. Adult specimens of two ant species (*Odontomachus hastatus* and *Ectatomma tuberculatum*), collected in the wild (French Guyana) and stored at −20 °C prior to venom gland dissection.
2. Scalpel for careful tissue dissection.
3. Acetonitrile (10% solution) for subsequent peptide extraction.
4. Vacuum concentrator (such as a SpeedVac™) for sample volume reduction and lyophilization.

### 2.2 SCX-HPLC Fractionation

1. Approximately 1 mg of crude venom.
2. Strong cation exchange (SCX) HPLC column (we use 200 mm × 2.1 mm, 5 μm particles, pore size 300 Å; polysulfoethyl A™, PolyLC Inc., Columbia, MD, USA).

3. HPLC system with UV detection (we use Waters 2695 separations module equipped with a Waters 996 photodiode array).
4. Fraction collector (we use Bio-Rad 2110, Hercules, CA, USA) online coupled to HPLC.
5. Elution gradients: 25% acetonitrile in 10 mM ammonium formate, pH 3.0 (solvent $A_{SCX}$) and 25% acetonitrile in 500 mM ammonium formate, pH 6.8 (solvent $B_{SCX}$).

### 2.3 LC-MS and LC-MS/MS

1. Freshly prepared solution of 100 mM Tris (2-carboxyethyl) phosphine (TCEP).
2. Semi-preparative HPLC system with cooled autosampler (we use Agilent 1200 module).
3. 0.6% acetic acid as solvent $A_{LC\text{-}MS}$ and 80% acetonitrile in 0.6% acetic acid as solvent $B_{LC\text{-}MS}$.
4. Reversed-phase trap column (we use 20 mm × 100 μm $C_4$ column, 5 μm particles).
5. Reversed-phase analytical column (we use 140 mm L × 75 μm $C_4$ column, 5 μm particles).
6. Primary high resolution MS system (we use LTQ Orbitrap Velos™, ThermoFisher Scientific, Bremen, Germany).
7. Additional high quality tandem MS system (we use quadrupole Orbitrap hybrid instrument, Q Exactive Plus; ThermoFisher Scientific, Bremen, Germany), for fragmentation and analyses of multiple charge states of peptide ions.

### 2.4 Data Analysis

1. ReAdW software (version 2.0, available at http://sourceforge.net/projects/sashimi/files/) to convert Thermo *.raw files to *.mzXML files with (*see* **Note 1**).
2. MATLAB and Bioinformatics Toolbox (release 2014a, The MathWorks, Inc.) to determine monoisotopic mass, NIS and NMD.
3. MS deconvolution software (we use Thermo Xtract).

## 3 Methods/Analytical Protocols

### 3.1 Peptide Samples

1. Dissect the ant defensive glands. The number required depends on the size of the gland and the sensitivity of the mass spectrometer. Previously, we were able to identify peptides from pools of 134 individual glands of *O. hastatus*, and from 18 glands of *E. tuberculatum*.
2. After careful dissection of the ant defensive glands, the isolated tissues are pooled in 10% acetonitrile in water.
3. Particles and insoluble materials are removed by centrifugation at 12,000 × *g*. Pellets are discarded.

4. Supernatants are lyophilized and stored at −20 °C until further use. Both pooled samples yield approximately 50 mg of dried venom. As only 1 mg is required for SCX fractionation, this is ample material.
5. Prior to SCX-HPLC, approximately 1 mg of each respective venom sample is dissolved in 60 μL solvent $A_{SCX}$ and transferred to an HPLC autosampler vial.

### 3.2 SCX-HPLC Fractionation

1. The SCX column is equilibrated with solvent $A_{SCX}$.
2. For fractionation of the *E. tuberculatum* venom, a linear gradient from 25% to 90% solvent $B_{SCX}$ is applied in 43 min (*see* **Note 2**). Flow rate is set at 0.4 mL/min. Eluting peptides are monitored at 280 nm.
3. To separate the *O. hastatus* sample, a linear gradient from 0% to 50% solvent $B_{SCX}$ is applied in 40 min, followed by a 15 min linear gradient to 100% solvent $B_{SCX}$ (*see* **Note 2**). The flow rate is equally fixed at 0.4 mL/min, and eluting peptides are monitored at 280 nm.
4. Fractions of 0.4 mL (1 fraction/min) are collected and stored in the freezer (−20 °C) until further analysis.

### 3.3 LC-(High Resolution) MS

1. Fractions are split in two parts (50 μL each), one of which is reduced with freshly prepared TCEP to a final concentration of 20 mM. Reduction is allowed to take place for at least 20 min.
2. Samples (50 μL) are transferred to HPLC vials and placed in the cooled autosampler of the semi-preparative HPLC system.
3. The HPLC-MS system is set up in a vented column configuration, where the switch valve directs the flow after the trap column either to the waste (trapping) or to the analytical column (eluting). Details of the vented column setup are listed in **Note 3**.
4. The effluent of the analytical column is directly electrosprayed into the primary high resolution MS system using a gold coated spray needle. The capillary ionization voltage is maintained at 2 kV, resulting in a more or less stable electrospray with only slight variations between systems and eluents (*see* **Note 3**).
5. Of each sample, 10 μL are injected onto the trap column for 10 min. This is done to ensure removal of salts from the sample after the SCX-HPLC.
6. Peptides are eluted employing a linear gradient from 0% to 100% solvent $B_{LC\text{-}MS}$ in 60 to 120 min, depending on sample complexity (flow rate 150–200 nL/min). A graphical overview of **steps 6–8** is given in **Note 3**.
7. Full Fourier Transform MS scan range is set to 400–1500 *m/z* in profile mode.

8. The eluting peptide ions with highest intensity are selected for four microscans (*see* **Note 4**). This type of scan is also referred to as selected ion monitoring (SIM) or zoom scan. The automatic gain control target value is set to $10^5$, maximum injection time is fixed to 500 ms and resolution is set to 30,000 at *m/z* 400. Isolation width is set to −1.25 to +2.25 Da with respect to the ion precursor *m/z* (*see* **Note 5**).
9. Fragmentation spectra are acquired at a resolution of 30,000 at *m/z* 400, starting with a fixed first mass at *m/z* 100. The isolation window is set to 2.5 *m/z*, selection signal threshold is set to $10^5$, automatic gain control target value is $5 \times 10^5$, and maximum injection time is set to 200 ms. Fragmentation is performed at 24, 28 and 32 CE (*see* **Note 5**).

### 3.4 Data Analysis

1. The *.raw file of each LC-MS run is converted to an *.mzXML file using the ReAdw software.
2. Data are imported into MATLAB. From each zoom scan the monoisotopic, average, and nominal masses are determined. Details about the processing of the mass spectra can be found in the supplementary information of our previous publication [9]. A brief description is given in **Note 6**.
3. Employing Eqs. 1 and 3 in MATLAB, the nominal mass, normalized isotope shift and the normalized mass defect are calculated respectively [11].

$$\text{Nominal mass} = \text{round}\left(0.9995 \times \text{monoisotopic mass}\right) \quad (1)$$

$$\text{NIS} = 1000 \times \left(\text{average mass} - \text{monoisotopic mass}\right) / \left(\text{monoisotopic mass}\right) \quad (2)$$

$$\text{NMD} = 1000 \times \left(\text{monoisotopic mass} - \text{nominal mass}\right) / \left(\text{monoisotopic mass}\right) \quad (3)$$

4. The NMD (*x*-axis), NIS (*y*-axis) and overall mass (*z*-axis) of each peptide are plotted to yield a 3D map of the respective LC-MS run. As an example the 3D mass map of the complete *O. hastatus* is depicted in Fig. 1.
5. Peptides with a relatively low NMD (<0.55) and relative high NIS (>0.65) are prioritized for further analysis, as they are suspected to contain (multiple) cysteine residues. The corresponding higher-energy collisional dissociation (HCD) fragmentation spectra are investigated for fine-tuning the collision energy in subsequent de novo sequencing experiments.
6. Comparative analysis of data from unreduced and reduced venom fractions is indicative of the number of disulfide bridges and secondary structure of some peptides. An example is given in Fig. 2.

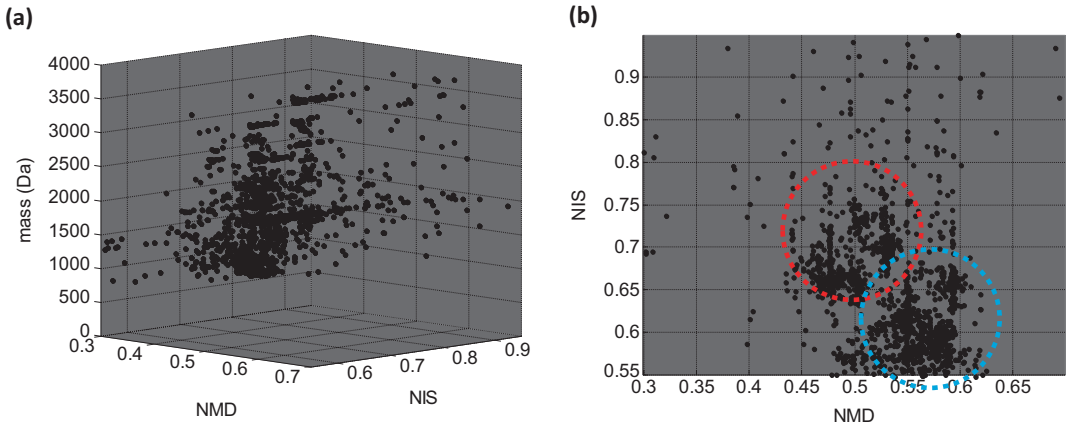



**Fig. 1** (**a**) Three-dimensional mass map of *O. hastatus* venom. As static representation of a 3D plot can be difficult to interpret, *see* (**b**) 2D representation (side view) of NMD versus NIS. In (**b**), two distinct clusters can be seen; one top left (encircled red), probably containing disulfide bond containing peptides. The other cluster (encircled blue) containing peptides without disulfide bonds. Comparison of samples before and after reduction indicated that from the cysteine rich cluster (red) 32 out of 34 most abundant peptides contain a disulfide bridge, whereas from the blue cluster only 2 out of 70 candidates contain a disulfide bridge

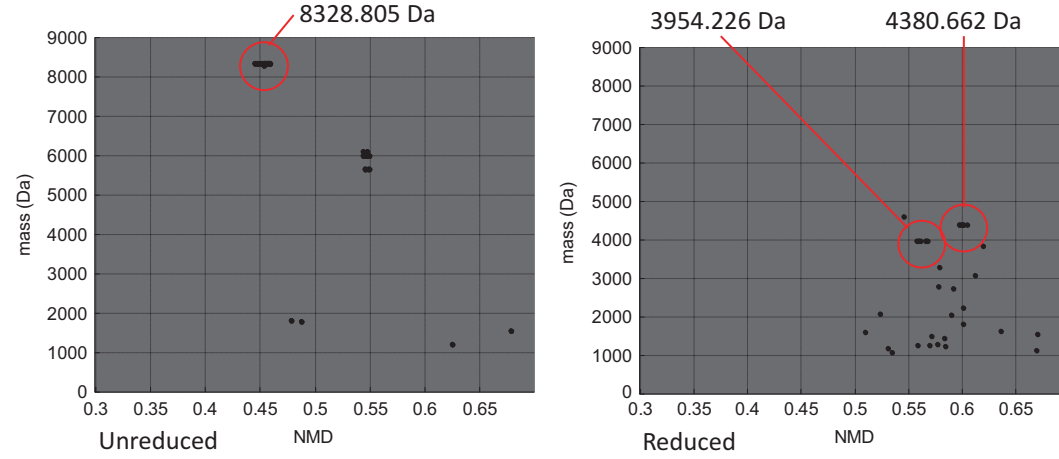



**Fig. 2** Detailed mass map of *E. tuberculatum* venom fraction. Left panel shows mass map (2D, NMD vs. mass) of a venom fraction which contains mainly peptide of 8328.805 Da. Right panel shows same sample after reduction with TCEP. Note disappearance of original spots the representing 8328.805 Da peptide, and appearance of two different abundant spots (masses 3954.226 Da and 4380.662 Da respectively). Mass difference before and after reduction (8328.805 − (3954.226 + 4380.66) = 6.083 Da) indicates this heterodimeric peptide contains 6 cysteine residues, a total of three disulfide bridges

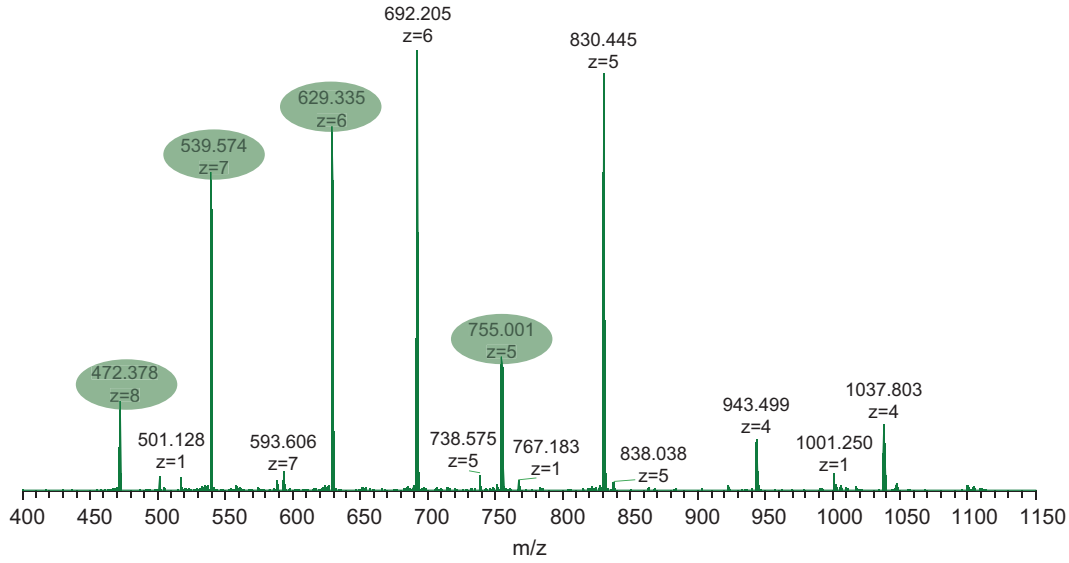



**Fig. 3** Elution of all charge states of peptide with mass monoisotopic mass 3768.975 Da. Fragmentation energies were fine-tuned for each charge state (5+ to 8+ ion, indicated in green). Several targeted MS runs were performed in which fragmentation energies were varied by 2 CE, until optimal CE was reached

### *3.5 LC-MS² Analysis*

1. SCX fractions containing peptides selected for de novo sequencing are reduced with TCEP, as described in Subheading 3.3.
2. Collision energy (CE) is fine-tuned for each eluting ion species (Fig. 3) on the additional high quality tandem MS system, yielding complementary fragmentation spectra (*see* Figs. 4 and 5). It is recommended to tune CE until the intensity of parent ion 10–30% relative abundance. As another example we provide in Fig. 6 fragmentation spectra of a 2 kDa peptide from *O. hastatus*. This peptide was suspected to contain a disulfide bridge based on its location in the 3D mass map, as depicted in Fig. 1.

## 4 Notes

1. Do not use a version of the ReAdW software later than 2.0, as the resulting *.mzXML files are incompatible with MATLAB software. There are no complications expected with using a version of MATLAB higher than 2014a.
2. For each sample, the SCX gradient parameters should be tuned. Essential in SCX-HPLC is that the majority of peptides in the venom get separated from each other in order to obtain relatively pure fractions. Keep in mind that prior to the actual MS measurement, another (online) LC step will be incorporated,

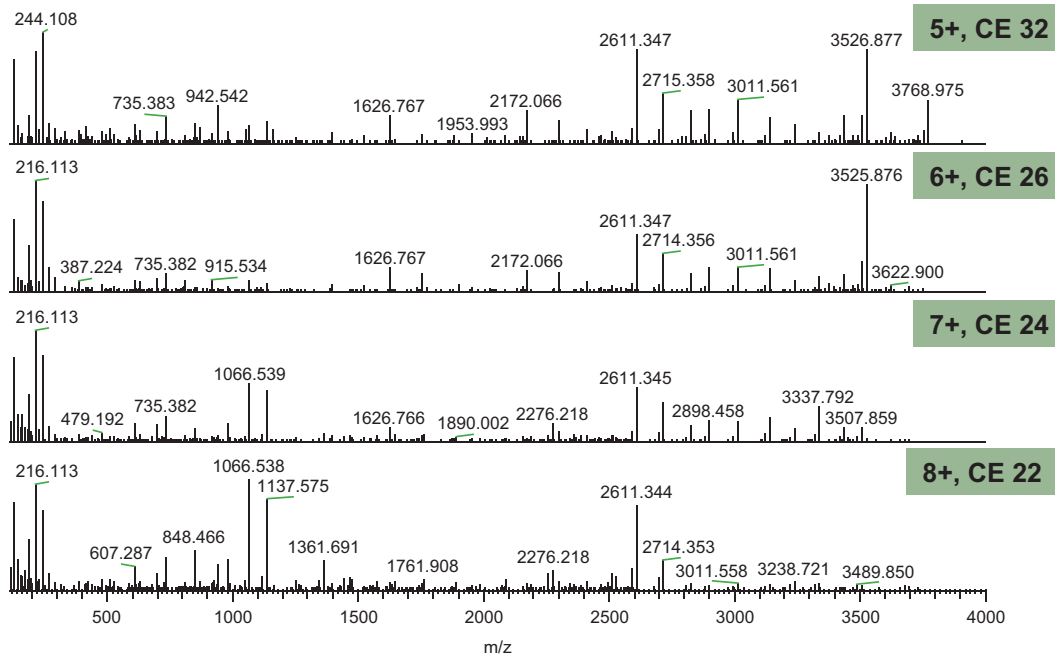



**Fig. 4** Fine-tuning HCD energy for peptide with mass monoisotopic mass 3768.975 Da. Ion species (5+ to 8+) yields complementary spectra. Note poor fragmentation of 8+ ion in higher *m/z* region (>2500), whereas 5+ ion yields excellent sequence coverage in same region. Reverse holds true for lower *m/z* region (<1000), where 8+ ion provides better sequence coverage compared to 5+ peptide ion

which will further assist in separating the peptides. Set up the vented column configuration as described in Fig. 7. This configuration is similar to Di Palma and coworkers [18]. The electrospray voltage might differ for other instrumentation.

3. The experimental LC-MS workflow to obtain data of monoisotopic mass, NMD and NIS can be roughly divided in three parts. Figure 8 gives a graphical representation of this data-dependent workflow.
4. With a maximum injection time at 0.5 s per microscan, the collection of 4 microscans (excluding dead time) takes around 2 s, which increases the duty cycle. In cases where the peptide already completely eluted during the zoom scan events, subsequent acquisition of fragmentation spectra can fail. In some instances it happens that during the time of microscans and subsequent fragmentation spectrum acquisition, another peptide elutes. This peptide will not be picked up by the mass spectrometer. In those cases adjusting the LC gradient will be required to improve capturing the missing peptides during another LC-MS run.
5. Depending on the size of the observed peptides, the isolation width might need adjustment. For peptides below 10 kDa, we

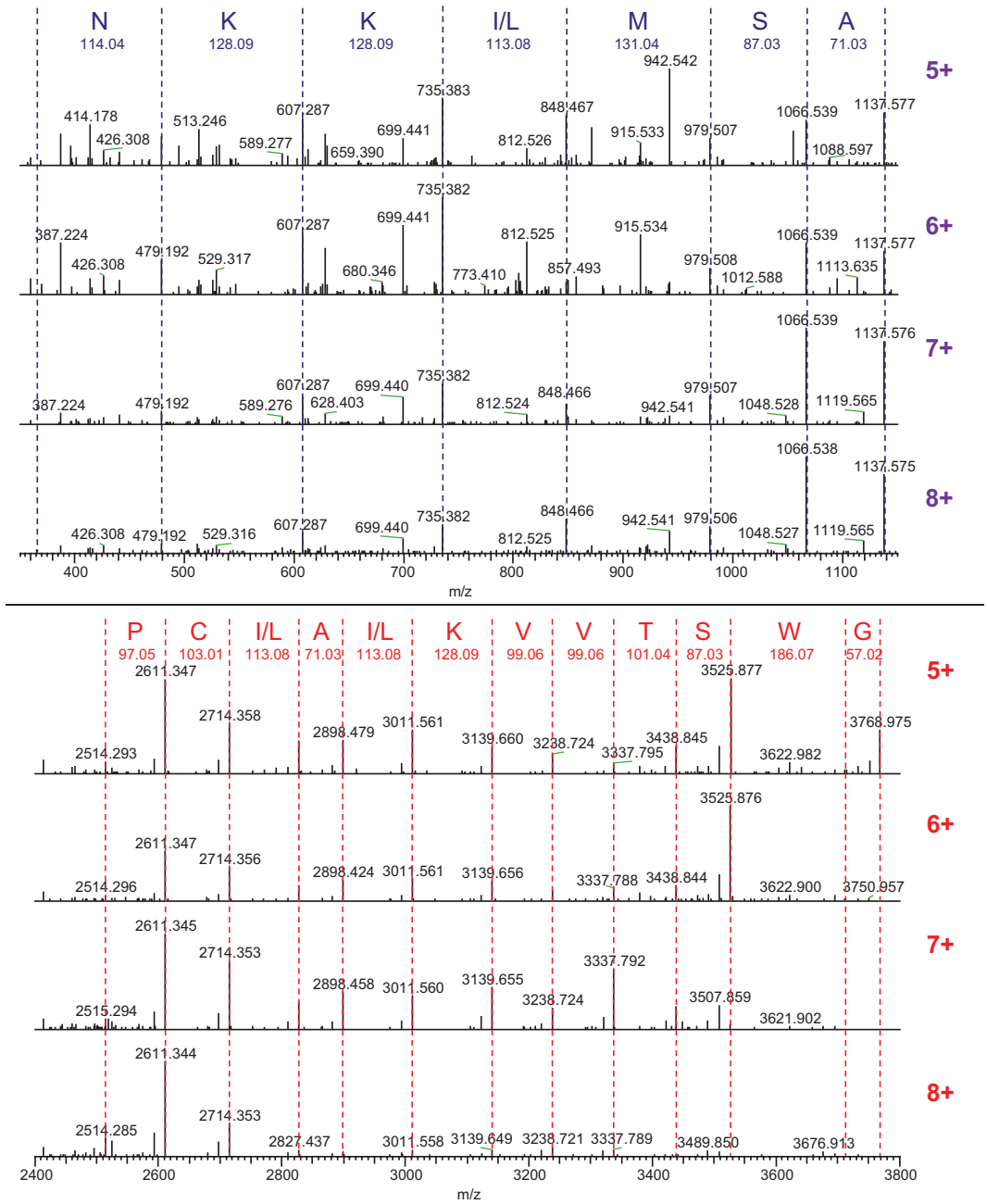



**Fig. 5** Sections of deconvoluted fragmentation spectra of *E. tuberculatum* peptide ETP-3768 $[M + 4H]^{4+}$ ion. In blue (top) lower *m/z* region (300–1200 Da) with improved fragmentation observed at higher charged ions. In red (bottom) sequence coverage in higher *m/z* region (2400–3800 Da) with improved sequence coverage in lower charged ions. Y-ions are indicated with dashed lines

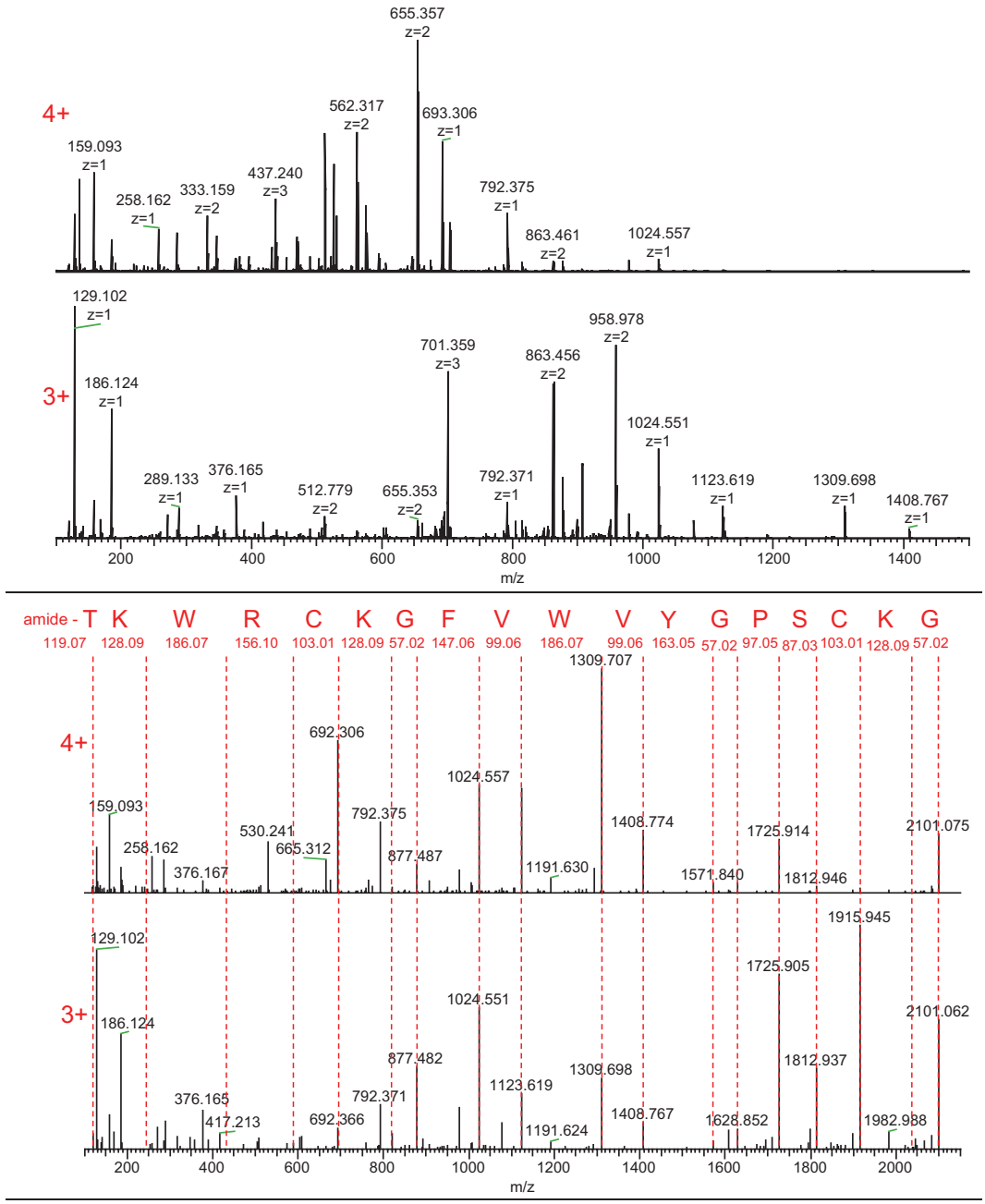



**Fig. 6** Top panel depicts fragmentation spectra of *O. hastatus* peptide OHP-2098 after fine-tuning CE for 3+ (*m/z* 701.0.28) and 4+ ion(*m/z* 526.021). Bottom depicts full sequence annotation of deconvoluted and de-isotoped spectra of 3+ and 4+ spectra. Red dashed lines indicate y-ions

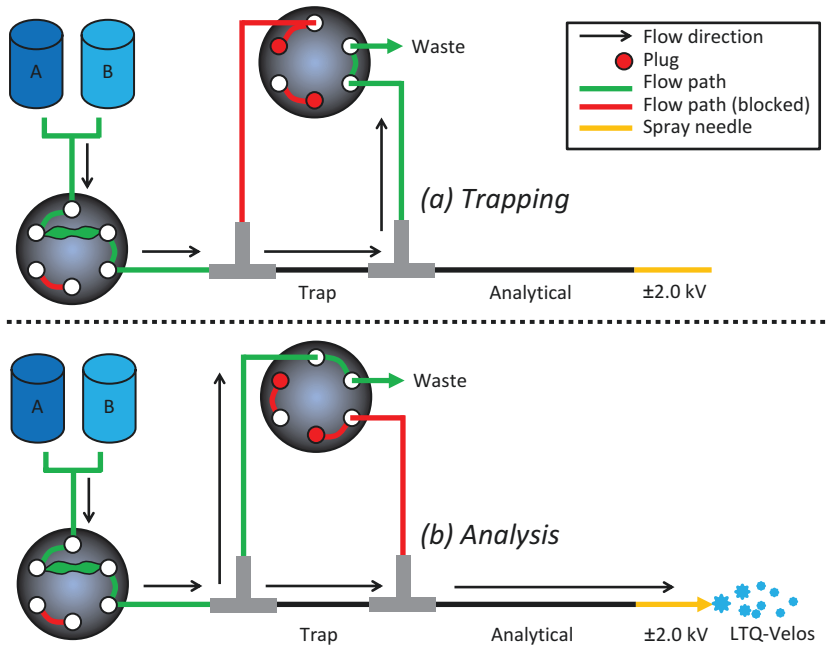



**Fig. 7** Vented column setup in trapping mode (**a**) and in analysis mode (**b**). Image adapted from Di Palma and coworkers [18]

recommend this isolation window. For larger peptides with low charges (<4+), it might occur that the full isotopic pattern of the peptide does not fall within this isolation width, due to the fact that the base peak (which is the reference for selection by the instrument) is not one of the first 2–3 peaks in the spectrum. The type and adjustments of parameters might differ for other MS instruments.

6. Zoom scan spectra were extracted from their *.mzXML files after import into MATLAB's workspace. Noise was removed from the spectrum by applying a simple threshold filter which removes all peaks with a relative intensity below 3%. From the mass difference between two highest peaks in each zoom spectrum the charge and monoisotopic mass were determined. From the full isotopic pattern of the peptides the average mass was calculated.

## Acknowledgments

Thanks to Dr. Axel Touchard (Université de Guyane, Kourou, France), Prof. Dr. Graham Nicholson (Sydney University of Technology, Sydney, Australia) for a generous supply of ant samples, collected in French Guyana.

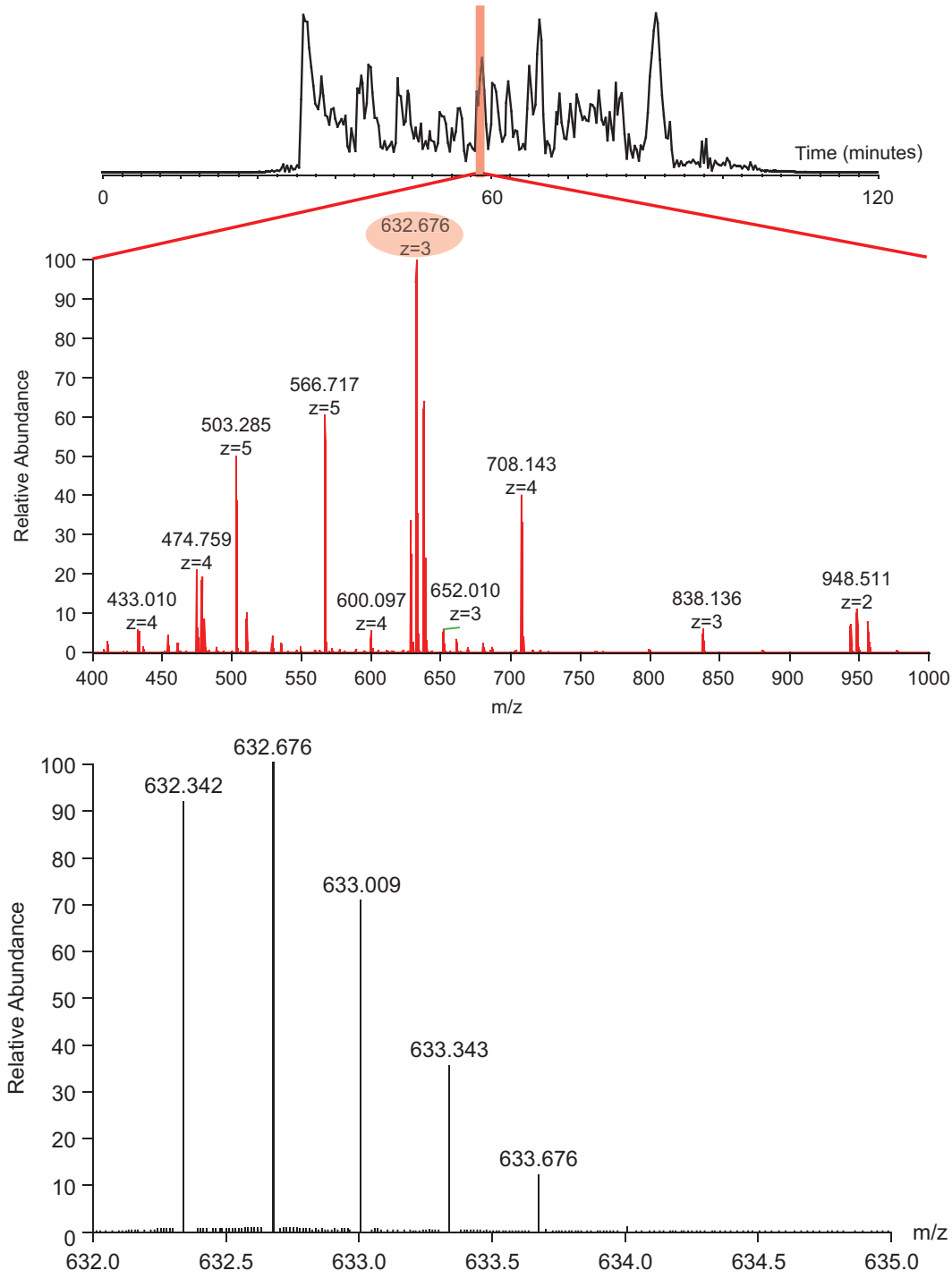



**Fig. 8** Depicted is an example of a typical elution profile of a venom sample (top chromatogram) with a full scan spectrum around retention time of 60 min (indicated in red) from 400 to 1000 *m/z* (middle). Most abundant peptide, in this case *m/z* 632.676 (3+ ion), was selected for microscan acquisition. Zoom spectrum (bottom) is recorded and is further processed in MATLAB to yield NMD and NIS of respective eluting peptide

## References


1. Jungo F, Bairoch A (2005) Tox-Prot, the toxin protein annotation program of the Swiss-Prot protein knowledgebase. Toxicon 45(3):293–301
2. Uhlig T, Kyprianou T, Martinelli FG et al (2014) The emergence of peptides in the pharmaceutical business: from exploration to exploitation. EuPA Open Proteom 4:58–69
3. Jungo F, Bougueleret L, Xenarios I et al (2012) The UniProtKB/Swiss-Prot Tox-Prot program: a central hub of integrated venom protein data. Toxicon 60(4):551–557
4. Escoubas P, King GF (2009) Venomics as a drug discovery platform. Expert Rev Proteomics 6(3):221–224
5. King GF (2011) Venoms as a platform for human drugs: translating toxins into therapeutics. Expert Opin Biol Ther 11(11):1469–1484
6. Touchard A, Aili S, Fox E et al (2016) The biochemical toxin arsenal from ant venoms. Toxins 8(1):30
7. Ueberheide BM, Fenyö D, Alewood PF et al (2009) Rapid sensitive analysis of cysteine rich peptide venom components. Proc Natl Acad Sci 106(17):6910–6915
8. Wedemeyer WJ, Welker E, Narayan M et al (2000) Disulfide bonds and protein folding. Biochemistry 39(15):4207–4216
9. van Oosten LN, Pieterse M, Pinkse MWH et al (2015) Screening method for the discovery of potential bioactive cysteine-containing peptides using 3D mass mapping. J Am Soc Mass Spectrom 26(12):2039–2050
10. Sleno L (2012) The use of mass defect in modern mass spectrometry. J Mass Spectrom 47(2):226–236
11. Artemenko KA, Zubarev AR, Samgina TY et al (2009) Two dimensional mass mapping as a general method of data representation in comprehensive analysis of complex molecular mixtures. Anal Chem 81(10):3738–3745
12. Zhang Y, Cui W, Zhang H et al (2012) Electrochemistry-assisted top-down characterization of disulfide-containing proteins. Anal Chem 84(8):3838–3842
13. Gygi SP, Rist B, Gerber SA et al (1999) Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. Nat Biotech 17(10):994–999
14. Murray CI, Uhrigshardt H, O'Meally RN et al (2012) Identification and quantification of S-nitrosylation by cysteine reactive tandem mass tag switch assay. Mol Cell Proteomics 11(2):M111.013441
15. Samgina TY, Vorontsov EA, Gorshkov VA et al (2011) Novel cysteine tags for the sequencing of non-tryptic disulfide peptides of anurans: ESI-MS study of fragmentation efficiency. J Am Soc Mass Spectrom 22(12):2246–2255
16. Thompson A, Schäfer J, Kuhn K et al (2003) Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. Anal Chem 75(8):1895–1904
17. Wang Z, Zhang Y, Zhang H et al (2012) Fast and selective modification of thiol proteins/peptides by N-(phenylseleno)phthalimide. J Am Soc Mass Spectrom 23(3):520–529
18. Di Palma S, Mohammed S, Heck AJR (2012) ZIC-cHILIC as a fractionation method for sensitive and powerful shotgun proteomics. Nat Protoc 7(11):2041–2055

# Chapter 23

# Analysis of the Snake Venom Peptidome

**Solange M.T. Serrano, André Zelanis, Eduardo S. Kitano, and Alexandre K. Tashima**

## Abstract

Snake venom peptidomes are known to be a large source of molecules with different pharmacological properties. The complexity and variability of snake venoms, the presence of proteinases, and the lack of complete species-specific genome sequences make snake venom peptidome profiling a challenging task that requires especial technical strategies for sample processing and mass spectrometric analysis. Here we describe a method for assessing the content of snake venom peptides and highlight the importance of sampling procedures, as they substantially influence the peptidomic complexity of snake venoms.

**Key words** Snake venom, Peptidome, Solid-phase extraction, Proteinase inhibition, Mass spectrometry

## 1 Introduction

Venomous snakes of the families *Elapidae*, *Viperidae*, and *Atractaspididae*, and some species of the *Colubridae* family, produce toxins (venom) that is used for immobilizing or killing prey and in defense against predators. Snake venom is produced in specialized secretory glands as an aqueous solution consisting of a high amount of proteins and peptides (up to 90% of dry weight) along with inorganic compounds (metal ions and inorganic anions) and organic components (citrate, carbohydrates, nucleosides, biogenic amines, amino acids, and lipids). Venom peptidomes are rich sources for drug discovery. Most biologically active peptides found in viperid snake venoms are derived from a precursor protein composed of seven bradykinin potentiating peptides (BPP) aligned in tandem after a hydrophobic signal peptide sequence followed by a poly-His-poly-Gly sequence (pHpG) and a C-type natriuretic peptide [1]. BPPs are inhibitors of the angiotensin-converting enzyme and are involved in the cardiovascular effects displayed by snake venoms [2]. Natriuretic peptides are involved in the venom-induced hypotension that contributes to rapid loss of consciousness in envenomed animals [3].

Michael Schrader and Lloyd Fricker (eds.), *Peptidomics: Methods and Strategies*, Methods in Molecular Biology, vol. 1719, https://doi.org/10.1007/978-1-4939-7537-2_23, 

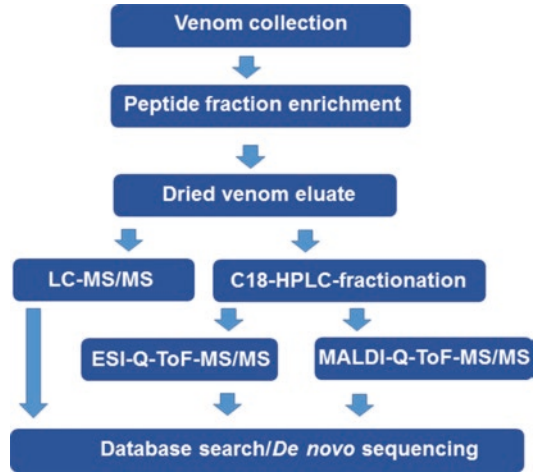



**Fig. 1** Schematic overview of the workflow applied to analyze the peptidome of the venom from *Bothrops* species

pHpG peptides were shown to inhibit the proteolytic activity of venom metalloproteinases [4]. Moreover, between the BPP sequences, this precursor protein also contains tripeptides with pyroglutamic acid at the N-terminus (pEKW; pEQW), which prevent the degradation of venom components by metalloproteinases within the gland.

Although there are few studies on the characterization of snake venom peptidomes, it is recognized that viperid venom peptidomes are complex [5–8] and ontogenetic and sex-based variation was demonstrated for the *Bothrops jararaca* venom peptidome [9–11]. Usually the analyses of snake venom proteomes and peptidomes are carried out using pooled samples from many specimens to overcome the individual variability that has been documented at both the protein and peptide levels. Moreover, fresh venom samples are frequently lyophilized to preserve the structural integrity of proteins. However, it has been shown that the lyophilization and resolubilization processes may impact the proteomes and peptidomes of snake venoms [6]. The conditions of venom extraction and sample processing for mass spectrometric analysis directly influence the levels of proteins, peptides, and citrate, as well as the pH of venom solutions. All of these factors can potentially impact the venom peptidome complexity due to the activity of venom proteinases upon venom components. The use of complementary methodologies of sample decomplexation, mass spectrometry, and data analysis, including de novo sequencing, can improve the results.

We established an analytical workflow to explore snake venom peptidomes (Fig. 1) that has successfully been used to investigate *Bothrops* species. This can be adapted to analyze other snake venoms. The general procedure described in this chapter for *Bothrops* snake venoms can be optimized for other species, depending on the amount of available venom peptide fraction and on the sensitivity, resolution, and mass accuracy of the mass spectrometer employed for the analysis.

## 2 Materials

All solutions must be freshly made. Use ultrapure water (resistivity of 18 MΩ•cm at 25 °C) and analytical grade reagents. Prepare and store all reagents at room temperature unless indicated otherwise. Fresh venom should be used whenever possible (*see* **Note 1**).

### 2.1 Venom Collection

1. 100–200 mL glass beaker immersed in ice bath (*see* **Note 2**).
2. Solution of 100 mM ethylenediaminetetracetic acid (EDTA) in ultrapure water, to inhibit metalloproteases.
3. Solution of 40 mM phenylmethanesulfonyl fluoride (PMSF) in isopropanol to inhibit serine proteases (*see* **Note 3**). Protect PMSF solution from light using amber flasks or wrap the solution container with aluminum foil.

### 2.2 Solid-Phase Extraction (SPE) of Peptides

1. SPE C18 cartridges (we use Sep-Pak C18 Plus from Waters, Milford, MA, USA).
2. Solution A: 0.1% (v/v) of trifluoroacetic acid (TFA).
3. Solution B: 0.1% TFA in water/acetonitrile (ACN) (70/30; v/v).

### 2.3 Liquid Chromatography–Tandem Mass Spectrometry (LC-MS/MS)

1. nanoLC chromatographer (we use nanoAcquity UPLC, Waters, Milford, MA) and ESI-Q-TOF mass spectrometer (we use Ultima API, Waters, Milford, MA).
2. 180 μm × 20 mm C18 trap column (we use nanoAcquity C18 trap column, Waters, Milford, MA).
3. 75 μm × 100 mm C18 analytical column (we use nanoAcquity C18 column, Waters, Milford, MA).
4. Prepare phase A1 with 0.1% formic acid in ultrapure water (v/v).
5. Prepare phase B1 with 0.1% formic acid in ACN (v/v).

### 2.4 RP-HPLC Fractionation

1. HPLC chromatographer (we use 10 AVp, Shimadzu, Kyoto, Japan).
2. 10 × 250 mm C18 column (we use Discovery 5 μm particle size; Supelco, Bellefonte, PA).
3. Prepare phase A2 with 0.05% TFA in ultrapure water (v/v).
4. Prepare phase B2 with 0.05% TFA in ACN (v/v).

### 2.5 ESI-Q-TOF Analysis of the Peptide Fractions

1. ESI-Q-TOF mass spectrometer (we use Ultima API, Waters, Milford, MA).
2. Prepare phase water/ACN/formic acid (~49.8/49.7/0.5, v/v/v).

### 2.6 MALDI-Q-TOF Analysis of the Peptide Fractions

1. MALDI-Q-TOF-MS/MS spectrometer (we use MALDI-Q-TOF Premier, Waters, Milford, MA).
2. Prepare saturated α-cyano-4-hydroxycinnamic acid solution (we use from Sigma, St. Louis, MO) in 0.1% TFA/ACN (70/30, v/v).

### 2.7 Data Analysis

1. MassLynx 4.1 (Waters, Milford, MA, USA).
2. ProteinLynx (Waters).
3. Uniprot database (www.uniprot.org).
4. NCBI database (www.ncbi.nlm.nih.gov/protein).
5. MASCOT Server (Matrix Science, London, UK) or comparable database search engine (PEAKS Studio, X! Tandem, etc.).

## 3 Methods

### 3.1 Venom Collection and Processing

Carry out all procedures at room temperature unless otherwise specified.

1. Prepare a 28 μL protease-inhibitor mix solution by adding 14 μL of the 100 mM EDTA and 14 μL of 40 mM PMSF stock solutions into a sterile microfuge tube (*see* **Note 4**). Carefully pipette 250 μL of the venom solution into the microfuge tube containing the protease-inhibitor mix solution. This will ensure that, after dilution, the final EDTA and PMSF concentrations are 5 mM and 2 mM, respectively.
2. Centrifuge the venom solution at 1700 × *g*, 4 °C, for 30 min, to remove any debris such as scales and mucus. Carefully pipette the clear supernatant after centrifugation and measure the protein concentration using a standard assay such as the Bradford method or comparable procedure. If not proceeding immediately with the next step, lyophilize and store venom aliquots at −20 °C until use (*see* **Note 1**).

### 3.2 Peptide Fraction Enrichment

1. Pipette the volume of fresh venom that contains 25 mg of protein (usually ~250 μL for *Bothrops* species), or weigh 25 mg of lyophilized venom, and acidify with 1 mL of Solution A to bring the pH to 2.0–3.0. Use C-18 cartridges for performing the SPE. For this purpose, perform the conditioning step according to manufacturer's instructions.
2. Load the acidified venom solution onto the conditioned SPE cartridge and wash the SPE cartridge with 2 mL of Solution A. For the elution of the peptide fraction, wash the SPE cartridge by adding 2 mL of Solution B (*see* **Note 5**).
3. Dry the samples using a vacuum centrifuge (ideally, under low temperature, e.g., 4 °C) and store the dried venom eluate at −20 °C until use.

### 3.3 LC-MS/MS Analysis

1. Dissolve the dried venom elute in 500 μL of solution A1.
2. Inject 5–10 μL of the peptide solution onto the reversed-phase C18 trap column and wash at 8 μL/min for 5 min (*see* **Note 6**).
3. After washing the trap column, connect the output of this column to flow into the main C18 column used for peptide separation.
4. Elute the peptide mixture from the C18 column into the ESI-MS using a linear gradient of 0–30% of solution B1 in 45 min, 30–80% in 10 min, and 80% B1 for 5 min at 300–600 nL/min.
5. Acquire MS spectra in the positive mode with capillary voltages of 3.5 kV and from $m/z$ of 250–2000 for 1 s.
6. Select the three most intense precursor ion signals with charge states between +2 and +5 for collision-induced dissociation (CID) fragmentation and acquire each MS/MS for 2.5 s from $m/z$ 50–2000. Vary collision energies from 15 to 56 eV, according to $m/z$ values and ion charges (*see* **Note 7**).
7. Apply a dynamic peak exclusion of 45 s (*see* **Note 8**).

### 3.4 Database Search

1. Convert the .raw MS data files to pkl or mgf format using for example the latest version of ProteinLynx.
2. Compile a Fasta database file with sequences of taxonomy Serpentes of Uniprot or of NCBI databases (*see* **Note 9**).
3. Perform database searches in the latest version of MASCOT Server with precursor and fragment tolerance masses of 0.3 Da (*see* **Note 10**), no enzyme specificity and pyroglutamic acid from N-terminal Gln or Glu as variable modifications.
4. Consider only peptide identifications with $p < 0.05$.

### 3.5 RP-HPLC Fractionation

Because the above analysis usually provides only partial identification of the snake venom peptidome, it is useful to perform additional fractionation and analysis of the samples at higher concentrations to acquire high quality spectra. The more concentrated peptide fractions can be directly infused to ESI-MS/MS or spotted to a plate and analyzed by MALDI-MS/MS. These additional steps are described in this and the next several sections.

1. Load 200 μL of the resuspended venom solution (from **step 1** of Subheading 3.3) into a 10 × 250 mm reversed-phase C18 column (5 μm particle size) using the binary pump HPLC system.
2. Elute the peptides using a gradient of 5% of solution B2 in 10 min, 5–40% in 35 min, 40–80% in 5 min, and 80% for 10 min at 1.5 mL/min.

3. Detect the peptides by UV absorbance at 215 nm and collect the fractions presenting high absorbance (*see* **Note 11**).
4. Dry the collected fractions in a vacuum centrifuge.

### 3.6 ESI-Q-TOF Analysis of the Peptide Fractions

1. Dissolve the dried RP-HPLC fractions in 20 μL of water/ACN/formic acid (~49.8/49.7/0.5, v/v/v).
2. Directly infuse the peptide solution with a syringe pump at 500 nL/min into the Q-TOF mass spectrometer coupled to a nano-ESI ionization source operated in positive mode (*see* **Note 7**).
3. Using MassLynx 4.1, acquire MS spectra in the $m/z$ range of 300–2000 for 1 min and observe multiply charged features (usually +2 to +4) for MS/MS analysis.
4. Select precursor ions under a 1 $m/z$ window and manually adjust collision energies from 15 to 55 eV for CID fragmentation (*see* **Note 12**). Acquire MS/MS spectra in the $m/z$ range of 50–2000.

### 3.7 MALDI-Q-TOF Analysis of the Peptide Fractions

1. Spot 1.5 μL of each RP-HPLC fraction (*see* **Note 13**) onto a MALDI sample plate, dry, mix with 1.5 μL of a saturated α-cyano-4-hydroxycinnamic acid solution in 0.1% TFA/ACN (70/30), and dry at room temperature.
2. Insert the MALDI sample plate in the MALDI-MS/MS spectrometer, wait until vacuum reaches proper level, and acquire MS spectra in the $m/z$ range of 800–4000, accumulating 15 laser shots.
3. Automatically select the ten most intense ions for MS/MS analysis and acquire spectra accumulating 30 laser shots.

### 3.8 De Novo Peptide Sequencing

1. From an MS spectrum of interest, calculate the peptide charge from the isotopic pattern and the monoisotopic mass from the precursor ion's first isotope.
2. Using MassLynx 4.1 and the MaxEnt 3 module to smooth and deisotope the MS/MS spectra acquired by ESI and MALDI (*see* **Note 14**).
3. Starting with the ESI-MS/MS spectrum, look for complementary $b_i$ and $y_{n-i}$ or $y_i$ and $b_{n-i}$ ion pairs, where $n$ is the peptide total number of amino acids and $i$ is the number of amino acids in the ion fragment. Check if the sum of the pair masses results in the precursor mass.
4. Once a $b_i$ or $y_i$ is found, search for adjacent ions that fit amino acid residue differences. Record the amino acids and repeat the procedure until the $b_1$ or $y_1$ is found.
5. Do the same procedure with the correspondent $y_{n-i}$ or the $b_{n-i}$, until the complete peptide sequence is obtained (Fig. 2) (*see* **Note 15**).

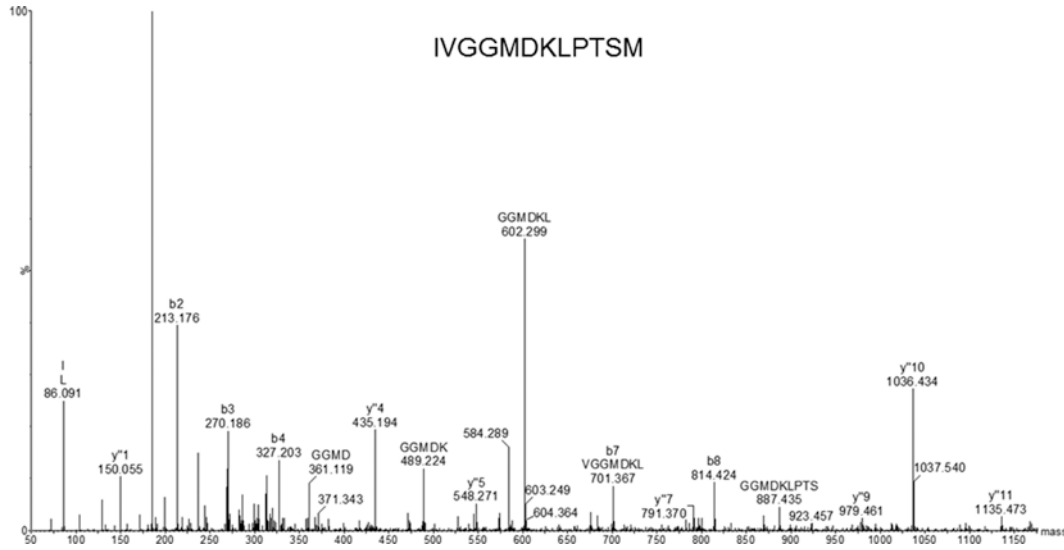



**Fig. 2** ESI-MS/MS spectrum of the peptide IVGGMDKLPTSM from an L-amino acid oxidase of *Bothrops cotiara* venom. Sequence obtained by manual de novo analysis of the precursor ion at m/z 624.80 ($[M + 2H]^{+2}$). Fragment ion series and internal fragments labeled in Biolynx module of MassLynx 4.1. Amino acids with identical masses (Ile/Leu) or very similar masses (Gln/Lys) were assigned according to BLAST homology search

6. In case of missing fragments look for fragment ions in the MALDI-Q-TOF spectra.
7. As a way of statistically validating the newly obtained de novo sequences, compile a Fasta database file with the peptides and perform an automated database search to check the peptide ion score and fragment assignments (*see* **Notes 16** and **17**).

## 4 Notes

1. The lyophilization and resolubilization processes may lead to degradation of venom proteins and generate distinct outcomes in the peptidome profile [6]. When long-term storage is needed, venoms can be lyophilized and stored at −20 °C; however, upon resolubilization, venoms should be left the minimum possible time in solution to avoid proteolysis. For the best evaluation of the venom peptidome, immediately add protease inhibitors to the collected venom and fractionate the peptides, as described in Subheadings 3.1 and 3.2.
2. Although the volume of venom extracted from a single adult *B. jararaca* specimen rarely exceeds 1 mL, and can vary depending on age and sex, a 100–200 mL beaker is preferred for the milking procedure as it better accommodates the snake's mouth. Since snake venoms are rich sources of biologically active

hydrolases, keeping the venom solution cold (e.g., in ice bath) is of paramount importance.

3. PMSF is soluble in anhydrous isopropanol at 35 mg/mL with heating, resulting in a clear to very slightly hazy, colorless to faint yellow solution. A 200 mM solution in dry solvent is stable for at least 9 months at 2–8 °C. However, its half-life changes dramatically in aqueous solutions: from 110 min to 35 min, at pH 7.0 and 8.0, respectively. Any leftover of the 40 mM stock solution should be discarded.
4. When studying the degradome of substrates of a given protease, its inhibitor should be avoided in the inhibitor mix solution. Additionally, avoid any protease inhibitor when analyzing the degradation profile of venom hydrolases (irrespective of the protease class).
5. Even though the washing volumes may vary according to different SPE cartridges, in most cases, using a lower organic solvent concentration (e.g., 30% of acetonitrile) ensures that the peptide fraction is promptly eluted. However, one must keep in mind that hydrophilic peptides may elute within the loading sample step (i.e., the flow through fraction). Therefore, this fraction may also be kept for further LC-MS/MS analysis.
6. It is recommended to perform analytical runs using a small sample volume to check the ideal peptide concentration to avoid either low or high intensity signals, which eventually lead to low ion fragmentation or detector saturation.
7. Other mass spectrometers, such as modern Q-TOF, or Orbitrap (Thermo Scientific, Bremen, GA) may be used for acquisition of MS and MS/MS spectra of the peptides. If possible, avoid acquiring MS/MS spectra using ion-trap mass analyzers due to their low mass cutoff, known as the "one third rule." The missing low mass fragments substantially limits de novo analysis of unknown sequences.
8. When dynamic peak exclusion time is enabled, the instrument will not repeat MS/MS scans of the same precursor ion already selected for fragmentation within the specified time. Also, the data-dependent acquisition mode favors the identification of co-eluting peptides by allowing the fragmentation of the low-abundant ones.
9. Other options may be evaluated in the compilation of sequence databases, such as the restriction to a genus group. After de novo analysis, the new sequences obtained may be included in the final, customized database.
10. The mass tolerances may need experimental adjustment according to the mass accuracy of the instrument used for MS analysis.
11. An analytical run with a small sample volume of 25–50 μL may be performed as a retention time guide of the peptide fractions to be collected in the definitive RP-HPLC run.

12. In general, 2–4 min is sufficient to acquire good spectra for de novo analysis.
13. The remaining volume prepared according to Subheading 3.6 can be recovered for MALDI-based analysis.
14. Snake venoms contain several proline-rich peptides, and these proline-containing fragments produce intense peaks in the ESI-MS/MS spectra, which, in some cases, suppress other neighboring fragments and prevent complete de novo analysis. Fragments from singly charged MALDI ions are richer in y-, b-, a-ions and internal fragments, and acquiring additional MS/MS spectra with MALDI ionization helps de novo sequencing, as the spectra from both ionization modes provide complementary information [12]. Both the MALDI and ESI-MS/MS analysis described in this chapter use CID to fragment the peptides; if available, a mass spectrometer with other fragmentation methods (such as electron-transfer dissociation or electron-capture dissociation) can provide additional complementary fragmentation information.
15. An Excel worksheet or a simple script as denovo.pl [6] to compute all amino acid differences on a computer screen may assist in manual de novo analysis.
16. Any proteomic database search engine (MASCOT Server, PEAKS Studio, X! Tandem, etc.) with the "no enzyme" specificity option can be used for this validation.
17. Automated de novo search tools may also be used to support and to accelerate de novo analysis; however, in this case, we recommend a manual validation step of the sequences obtained with peaks observed in the original spectra.

## References


1. Murayama N, Hayashi MA, Ohi H, Ferreira LA, Hermann VV, Saito H, Fujita Y, Higuchi S, Fernandes BL, Yamane T, de Camargo AC (1997) Cloning and sequence analysis of a Bothrops jararaca cDNA encoding a precursor of seven bradykinin-potentiating peptides and a C-type natriuretic peptide. Proc Natl Acad Sci U S A 94:1189–1193
2. Camargo AC, Ianzer D, Guerreiro JR, Serrano SM (2012) Bradykinin-potentiating peptides: beyond captopril. Toxicon 59:516–523
3. Fry BG, Vidal N, Norman JA, Vonk FJ, Scheib H, Ramjan SF, Kuruppu S, Fung K, Hedges SB, Richardson MK, Hodgson WC, Ignjatovic V, Summerhayes R, Kochva E (2006) Early evolution of the venom system in lizards and snakes. Nature 439:584–588
4. Wagstaff SC, Favreau P, Cheneval O, Laing GD, Wilkinson MC, Miller RL, Stöcklin R, Harrison RA (2008) Molecular characterisation of endogenous snake venom metalloproteinase inhibitors. Biochem Biophys Res Commun 365:650–656
5. Munawar A, Trusch M, Georgieva D, Spencer P, Frochaux V, Harder S, Arni RK, Duhalov D, Genov N, Schlüter H, Betzel C (2011) Venom peptide analysis of *Vipera ammodytes meridionalis* (Viperinae) and *Bothrops jararacussu* (Crotalinae) demonstrates subfamily-specificity of the peptidome in the family Viperidae. Mol BioSyst 7:3298–3307
6. Tashima AK, Zelanis A, Kitano ES, Ianzer D, Melo RL, Rioli V, Sant'anna SS, Schenberg AC, Camargo AC, Serrano SM (2012) Peptidomics of three Bothrops snake venoms: insights into the molecular diversification of proteomes and peptidomes. Mol Cell Proteomics 11:1245–1262

7. Zelanis A, Tashima AK (2014) Unraveling snake venom complexity with 'omics' approaches: challenges and perspectives. Toxicon 87:131–134
8. Nicolau CA, Carvalho PC, Junqueira-de-Azevedo IL, Teixeira-Ferreira A, Junqueira M, Perales J, Neves-Ferreira AG, Valente RH (2016) An in-depth snake venom proteopeptidome characterization: benchmarking *Bothrops jararaca*. J Proteomics 151:211–231
9. Pimenta DC, Prezoto BC, Konno K, Melo RL, Furtado MF, Camargo AC, Serrano SM (2007) Mass spectrometric analysis of the individual variability of *Bothrops jararaca* venom peptide fraction. Evidence for sex-based variation among the bradykinin-potentiating peptides. Rapid Commun Mass Spectrom 21:1034–1042
10. Zelanis A, Tashima AK, Rocha MM, Furtado MF, Camargo AC, Ho PL, Serrano SM (2010) Analysis of the ontogenetic variation in the venom proteome/peptidome of Bothrops jararaca reveals different strategies to deal with prey. J Proteome Res 9:2278–2291
11. Zelanis A, Menezes MC, Kitano ES, Liberato T, Tashima AK, Pinto AF, Sherman NE, Ho PL, Fox JW, Serrano SM (2016) Proteomic identification of gender molecular markers in Bothrops jararaca venom. J Proteome 139:26–37
12. Tashima AK, Zelanis A (2016) Snake venom peptidomics. In: Gopalakrishnakone P, Calvete JJ (eds) Venom genomics and proteomics. Springer, Dordrecht, pp 317–331

# Chapter 24

# Identification of Peptides in Spider Venom Using Mass Spectrometry

**Rafael L. Lomazi, Erika S. Nishiduka, Pedro I. Silva Jr, and Alexandre K. Tashima**

## Abstract

Spider venoms are composed of hundreds of proteins and peptides. Several of these venom toxins are cysteine-rich peptides in the mass range of 3–9 kDa. Small peptides (<3 kDa) can be fully characterized by mass spectrometry analysis, while proteins are generally identified by the bottom-up approach in which proteins are first digested with trypsin to generate shorter peptides for MS/MS characterization. In general, it is sufficient for protein identification to sequence two or more peptides, but for venom peptidomics it is desirable to completely elucidate peptide sequences and the number of disulfide bonds in the molecules. In this chapter we describe a methodology to completely sequence and determine the number of disulfide bonds of spider venom peptides in the mass range of 3–9 kDa by multiple enzyme digestion, mass spectrometry of native and digested peptides, de novo analysis, and sequence overlap alignment.

**Key words** Spider venom, Peptidomics, Mass spectrometry, De novo analysis

## 1 Introduction

Peptides in the mass range of 3–9 kDa compose a large part of spider venom toxins [1–3]. In general, these toxins are cysteine-rich peptides containing the inhibitor cysteine knot [4], the disulfide-directed β-hairpin [5], or the Kunitz-type inhibitor domains [6]. It has been shown that these peptides are blockers of ion channels [7], inhibitors of proteases [8], antimicrobial agents [3, 9], and, in some cases, present very specific paralytic activities against insects [10, 11].

Heavy peptides are not efficiently fragmented by collision-induced dissociation with inert gases [3, 12, 13], giving limited information of primary structures. However, even without complete fragmentation, MS analysis of native and reduced peptides often gives important information about disulfide bond connections. Moreover, complementary MS/MS data of digested peptides help elucidate the amino acid sequence. In

Michael Schrader and Lloyd Fricker (eds.), *Peptidomics: Methods and Strategies*, Methods in Molecular Biology, vol. 1719, https://doi.org/10.1007/978-1-4939-7537-2_24, 

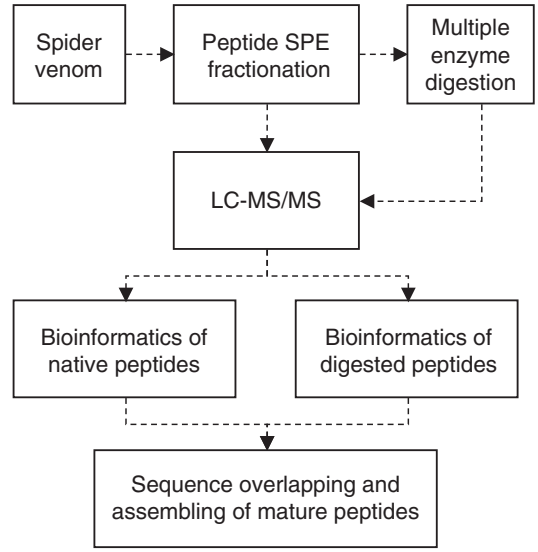



**Fig. 1** Scheme of the spider venom peptidomic strategy combining native and digested peptide analysis

this chapter, we describe a methodology combining mass spectrometry of native venom peptides, digestion of reduced and alkylated peptides with multiple enzymes for MS/MS analysis, de novo sequencing, database search, and assembling of the mature peptides based on overlapping of toxin fragments and on information from RNA-seq (Fig. 1). We show that with this methodology it is possible to obtain full sequences of venom toxins, determine the number of disulfide bonds, and identify potential structural isomers among venom toxins.

## 2 Materials

### *2.1 Quantification*

1. Bradford reagent.
2. 96-well polystyrene plate.
3. Microplate reader with 595 nm filter (we use Synergy HT, Biotek Instruments, Winooski, VT, USA).

### *2.2 Solid Phase Extraction*

1. C18 reversed-phase cartridges (we use Sep-pak Light, Waters, MA, USA).
2. Ultrapure water (resistivity at 25 °C: 18.2 MΩ•cm).
3. Methanol: HPLC or superior grade.
4. Acetonitrile (ACN) of HPLC or superior grade.
5. Trifluoroacetic acid (TFA) solutions: 0.1% TFA in water, and 0.1% TFA in 30% ACN (*see* **Note 1**).
6. Plastic syringes.

### 2.3 Digestion

1. Enzymes: trypsin, Glu-C, chymotrypsin, and thermolysin (*see* **Note 2**).
2. Dithiothreitol (DTT) solution, 100 mM.
3. Iodoacetamide (IAA) solution, 300 mM.
4. Ammonium bicarbonate solution (*see* **Note 3**), 50 mM.
5. Tris–HCl buffer, 100 mM, with 10 mM $CaCl_2$. Adjust pH to 7.4–8.0 with 1 M HCl.
6. Sodium phosphate buffer, 50 mM. Dissolve 60.0 mg of sodium phosphate monobasic ($NaH_2PO_4$) in 10 mL of water. Dissolve 71.0 mg of sodium phosphate dibasic ($Na_2HPO_4$) in 10 mL of water. Mix 2.26 mL of the $NaH_2PO_4$ solution with 7.74 mL of the $Na_2HPO_4$ solution to produce a 50 mM sodium phosphate buffer of pH 7.4.
7. Cleavable surfactant (we use Rapigest, Waters, MA, USA).

### 2.4 Mass Spectrometry

1. ACN for LC-MS.
2. Formic acid (FA): add 100 μL FA to 100 mL of water (phase A, 0.1% FA) or to 100 mL of ACN (phase B, 0.1% FA in ACN).
3. Reversed-phase trap column (*see* **Note 4**): we use nanoACQUITY UPLC Symmetry C18 Trap Column, 5 μm, 180 μm × 20 mm (Waters, MA, USA).
4. Analytical reversed-phase capillary column (*see* **Note 5**): we use nanoACQUITY CSH130 C18 1.7 μm, 75 μm × 200 mm.
5. Glu-fibrinopeptide B (GFP-B) ≥90% (sequence EGVNDNEEGFFSAR).
6. Picotip and Tapertip (New Objective) emitters.
7. Instrumentation for LC-MS experiments: we use a nanoAcquity UPLC system (Waters, MA, USA) coupled to a Q-TOF Synapt G2 HDMS (Waters, MA, USA) or an LTQ-Orbitrap Velos (Thermo Scientific, Bremen, GA, USA).

### 2.5 Bioinformatics

1. Software for retention time alignment, peak picking, and deisotoping peptide ions: we use Progenesis QI for Proteomics (Nonlinear Dynamics, Newcastle, UK).
2. Software for MS data interpretation, database, post-translational modification (PTM), and homology searches: we use PEAKS Studio 7.5 (Bioinformatics Solution Inc., Waterloo, Canada).

## 3 Methods

### 3.1 Venom Milking

1. Keep the spider in a small vented plastic box.
2. Place the box in a container filled with carbon dioxide for 2–10 min to anesthetize the animal (*see* **Notes 6** and **7**).

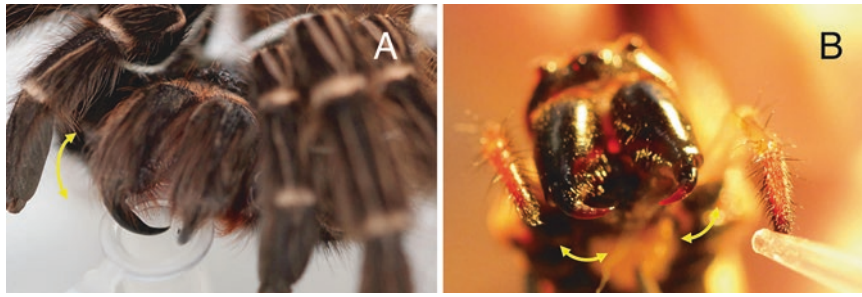



**Fig. 2** Cheliceras of orthognath and labdognath spiders. *Arrows* show movement directions. (**a**) *Acanthoscurria juruenicola*, an *orthognath* spider. (**b**) Nephilingis cruentata, a *labdognath* spider

3. After anesthetizing, hold the spider and apply mild electric stimulation with wet electrodes in each venom gland (10–25 V, 10 Hz) [14]. The extraction procedure depends on the chelicera position. *Orthognath* spiders (e.g., Mygalomorphae) bite with cheliceras in frontal movements while *labdognath* spiders (e.g., Araneomorphae) attack in lateral movements (Fig. 2). Cheliceras of *orthognath* spiders can be positioned in microtubes. Venoms of *labdognath* spiders have to be collected directly with pipettes (*see* **Note 8**).
4. Centrifuge venoms at 14,000 × *g* and 4 °C for 5 min, collect the supernatant, and store at −80 °C until use.

### 3.2 Quantification

1. Pipette 5 μL of venom sample in duplicate or triplicate and add 250 μL of the Bradford reagent in each well (*see* **Note 9**).
2. Prepare the standard curve with a pure protein as a reference (*see* **Note 10**).
3. Incubate plate (30 min, 37 °C) before reading.
4. Read absorbance at 595 nm.

### 3.3 Solid Phase Extraction (SPE)

1. Condition the C18 reversed-phase SPE cartridge with 3 mL of methanol.
2. Wash the cartridge with 3 mL of 0.1% TFA (*see* **Note 11**).
3. Dilute a volume containing ~100–1000 μg of venom protein in 500 μL of 0.1% TFA (*see* **Note 12**).
4. Load the venom solution in the SPE cartridge. Discard the excess volume.
5. Wash the cartridge with 3 mL of 0.1% TFA (*see* **Note 13**).
6. Elute the venom peptides with 3 mL of 30% ACN (*see* **Note 14**). Split the venom peptide eluate in five fractions, one for native peptide analysis and the others for digestions. Dry the peptide fractions in a vacuum concentrator (*see* **Note 15**).

### 3.4 Peptide Digestion

1. Dissolve four of the venom peptide fractions in 50 μL of digestion solutions: 50 mM $NH_4HCO_3$ for trypsin; 100 mM Tris-HCl, 10 mM $CaCl_2$ for chymotrypsin; 50 mM sodium phosphate buffer, 0.5 mM $CaCl_2$ for Glu-C and thermolysin.
2. Add 25 μL of 0.2% Rapigest or equivalent mass spectrometry compatible cleavable surfactant (*see* **Note 16**) to each tube and incubate the peptide solutions at 80 °C for 15 min.
3. Add 2.5 μL of DTT solution and incubate the peptide solutions at 60 °C for 30 min to reduce disulfide bonds.
4. Add 2.5 μL of IAA solution and incubate at room temperature in the dark for 30 min to alkylate cysteines.
5. Add each enzyme (separately) in the respective tube at enzyme:peptide ratios of 1:100, except for thermolysin, whose ratio is 1:250. Incubate the thermolysin tube at 75 °C and all other tubes at 37 °C for 15 min (*see* **Note 17**).

### 3.5 Mass Spectrometry

1. Calibrate the mass spectrometer with the GFP-B solution (100–500 fmol/μL) infused in the lockspray nano-source at 300–500 nL/min and 3.5 kV.
2. Dissolve the remaining dried venom peptide fraction from SPE in 25 μL of 0.1% FA and pipette in total recovery glass vials. For the digested peptides, filter solutions with 0.2 μm syringe filters before LC-MS/MS analysis to avoid clogging.
3. Load 5 μL of the peptide sample in the trap column using an autosampler (we use nanoAcquity UPLC) and wash with phase A at 8 μL/min for 5 min.
4. After loading and washing the peptide mixture in the trap column, switch the flow to the analytical capillary column in order to elute the peptides with a gradient of 7–35% of phase B in 90 min at 275 nL/min (*see* **Note 18**).
5. If a Q-TOF instrument equipped with ion mobility separation (IMS) is used for LC-MS/MS analysis, acquire data using the data-independent acquisition mode and IMS in the $m/z$ range of 50–2000 for precursors and fragments, with scan times of 1.25 s and in the resolution mode (resolution of ~20,000 at $m/z$ 785.8). Use collision energy of 4 eV for the precursor ions and a ramp of 17–60 eV for the fragment ions (*see* **Note 19**). For other instruments, acquire data in the data-dependent acquisition mode, scanning precursor ions in the $m/z$ range of 400–2000, followed by CID of the 3–10 most intense ions at the highest possible resolution. Use a dynamic exclusion time of 90 s (*see* **Note 20**).

### 3.6 Bioinformatics of Native Peptides

1. Load the MS raw data of native peptides in an appropriate proteomics software (we use Progenesis QI for Proteomics, Nonlinear Dynamics, Newcastle, UK).

2. Perform automatic peak picking and retention time alignment.
3. After analysis, manually check the isotopic peak and charge assignments of the peptide ions (*see* **Note 21**).
4. Export the final peptide ion worksheet with information of *m/z*, mass, retention time, and intensities to produce mass maps and compare native peptide masses with the overlapping sequences from digested peptide identifications.
5. Check retention times of the peptide ions. Multiple chromatographic peaks for the same ion may indicate the existence of conformational isomers.

### 3.7 Bioinformatics of Digested Peptides

1. Convert the MS raw data to the .mzML, .pkl, or .mgf formats.
2. Load the converted files to an MS data interpretation software (such as PEAKS Studio 7.5 or equivalent program), separating the datasets by type of enzyme.
3. Run automated de novo analysis using tolerances of 10 ppm and 0.025 Da to precursor and fragment mass, respectively.
4. Run database search of all files simultaneously (*see* **Note 22**) using the same mass tolerances of de novo analysis and set Cys carbamidomethylation as fixed modification and Met oxidation, protein N-terminal acetylation and deamidation of Asn/Gln as variable modifications. Set the enzymes as "specified for each sample," allow nonspecific cleavages at one end of the peptide and up to three missed cleavages (*see* **Note 23**).
5. From the overlapping alignments, calculate the masses of the continuous sequences and compare with the native peptide masses to match possible mature peptides (*see* **Note 24**).
6. After database search, run PTM and homology searches using PEAKS Studio 7.5 or equivalent software to find more PTM and possible sequence mutations.

## 4 Notes

1. Be careful when pipetting TFA, which is a volatile and corrosive acid. Prepare the TFA solutions in an exhaust hood.
2. Use mass spectrometry or proteomics grade enzymes. In addition to the enzymes described in the above protocol, other enzymes may also be used for digestions. The aim is to obtain peptides cleaved at different amino acids to assemble overlapping sequences.
3. Always prepare fresh solutions for digestions. This is valid for all digestions, but especially important for those using volatile salts such as ammonium bicarbonate.

4. The trap column is used to quickly load the peptides in the chromatographic system and also to desalt the peptide mixture online.
5. Other columns may be used for chromatographic separation. If your system can withstand high pressures associated with columns with smaller internal diameters, the smaller diameter columns generally allow lower flow rates and higher sensitivities.
6. To improve extraction yield, it is recommended to keep the spider in captivity without feeding for 5–7 days before venom extraction. Spiders can stay long periods without food.
7. The container with carbon dioxide to anesthetize the spider can be prepared with a beaker of water and dry ice in a deep box. The time to anesthetize the spider may be adjusted and depend on the animal size.
8. It is important to observe that the venom is not being contaminated with other fluids that may be expelled during the venom extraction process. For *orthognaths*, cheliceras can be positioned in the tubes; however, it is not possible for *labdognaths* due to their positions. In this case, it is convenient during extraction to observe the venom drops with a stereomicroscope to ensure correct collection.
9. Spider venom protein concentrations may vary from about 10–30 μg/μL. Dilutions of 10–40-fold may be necessary to attain the linear range of most colorimetric methods. Other reagents for protein quantification as the bicinchoninic acid can also be used.
10. Several proteins can be used as standards in this step for quantification, as crude venoms contain peptides and proteins. It is for an estimate of venom amounts to be used as starting material for peptide fractionation.
11. It is important to remove the organic solvent from the SPE cartridges; otherwise the venom peptides will not adsorb to the C18 chains.
12. The volume to obtain 100 μg of venom proteins may be as low as 3 μL. The dilution is a way to have sufficient solution volume to load the protein/peptide mixture in the SPE cartridges. Higher amounts of venom may be used depending on the LC-MS setup and on the sensitivity of the mass spectrometer used for characterization.
13. At this point, salts and small polar molecules can be separated from the venom peptides.
14. The 30% ACN solution is usually sufficient to elute peptides in the mass range of 3–9 kDa.
15. The remaining proteins from the SPE cartridge can be eluted with 80% ACN solution. The protein and peptide fractions

may be dried and stored for several months at −80 °C for further analysis.

16. This step is optional, but a surfactant may be useful to denaturate peptides and expose disulfide bond sites. Some peptides are protease inhibitors and denaturation may be necessary to assist enzymatic cleavage.
17. These short time digestions result in several fragments with missed cleavages and they are useful for the final purpose of sequence overlapping.
18. Switch of column flows after loading peptides in the trap column is usually performed automatically by the chromatographic system through divert valves and are setup in the chromatography method. The gradient time and flow rate may be changed depending on the capillary column internal diameter and on the sample complexity.
19. In our experience, the mass spectrometer Synapt G2 (Waters) presents better acquisition performance for spider venom peptidomics when working in the data-independent acquisition mode and with IMS.
20. Other mass spectrometers and acquisition modes may be used for these spider venom peptidomics experiments. However, it is important to use high resolution (>10,000) to analyze the native peptide samples and to resolve the isotopic patterns of the highly charged peptides, frequently in the range of +4 to +8.
21. Progenesis QI for Proteomics matches appropriately the charges and monoisotopic peaks of tryptic peptides. However, some heavy and highly charged nontryptic peptides may need manual inspection and validation due to the increasing difficulty in resolving their isotopic patterns and charge assignments.
22. With simultaneous database searches, peptides generated from different enzymes are identified in the same run and can be overlapped to examine sequence coverage. This is a key step of the method. When several peptide-spectrum matches from different enzymes cover a restricted and continuous region of the protein sequence, it indicates a possible mature peptide full coverage.
23. We recommend the preparation of a venom gland transcriptomic database to support peptide-spectrum matches. If a transcriptome is not available for the species of spider being studied, compile a database with sequences from species of the closest genus. In case a high sequence coverage is not attained, manual overlap of the de novo fragments may be tried, considering the native masses as guides to mature peptide sequences.
24. To perform the calculations, simulate Cys-Cys bonds for each pair of Cys identified to compare with the native peptide masses. Commonly, three to five disulfide bonds are found in these spider venom peptides.

## Acknowledgments

This work was supported by grants from Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP, 2012/19321-9 and 2016/03839-0 to A. K. T.), Financiadora de Estudos e Projetos (FINEP) and Master's degree fellowships from FAPESP (2014/17140-2) to R. L. L., and from Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) to E. S. N.

### References

1. Escoubas P, Sollod B, King GF (2006) Venom landscapes: mining the complexity of spider venoms via a combined cDNA and mass spectrometric approach. Toxicon 47:650–663
2. Palagi A, Koh JMS, Leblanc M et al (2013) Unravelling the complex venom landscapes of lethal Australian funnel-web spiders (Hexathelidae: Atracinae) using LC-MALDI-TOF mass spectrometry. J Proteome 80:292–310
3. Abreu TF, Sumitomo BN, Nishiyama MY et al (2017) Peptidomics of Acanthoscurria gomesiana spider venom reveals new toxins with potential antimicrobial activity. J Proteome 151:232–242
4. Norton RS, Pallaghy PK (1998) The cystine knot structure of ion channel toxins and related polypeptides. Toxicon 36:1573–1583
5. Wang X, Connor M, Smith R et al (2000) Discovery and characterization of a family of insecticidal neurotoxins with a rare vicinal disulfide bridge. Nat Struct Biol 7:505–513
6. Peng K, Lin Y, Liang SP (2006) Nuclear magnetic resonance studies on Huwentoxin-XI from the Chinese bird spider Ornithoctonus huwena: 15N labeling and sequence-specific 1H, 15N nuclear magnetic resonance assignments. Acta Biochim Biophys Sin Shanghai 38:457–466
7. Rates B, Prates MV, Verano-Braga T et al (2013) μ-Theraphotoxin-An1a: primary structure determination and assessment of the pharmacological activity of a promiscuous anti-insect toxin from the venom of the tarantula Acanthoscurria natalensis (Mygalomorphae, Theraphosidae). Toxicon 70:123–134
8. Wan H, Lee KS, Kim BY et al (2013) A spider-derived Kunitz-type serine protease inhibitor that acts as a plasmin inhibitor and an elastase inhibitor. PLoS One 8:e53343
9. Ayroza G, Ferreira ILC, Sayegh RSR et al (2012) Juruin: an antifungal peptide from the venom of the Amazonian Pink Toe spider, Avicularia juruensis, which contains the inhibitory cystine knot motif. Front Microbiol 3:324
10. Windley MJ, Herzig V, Dziemborowicz SA et al (2012) Spider-venom peptides as bioinsecticides. Toxins (Basel) 4: 191–227
11. Mourão CB, Heghinian MD, Barbosa EA et al (2013) Characterization of a novel peptide toxin from Acanthoscurria paulensis spider venom: a distinct cysteine assignment to the HWTX-II family. Biochemistry 52:2440–2452
12. Tashima AK, Zelanis A, Kitano ES et al (2012) Peptidomics of three bothrops snake venoms: insights into the molecular diversification of proteomes and peptidomes. Mol Cell Proteomics 11:1245–1262
13. Sasaki K, Osaki T, Minamino N (2013) Large-scale identification of endogenous secretory peptides using electron transfer dissociation mass spectrometry. Mol Cell Proteomics 12: 700–709
14. Rocha-e-Silva TAA, Sutti R, Hyslop S (2009) Milking and partial characterization of venom from the Brazilian spider Vitalius dubius (Theraphosidae). Toxicon 53:153–161

# Chapter 25

# Single Cell Peptidomics: Approach for Peptide Identification by N-Terminal Peptide Derivatization

Susanne Neupert

## Abstract

In recent years, single cell microanalysis techniques have moved into the center stage to study fundamental intracellular interactions and cell–cell communication events, and have led to a better understanding of physiological processes and behavioral patterns. The availability of more sensitive, robust, and precise mass spectrometers improved the detection and characterization of putative neuroactive substances from individual cells. For sequence characterization, particularly when working with samples as small as a single cell, the most crucial step to obtain usable data is sample preparation. For some studies, genetic or molecular data are not available to confirm an amino acid sequence of a putative neuropeptide, and it is necessary to sequence the peptide from the mass spectrometry analysis alone (i.e., de novo sequencing). In this chapter, a protocol is described for de novo sequencing of neuropeptides from individual single cells by N-terminal derivatization using 4-sulfophenyl isothiocyanate and subsequent mass spectrometric analysis.

**Key words** Single cell peptidomics, N-Terminal derivatization for de novo sequencing, 4-Sulfophenyl isothiocyanate (SPITC), Neuropeptides, MALDI-TOF/TOF

## 1 Introduction

Complex structures such as tissues and organs are made of sometimes a few hundred up to millions or billions of cells, depending on the organism and tissue. Different cell types are involved in a variety of processes, including development, embryogenesis, regeneration, and the organization of neuronal networks. To communicate and interact with each other, many cells synthesize, store, and secrete biomolecules. Microchemical analysis of these compounds from individual cells is becoming increasingly important to achieve a detailed understanding of cell function in biomedical, pharmacological, and neurophysiological research [1, 2]. A successful performance of biomolecule characterization at the single-cell level depends on many different conditions: (1) a clear cell identification among thousands to millions of similar cells; (2) contamination-free cell isolation; (3)

Michael Schrader and Lloyd Fricker (eds.), *Peptidomics: Methods and Strategies*, Methods in Molecular Biology, vol. 1719, https://doi.org/10.1007/978-1-4939-7537-2_25, 

an optimized sample preparation protocol for subsequent biochemical analysis; and (4) the availability of technologies that are capable of analyzing micrometer-sized structures and compounds at the attomole level. Commercially available instruments for biomolecule characterization are very sensitive and have high throughput, reproducibility and mass accuracy. While the currently available equipment is capable of single-cell sequence determinations, more sensitive and precise next generation sequencing technologies are being developed and will likely become indispensable for single-cell analyses.

This chapter describes an optimized sample preparation, which is the key issue for producing high quality and unambiguous data. Our method of choice to study the peptidergic content of neurons is matrix-assisted laser desorption/ionization tandem time of flight mass spectrometry (MALDI-TOF/TOF MS) [2, 3]. However, for de novo sequencing of putative neuropeptides, direct single cell analysis does not always provide sufficient information for determination of the complete amino acid sequence. Furthermore, the identification of the complete primary sequence of a peptide is often prejudiced by differences in fragmentation efficiencies and by the amino acid sequence of a peptide. In 2003, an effective N-terminal derivatization method using 4-sulfophenyl isothiocyanate (SPITC) was published which enhances peptide sequencing by post-source decay MALDI-TOF mass spectrometry [4]. Only N-terminally unblocked peptides react with SPITC and are detected in MS mode with an increase of 215 Da due to the SPITC group. Furthermore, the resulting peptides carry a fixed negative charge at the N-terminal end and the resulting $MS^2$ spectrum is dominated by C-terminal *y*-type ions. For many peptides, this allows the complete sequence to be determined either manually or with simple software tools such as MASCOT [4]. In this chapter, the N-terminal derivatization method using SPITC is described along with an approach to optimize de novo sequencing of neuropeptides from single cell preparation. Such approach was successfully used for neuropeptide sequence characterization from single capa neurons of the cockroach *Periplaneta americana* [5].

## 2 Materials

Use highly purified water for all solutions. We use a water purification system from TKA CANDOR Bioscience GmbH (Wangen, Germany) with a polisher module (0.059 μS/cm), continuous UV-disinfection (254 nm), and an inline sterile filtration unit (0.2 μm) at 24.5 °C. Use high purity organic solvents for HPLC (≥99.9% purity) with analytical grade reagents. Prepare and store all reagents at room temperature unless indicated otherwise. Follow appropriate safety instructions when using chemicals.

### 2.1 Preparation

1. Dissecting saline: (in mM) NaCl 126, KCl 5.4, $NaH_2PO_4$ 0.17 and $KH_2PO_4$ 0.22, with pH adjusted to 7.4 [6] using 0.1 M HCl.
2. Fine forceps (e.g., sharpened Dumont No. 5, Fine Science Tools GmbH, Heidelberg, Germany).
3. Ultrafine spring or clipper scissors (such as from Fine Science Tools GmbH, Heidelberg, Germany).
4. Tungsten micro needles (such as from Fine Science Tools GmbH, Heidelberg, Germany).
5. Uncoated glass capillaries (e.g., Hilgenberg GmbH) fitted to a tube with mouthpiece (e.g., a sterile pipette tip) for cell dissection and cell extraction.
6. Sylgard-coated black glass preparation dish.
7. Fluorescence dye for cell labeling: 10% dextran, tetramethylrhodamine, 3000 MW, anionic, lysine fixable (we use Molecular Probes, Leiden; Netherlands) in saline, storing and using at 4 °C.
8. Home-made two chamber device for backfilling.
9. Petroleum jelly (commonly referred to by the brand name Vaseline, and which can be purchased from a grocery store or pharmacy). Use petroleum jelly without addition of colorant, preservative, fragrance, aromas, or any other additives.
10. Stereo fluorescence dissecting microscope with high magnification (e.g., SteREO Lumar.V12, Carl Zeiss AG, Goettingen, Germany) equipped with an Ex 550/25 filter or similar filter set for dextran-tetramethylrhodamine (Excitation: 555 nm; Emission: 580 nm) detection and a AxioCam MrC camera set or similar equipment.

### 2.2 Peptide Derivatization Using 4-Sulfophenyl Isothiocyanate (SPITC)

1. 4-Sulfophenyl isothiocyanate (SPITC) solution, 1.5 mg/mL.
2. Acetic acid (AA), 10% and 0.5% in purified water, respectively.
3. Sodium bicarbonate ($NaHCO_3$) solution in purified water, 20 mM, with a pH of 9.0.
4. Thermomixer compact (we use Eppendorf, Wesseling-Berzdorf, Germany).
5. Mini Centrifuge (we use Tomy PMC-060 Capsulefuge).
6. Microfuge tubes (we use 0.5 μL Eppendorf tubes, Wesseling-Berzdorf, Germany.

### 2.3 StageTip

1. Reversed-phase C18-StageTips (we use Empore 3 M, IVA Analysetechnik e.K., Meerbusch, Germany).
2. 1–10 μL gel-loader tips (we use Eppendorf, Wesseling-Berzdorf, Germany).
3. 100% methanol (MeOH).
4. Solution A: 0.5% AA.

5. Solution B: 80% acetonitrile (ACN) with 0.5% AA.
6. 10%, 20%, 30%, 40%, and 50% solutions of ACN in water with 0.5% AA.
7. 80% ACN 20% isopropanol.
8. 1.25 mL syringe (we use Eppendorf, Wesseling-Berzdorf, Germany).

### 2.4 MALDI-TOF Matrix Application

1. Re-crystallized α-cyano-4-hydroxycinnamic acid (CHCA).
2. 2,5-Dihydroxybenzoic acid (DHB).
3. 20% ACN with 1% formic acid (FA).
4. 50% methanol (MeOH) in water.
5. 60% ethanol (EtOH), 36% ACN, 4% water.

### 2.5 MALDI-TOF Mass Spectrometry

1. MALDI target plate, we use MTP 384 ground steel BC fitted in a MTP target frame III or home-made stainless steel 100 spot MALDI target fitted in a MTP Adapter (BRUKER Daltonics, Bremen, Germany).
2. MALDI-TOF/TOF mass spectrometer (ultrafleXtrem, Bruker Daltonics, Bremen, Germany).

### 2.6 Software

1. MS software (we use flexAnalysis 3.4, Bruker Daltonics, Bremen, Germany).
2. Microscopic software package (we use AxioVision Rel. 4.8, Carl Zeiss AG, Goettingen, Germany).
3. Regular pocket calculator.

### 2.7 Internet Sources

1. Proteomics tools for mining sequence databases in conjunction with mass spectrometry experiments (we use ProteinProspector, http://prospector.ucsf.edu).
2. Tool for searching and comparing primary biological sequence information, such as amino acid sequences (we use NCBI/BLAST, http://blast.ncbi.nlm.nih.gov/Blast.cgi).

## 3 Methods

### 3.1 Tissue Dissection

Dissect insect ganglion containing cells of interest in insect saline. Remove the connective tissue (muscles or fat body) carefully without damaging the ganglion. If retrograde labeling has to be performed for cell identification, attached nerves should be unscathed and severed as distal as possible (*see* **Note 1**).

### 3.2 Cell Identification

Depending on the cell type, there are several approaches for cell identification including cell size and localization, active and passive dye injection, Tyndall effect, and UAS-GAL4-system.

### 3.2.1 Cell Size and Localization

For larger cells (between 40 and 500 μm) which are characterized by an unambiguously tractable localization in a ganglion or tissue sample, these parameters alone are often sufficient for identification and additional tracing, biochemical, or genetic methods are not necessary.

### 3.2.2 Dye Injection

Cells can be labeled by retrograde labeling, intracellular dye application, whole cell patch clamp, or perforated patch clamp.

#### Retrograde Labeling of Neurons by External Nerves

Place the ganglion into a chamber filled with insect saline (4 °C) and transfer the attached nerve that contains the axon of cells of interest in a second nearby chamber filled with 10% dextran-tetramethylrhodamine using a needle fixed in a needle holder. Cover the nerve between the two chambers with petroleum jelly (*see* **Note 2**). Staining of cell bodies via passive diffusion into the ganglion requires 5–48 h (Fig. 1a; *see* **Note 3**).

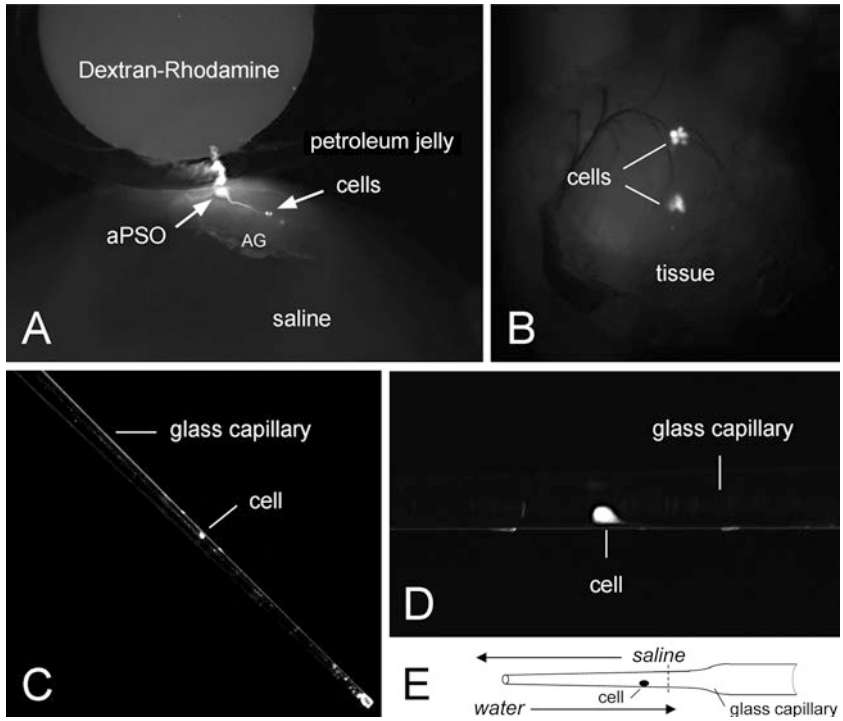



**Fig. 1** Cell dissection for SPITC derivatization and subsequent MALDI-TOF MS analysis. (**a**) Retrograde filling of neurons via external nerve, which contain the respective neurites. Tissue with cell bodies of interest are placed in chamber filled with saline. An external nerve is placed in a separate chamber containing fluorescence dye. The part of the external nerve between the chambers is covered with petroleum jelly (i.e., Vaseline). Labeled cells are detectable in the ganglion using a Stereo-Lumar V12 fluorescence microscope. (**b**) Overview of stained *capa* somata in the abdominal ganglion of the cockroach *Periplaneta americana*. (**c**) Dissect a labeled cell using an uncoated glass capillary. (**d**) After removing of saline labeled cell is attached on the inner side of the glass capillary. (**e**) Schematic drawing of removing saline and adding purified water for peptide extraction. Water should cover the cell completely (dashed line). AG, abdominal ganglion; aPSO, abdominal perisympathetic organs

*Retrograde Labeling of Neurons by Dye Injection into Axonal Projection Areas*

Use a glass capillary connected to a tube and a mouth piece. Fill the sharp tip with a florescence dye and inject the dye into the projection area of the cell of interest by treating the area with the sharp tip several times under continuous dye application (*see* [7], **Note 4**).

*Intracellular Dye Application*

Insert a microelectrode filled with a dye into an individual cell body to measure electrical activity while injecting the dye into usually a neuron.

*Whole Cell Patch Clamp*

Place a patch clamp electrode which is filled with a fluorescence dye on the membrane of the cell of interest and load the dye by application of a hyperpolarizing current. Injection time and current settings are depending on the cell type as well as the dye (*see* [8]).

*Perforated Patch Clamp*

Fill the neuron juxtasomal with a dye by giving electroporating stimuli via the patch pipette (*see* for more details [9]).

#### *3.2.3 UAS-GAL4-System*

For insects, such as *Drosophila melanogaster* or *Tribolium castaneum*, different GAL-4 driver lines are available to identify specific cells by expression of fluorescent marker proteins such as green fluorescence protein (GFP) under upstream activating sequence (UAS) control [10, 11].

#### *3.2.4 Tyndall Effect*

Use this phenomenon to identify neurosecretory cells in the nervous tissue based on their slightly bluish color. In transmitted light the neurons appear blue because of the light scattering due to peptide vesicles that they contain. The intensity of the color depends on the concentration of these dense core vesicles. Examples of neurons with distinct Tyndall effect are PBAN-expressing neurons in the subesophageal ganglion of moths [12] and neurons of the *pars intercerebralis* in many fly species.

### *3.3 Cell Dissection and Peptide Extraction*

1. Fix the ganglia with microneedles, search for the labeled neurons under a stereo fluorescence microscope (cell identification methods *see* Subheadings 3.2.1, 3.2.2, and 3.2.3; Fig. 1b) or a stereo microscope without fluorescence option (cell identification method *see* Subheading 3.2.4) and remove the ganglionic sheath close to the labeled cells using a fine scissor or ultrafine forceps.
2. Without prior enzyme treatment, remove a labeled cell manually using an uncoated glass capillary fitted to a tube with a mouthpiece. Let the cell settle and allows it to attach to the inner side of the glass capillary (Fig. 1c, d, *see* **Note 5**). Remove the saline complete under constantly visual control (*see* **Note 6**). Extract the cell by adding purified water into the glass capillary (Fig. 1e). Incubate the sample on ice for 15–30 min in a humid chamber (*see* **Note 7**).

3. For documentation, take photographs of the ganglia before and after cell dissection using the stereo microscope equipped with appropriate camera set each time a single cell is removed for reconstruction of the cell localization.

### 3.4 Peptide Derivatization Using 4-Sulfophenyl Isothiocyanate (SPITC)

1. Transfer the single cell extract in 40 µL SPITC solution dissolved in 20 mM $NaHCO_3$ (pH 9.0) at a concentration of 1.5 mg/mL. The sulfonation reaction should be carried out in a 0.5 mL microfuge tube for 1 h at 55 °C in the Thermomixer at 300 rpm.
2. Add 2.5 µL 10% AA followed by 50 µL 0.5% AA to acidify [4, 13]).
3. Activate and equilibrate a home-made StageTip Empore 3 M C18 column (*see* [14], **Note 8**). Rinse the column with 50 µL 0.5% AA without drying the C18 material. Load sample solution on the column by using a mini centrifuge. Elute peptides from the column by adding 1.5 µL 10/20/30/40/50/80% ACN 0.5% AA sequentially using a 1.25 mL syringe which is fitting on the StageTip.
4. Drop approximately 1.4 µL of each elute onto a MALDI target. For a better separation distribute each of the 1.4 µL sample solutions on three spots (approximately 0.4 µL per spot). Apply 0.3 µL matrix solution to each sample spot before drying using a 0.1–10 µL nanopipette.

### 3.5 Matrix Composition and Application

1. Prepare a stock solution of 10 mg/mL CHCA dissolved in 60% EtOH, 36% ACN, 4% purified water. For MALDI-TOF MS analysis, dilute one part of the CHCA stock solution with three parts 50% methanol/water.
2. Alternatively, use 10 mg/mL DHB dissolved in 20% ACN and 1% FA (*see* **Note 9** and Fig. 2).
3. To remove free matrix crystals, rinse the dried spot with purified water or 0.1% TFA and remove it after a few seconds by cellulose paper (*see* **Note 10**).
4. Place an appropriate peptide standard close to the sample spots on the sample plate (*see* **Note 11**).

### 3.6 MALDI-TOF Analysis

1. Load the sample plate into a MALDI-TOF mass spectrometer and use the synthetic peptide standard for calibration and optimization of the settings of the mass spectrometer.
2. Analyze the sample in the reflector positive mode as recommended for peptide samples. Acquire mass fingerprint spectra with a sufficient signal-to-noise ratio by using an appropriate laser power and a minimum of laser shots for each spectrum. Remember that you will use the same sample spot for subsequent fragmentation experiments (*see* **Note 12**).

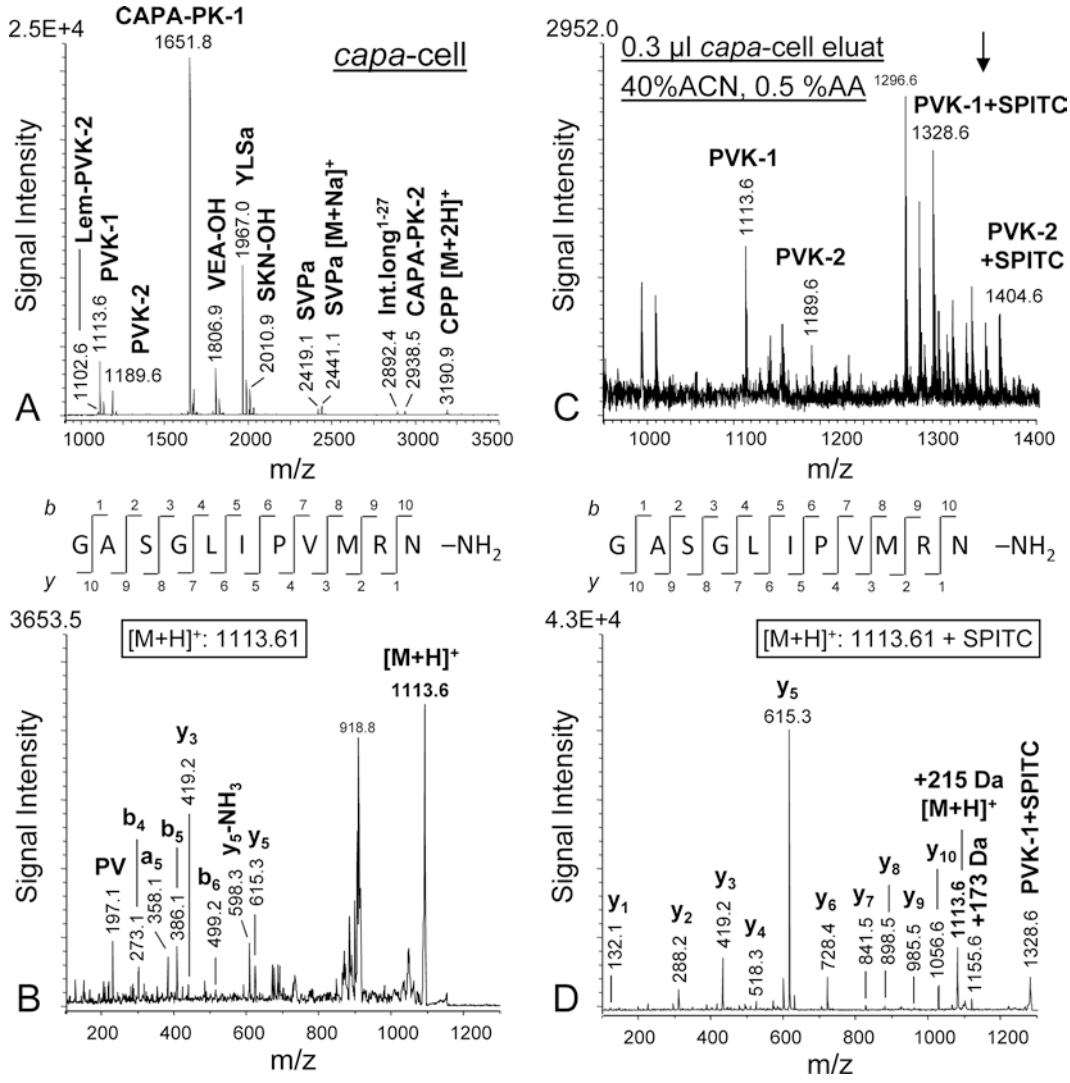



**Fig. 2** Direct single cell MALDI-TOF MS and peptide characterization from derivatized and non-derivatized putative neuropeptides. (**a**) Mass spectrum obtained by direct profiling of a single *capa* cell in a mass range of 900–3500 Da. Marked ion signals represent product from two *capa* cDNA transcripts (*see* also [5]). (**b**) Mass spectrum of the ion signal at *m/z* 1113.6. Resulting b and y fragments are labeled which contain enough information for sequence confirmation but not for de novo sequencing. (**c**) 0.3 µL of 40% ACN 0.5% AA eluate of a single cell extraction treated by SPITC. The non-derivatized ion at *m/z* 1113.6 and the associated derivatized ion at *m/z* 1328.6 are detectable. (**d**) Fragmentation mass spectrum of the ion *m/z* 1328.6. Derivatization induces neutral losses of the modifying group (–173 Da, –215 Da) and almost exclusive formation of y-series fragment ions for de novo sequencing

3. For fragmentation, change the settings on the mass spectrometer as recommended for peptide fragmentation, select a derivatized peptide ion signal of interest (*see* **Note 13**) and start fragment analysis (*see* **Note 14**).

## 4 Notes

1. Be careful not to damage the nerve attached to the ganglion, otherwise backfilling will not work effectively.
2. Most of the fluorescence dyes used for retrograde labeling do not suppress ion signals during MALDI-TOF MS analysis. Do not use Lucifer Yellow which can interfere with MALDI-TOF.
3. The amount of time for filling a cell body with a dye depends on the distance between dye application in the axon and the location of the corresponding cell body of interest within the tissue.
4. If cell bodies become damaged during the dissection procedure, the fluorescence signal of the labeled cells disappears and the subsequent mass spectra usually have lower signal intensities even if the cell is properly placed on the inner side of the glass capillary.
5. To avoid contamination, use a new glass capillary for each cell preparation even if a cell does not attach in the capillary or an already dried cell could not be visualized before water application for peptide extraction.
6. The fluorescence marker should be visible. If not, do not use the sample for subsequent analysis.
7. Use double faced adhesive tape to fix the glass capillary precisely in a Petri dish. Be careful not to damage the tip of the capillary. Place into the Petri dish moist cellulose or tissue paper; not wet. Put a lid on the Petri dish and incubate the samples in a refrigerator at 8 °C.
8. Fill 0.4–0.5 mg StageTip Empore 3 M C18 material in a 20 μL Gel-loader Tip using an appropriate small metal stick. Cut off the tip in an appropriate distance to the C18 material. To hold the home-made C18 column during centrifugation steps use a closed 2 mL microfuge tube with a hole in the lid.
9. For an even distribution of DHB matrix crystals, dry samples using a regular hot air hairdryer.
10. Include rinsing step after matrix application only if using CHCA as matrix. Water can be removed after rinsing by cellulose paper. Do not rinse DHB samples. DHB is recrystallized using water or 0.1% TFA and will be removed from the spot by cellulose paper.
11. Analyze synthetic peptides using the same settings as expected for the cell samples (use very low peptide concentrations). The single cell preparations do not contain enough material for testing and selecting the optimal conditions. A standard peptide mixture is important not only for an optimal calibration of the mass spectrometer but also for the tuning of the hardware settings. Mixtures are commercially available or can be individually prepared.

12. Start analyses with relatively low laser energy and few laser shots to save material for subsequent fragmentation experiments.
13. Depending on the derivate, the parent mass of your ion of interest is changed. For SPITC derivatization ions are detected with a mass shift of +215 Da.
14. For fragmentation experiments of derivatized ions, the laser power has to be reduced at 20–30% (settings adjustable in the software of the mass spectrometer) comparable to the laser power which is necessary for fragmentations of non-derivatized ions.

## References


1. Rubakhin SS, Lanni EJ, Sweedler JV (2013) Progress toward single cell metabolomics. Curr Opin Biotechnol 24:95–104. Review
2. Romanova EV, Sweedler JV (2015) Peptidomics for the discovery and characterization of neuropeptides and hormones. Trends Pharmacol Sci 36:579–586. Review
3. Hummon AB, Amare A, Sweedler JV (2006) Discovering new invertebrate neuropeptides using mass spectrometry. Mass Spectrom Rev 25:77–98
4. Marekov LN, Steinert PM (2003) Charge derivatization by 4-sulfophenyl isothiocyanate enhances peptide sequencing by post-source decay matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. J Mass Spectrom 38:373–377
5. Neupert S, Derst C, Sturm S, Predel R (2014) Identification of two *capa* cDNA transcripts and detailed peptidomic characterization of their peptide products in *Periplaneta americana*. EuPA Open Proteom 3:195–205
6. Su H, O'Dowd DK (2003) Fast synaptic currents in drosophila mushroom body Kenyon cells are mediated by alpha-bungarotoxin-sensitive nicotinic acetylcholine receptors and picrotoxin-sensitive GABA receptors. J Neurosci 23:9246–9253
7. Reischig T, Stengl M (2002) Optic lobe commissures in a three-dimensional brain model of the cockroach *Leucophaea maderae*: a search for the circadian coupling pathways. J Comp Neurol 443:388–400
8. Fusca D, Schachtner J, Kloppenburg P (2015) Colocalization of allatotropin and tachykinin-related peptides with classical transmitters in physiologically distinct subtypes of olfactory local interneurons in the cockroach (*Periplaneta americana*). J Comp Neurol 443:388–400
9. Linley JE (2013) Perforated whole-cell patch-clamp recording. Methods Mol Biol 998:149–157
10. Neupert S, Johard HA, Nassel DR, Predel R (2007) Single-cell peptidomics of *Drosophila melanogaster* neurons identified by Gal4-driven fluorescence. Anal Chem 79:3690–3694
11. Brand A (1999) GFP as a cell and developmental marker in the drosophila nervous system. Methods Cell Biol 58:165–181
12. Neupert S, Huetteroth W, Schachtner J, Predel R (2009) Conservation of the function counts: homologous neurons express sequence-related neuropeptides that originate from different genes. J Neurochem 111:757–765
13. Wang D, Kalb SR, Cotter RJ (2004) Improved procedures for N-terminal sulfonation of peptides for matrix-assisted laser desorption/ionization post-source decay peptide sequencing. Rapid Commun Mass Spectrom 18:96–102
14. Rappsilber J, Mann M, Ishihama Y (2007) Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. Nat Protoc 2:1896–1906

# Chapter 26

# Peptidomic Identification of Cysteine-Rich Peptides from Plants

**Xinya Hemu, Aida Serra, Dina A. Darwis, Tobias Cornvik, Siu Kwan Sze, and James P. Tam**

## Abstract

Plant cysteine-rich peptides (CRPs) constitute a majority of plant-derived peptides with high molecular diversity. This protocol describes a rapid and efficient peptidomic approach to identify a whole spectrum of CRPs in a plant extract and decipher their molecular diversity and bioprocessing mechanism. Cyclotides from *C. ternatea* are used as the model CRPs to demonstrate our methodology. Cyclotides exist naturally in both cyclic and linear forms, although the linear forms (acyclotide) are generally present at much lower concentrations. Both cyclotides and acyclotides require linearization of their backbone prior to fragmentation and sequencing. A novel and practical three-step chemoenzymatic treatment was developed to linearize and distinguish both forms: (1) N-terminal acetylation that pre-labels the acyclotides; (2) conversion of Cys into pseudo-Lys through aziridine-mediated S-alkylation to reduce disulfide bonds and to increase the net charge of peptides; and (3) opening of cyclic backbones by the novel asparaginyl endopeptidase butelase 2 that cleaves at the native bioprocessing site. The treated peptides are subsequently analyzed by liquid chromatography coupled to mass spectrometry using electron transfer dissociation fragmentation and sequences are identified by matching the MS/MS spectra directly with the transcriptomic database.

**Key words** Cysteine-rich peptides, Plant peptides, Peptidomics, Molecular diversity, Butelase, Cyclotides

## 1 Introduction

The limitations to develop bioactive peptides into drugs are their structural and poor metabolic stability. Cysteine-rich peptides, such as plant defensins, heveins, and knottins, are highly recalcitrant structures stabilized by multiple intramolecular disulfide bonds to serve as a source of diverse, druggable compounds [1, 2]. Cyclotides are end-to-end macrocyclic CRPs without loose ends and therefore possess additional molecular stability against exopeptidase degradation [3]. Hence, they are useful leads and scaffolds for the

Xinya Hemu and Aida Serra contributed equally to this work.

Michael Schrader and Lloyd Fricker (eds.), *Peptidomics: Methods and Strategies*, Methods in Molecular Biology, vol. 1719, https://doi.org/10.1007/978-1-4939-7537-2_26, 

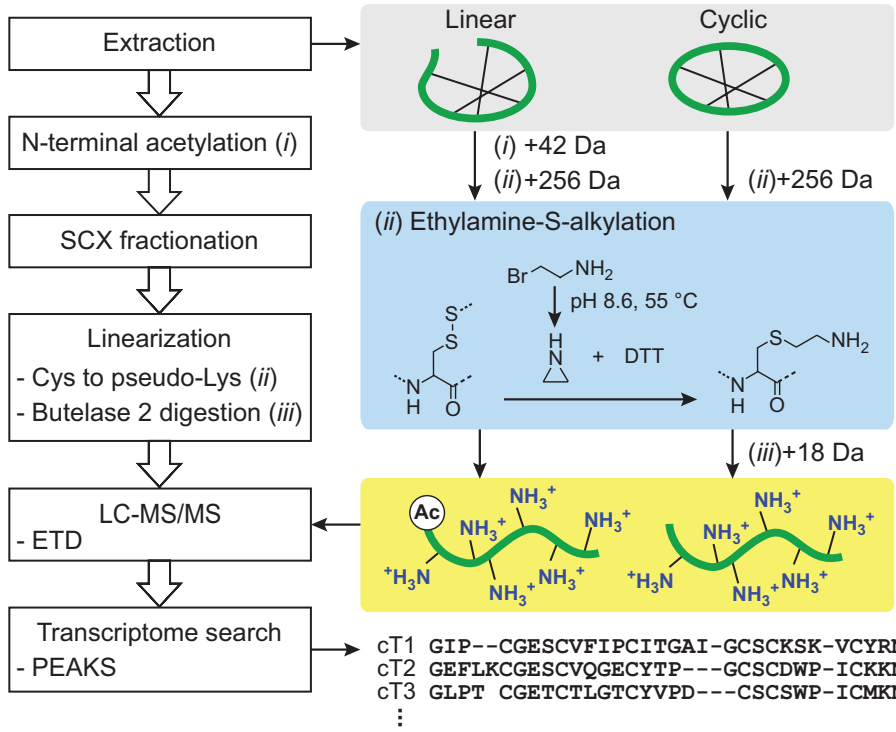



**Fig. 1** Flow chart for peptidomic analysis of plant cyclotide extract containing both cyclotides and acyclotides

development of orally active peptidyl drugs [4]. Our recent study suggests that a cyclotide-producing plant could produce hundreds of cyclotides through hypervariable sequences, which are further modified by posttranslational modifications (PTMs). A novel mechanism of PTM is the fuzzy bioprocessing that results in both cyclic and linear products from the same precursor in cyclotide biosynthesis.

To characterize a full spectrum of CRPs and to elucidate their molecular diversity, we developed a rapid and accurate peptidomic approach and demonstrated its efficiency using cyclotides from *Clitoria ternatea* [5]. In our approach the peptides are chemically and enzymatically modified in three steps (Fig. 1): First, the naturally occurring linear cyclotides in the mixture are pre-labeled by N-terminal acetylation so that they can be distinguished from their cyclic counterparts that will be linearized in a later treatment. Second, the intramolecular disulfide bonds are reduced and thiols are S-alkylated with ethylamine. The reduction and alkylation are performed with bromoethylamine (BrEA) and dithiothreitol (DTT) in a one-pot reaction triggered by heating and basic pH. Third, to open the cyclic backbones with regiospecificity, the Asx-specific endopeptidase butelase 2 is used to linearize cyclotides at the biosynthetic ligation site. Butelase 2 is a protease that is found in the transcriptome of

*C. ternatea* with sequence homology to butelase 1, a ligase that forms cyclic peptides. Because butelase 2 recognizes the same reactive site as butelase 1, the proteolytic action of butelase 2 results in hydrolysis and the generation of sequences homologous to those contained in the transcriptome. The three-step treatment allows peptide to be separated and fragmented by liquid-chromatography tandem mass spectrometry (LC-MS/MS) and side-chain modifications can be identified by direct transcriptomic analysis (Fig. 1).

## 2 Materials

### 2.1 Plant Material, Cell Line, and Microorganisms

1. Fresh *Clitoria ternatea* plant (weight ~ 30 g).
2. Sf9 insect cells adapted to serum-free medium (we use cells from Invitrogen).
3. *E. coli* Mach1 Phage-Resistant Chemically Competent (ThermoFisher Scientific).
4. *E. coli* DH10BAC (Invitrogen).

### 2.2 Chemicals, Reagents, and Kits

Most of the reagents needed are described in the following sections, and should use high purity HPLC-grade or better. Water used for procedures should be deionized.

1. Antibiotics: ampicillin, penicillin, streptomycin, gentamycin, tetracyclin, kanamycin.
2. Benzonase, purity grade II (>90%) for biotechnology.
3. BseRI restriction enzyme.
4. Cationic-lipid formulation for transfection of insect cells (we use Cellfectin® II reagent from Invitrogen).
5. dCTP and dGTP solution.
6. High-fidelity PCR master mix with GC buffer (we use Phusion®, Thermofisher).
7. pFB-LIC-Bse vector (available from Addgene, Cambridge, MA, USA).
8. Protease Inhibitor Cocktail (we use Nacalai Tesque).
9. T4 DNA polymerase.
10. Tobacco Etch Virus (TEV) protease.
11. Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) 10% acrylamide gel.

### 2.3 Buffers and Solutions

#### 2.3.1 Extraction, Clarification, and Fractionation of CRPs

1. 1 N HCl.
2. 1 N NaOH solution.
3. 20% ethanol in water with 10 mM HCl, pH 2.
4. 80% acetonitrile (ACN) in water with 0.1% trifluoroacetic acid (TFA).

5. 10% ACN in water with 0.1% TFA.
6. MALDI-TOF matrix solution: α-Cyano-4-hydroxycinnamic acid (CHCA) saturated solution in 80% ACN with 0.1% TFA.

*2.3.2 Expression of Butelase 2 and RNA Extraction*

1. P1, P2, P3, TE buffers for bacmid purification (Qiagen).
2. Lysogeny broth.
3. Insect cell culture media (we use SF900-III, Gibco).
4. Grace's insect medium.
5. Fetal-bovine serum (FBS).
6. Lysis buffer: 100 mM HEPES, 500 mM NaCl, 10 mM imidazole, 10% glycerol, 0.5 mM TCEP, adjust pH to 8.0 with 1 N NaOH, benzonase (8 μL for each 10 g cell pellet), protease inhibitor cocktail (50 μL in 100 mL lysis buffer).
7. Immobilized metal ion affinity chromatography (IMAC) wash and elution buffers.
8. IMAC wash 1 buffer: 20 mM HEPES, 500 mM NaCl, 10 mM imidazole, 10% (v/v) glycerol, 0.5 mM TCEP, adjust pH to 7.5 with 1 N NaOH.
9. IMAC wash 2 buffer: 20 mM HEPES, 500 mM NaCl, 25 mM imidazole, 10% (v/v) glycerol, 0.5 mM TCEP, adjust pH to 7.5 with 1 N NaOH.
10. IMAC Elution buffer: 20 mM HEPES, 500 mM NaCl, 500 mM imidazole, 10% (v/v) glycerol, 0.5 mM TCEP, adjust pH to 7.5 with 1 N NaOH.
11. Gel Filtration Buffer: 20 mM HEPES, 300 mM NaCl, 10% (v/v) glycerol, 0.5 mM TCEP, adjust pH to 7.5 with 1 N NaOH.
12. Enzyme activation buffer: 20 mM citrate buffer prepared by mixing citric acid and sodium citrate solution to pH 5, 1 mM DTT.
13. Digestion buffer: 0.2 M sodium phosphate buffer prepared by mixing monosodium phosphate solution and disodium phosphate solution to pH 6.5, 10 mM EDTA.
14. RNA extraction solution (we use Trizol reagent, invitrogen).
15. 10 mM Tris-Cl, adjust pH to 8.5 using 1 N HCl.
16. SDS-PAGE running buffer: 6 g Tris-base, 28.8 g glycine, 0.1% w/v SDS in 1 L water, adjust pH to 8.5 with 1 N HCl.
17. Protein sample loading buffer for SDS-PAGE (we use Laemmli from Bio-rad).

*2.3.3 Chemical and Enzymatic Derivatization of Cyclotides*

1. 0.2 M Ammonium acetate buffer, adjust pH to 5.0 with glacial acetic acid.
2. 1 M acetic anhydride solution in ACN.
3. 20 mM sodium phosphate monobasic ($NaH_2PO_4$) containing 20% (v/v) ACN, adjust pH to 2.9 with 1 N HCl.

4. 0.3 M DTT stock solution in water, store at −20 °C.
5. 0.3 M 2-Bromoethylamine hydrobromide (BrEA) stock solution in water, store at −20 °C.
6. 1 M Tris–HCl buffer (pH 8.6).

#### 2.3.4 Mobile Phases for Sample Preparation and Liquid Chromatography

1. Solvent for sample preparation: 3% ACN, 0.1% formic acid (FA) in water.
2. Eluent A: 0.1% FA. Sonicate for 20 min using an ultrasonic bath.
3. Eluent B: 0.1% FA in 90% ACN. Sonicate for 20 min using an ultrasonic bath.
4. Buffer RP-A: 0.11% TFA in deionized water.
5. Buffer RP-B: 0.1% TFA in ACN.
6. Buffer SCX-A: 20 mM $NaH_2PO_4$ in 20% ACN, adjust pH to 2.9 with 1 N HCl.
7. Buffer SCX-B: 1 M KCl, 20 mM $NaH_2PO_4$, 20% ACN, adjust pH to 2.9 with 1 N HCl.

### 2.4 Instruments (See Note 1)

1. Bioanalyzer Agilent 2100 (Agilent Technologies).
2. Blender.
3. Centrifuge, tabletop.
4. Illumina HiSeq 2000 Sequencing System.
5. Incubator.
6. Lyophilizer.
7. PAGE system.
8. Shaker incubator.
9. Ultrasonic bath.
10. Vortex mixer.
11. Vacuum centrifuge concentrator (such as Speedvac).
12. Ultra-High Performance Liquid Chromatography (UHPLC) system such as the Dionex UltiMate 3000 from Thermo Scientific Inc. (Bremen, Germany), or comparable system.
13. Mass spectrometer with electrospray ionization interface (ESI) such as the Michrom CaptiveSpray source and Orbitrap Elite Hybrid Ion Trap-Orbitrap from Thermo Scientific Inc.
14. Mass spectrometer with Matrix Assisted Laser Desorption/Ionization Time-of-Flight, MALDI-TOF, such as the ABI 4800 from Applied Biosystems (Waltham, MA, USA).

### 2.5 Other Materials

1. 24-well tissue culture plate.
2. Flat-bottomed 96-well microtiter plate.

3. Ni-nitrilotriacetic acid (Ni-NTA) beads (Invitrogen), 1 mL packed in gravity column.
4. HiLoad 16/60 Superdex 75 (or 200) prep grade column (GE Healthcare).
5. 10 kDa molecular-weight cutoff ultracentrifugal filter unit.
6. Desalting column: C-18 Sep-pack, 50 mg.
7. Rapid-flow filter unit 0.45 μm.
8. Oligo(dT) magnetic beads.
9. Cation-exchange column, 200 × 4.6 mm (such as the PolySULFOETHYL A column from PolyLC, Columbia, MD, USA).
10. C-18 Ziptip with 0.6 μL bed volume.
11. Reversed phase column for UHPLC (such as Acclaim™ PepMap RSL C18, 2 μm, 100 Å, 75 μm ID × 15 cm from Thermo Fisher Scientific, Waltham, USA).

### 2.6 Software

1. Transcriptome analysis: Trinity Software (Trinity Software Inc., Arlington, USA).
2. De novo MS/MS sequencing: PEAKS Studio version 7.5 [6] (Bioinformatics Solutions, Waterloo, Canada).

## 3 Methods

### 3.1 Extraction and Clarification

#### 3.1.1 Small Scale Extraction Optimization

1. Collect fresh plant tissues and gently wash it with water. Air-dry the remaining water but prevent plant tissue from dehydration. Separate different plant organs (leaf, stem, flower, pod, and root) and weigh them.
2. Optimize extraction buffer by preparing different concentrations of organic solvent/water mixture and mix with small amount of plant tissue (about 100 mg) in a ratio of 10 mL/g. Different organic solvents such as ethanol, methanol, and ACN can be tested during optimization. Ethanol was chosen for *C. ternatea* extraction.
3. Grind the mixture in a manual grinder to extract soluble ingredient. Spin down debris, at 13,000 × g, 4 °C for 10 min, and take the clear supernatant.
4. Dilute supernatant with water to 20% ethanol and acidify the solution to pH 2 with 1 N HCl and keep the mixture at 4 °C for 1 h to let proteins slowly precipitate.
5. Filter the mixture with a 0.45 μm filter unit.
6. Analyze MS profile by MALDI-TOF MS (*see* **Note 2**). The extraction buffer that provides the highest abundance of peaks in the range of 3000–4000 Da is chosen for the scale-up experiment.

#### 3.1.2 Scale-Up Extraction

1. Mix plant tissues with chilled extraction buffer in the ratio of 10 mL/g at final volume ranging from 100 mL to 1 L.
2. Blend the mixture for 10 min at high speed and high intensity.
3. Centrifuge blended mixture at 4000 × *g* for 30 min at 4 °C.
4. Collect supernatant and dilute into 20% ethanol with distilled water.
5. Acidify the mixture to pH 2 using 1 N HCl and keep the mixture at 4 °C for 1 h to let proteins slowly precipitate. If the mixture contains chlorophyll or other pigments, use equal volume of dichloromethane in the extraction.
6. Filter the mixture with 0.45 μm filter to obtain raw extract (E-I).
7. Analyze MS profile by MALDI-TOF MS (*see* **Note 3**).
8. Load the raw extract E-I in a self-pack C-18 column (20 mL), wash the column with 20% ethanol (with 10 mM HCl, pH 2), and elute peptides with 100 mL of 80% ACN with 0.1% TFA in water.
9. Freeze the reaction mixture in liquid nitrogen.
10. Lyophilize the eluent into dry powder.
11. Weigh the lyophilized material and redissolve in 10% ACN with 0.1% TFA in water to a final concentration of 3.5 mg/mL to obtain the clarified extract (E-II).

### 3.2 Recombinant Expression of Butelase 2

Butelase 2 can be expressed by a baculovirus expression vector system in Sf9 cells, following the protocol reported by Shrestha et al. [7] with minor modifications.

#### 3.2.1 Cloning and Transformation

1. Cloning of Butelase 2 into pFB-Lic-Bse results in the addition of 6xHis-TEV amino acid sequence (underlined) at the N-terminal of the translated protein.

   Amino acid sequence:
   MGHHHHHHSSGVDLGTENLYFQSMARLNPQKEWDSVIRLPTEP
   VDADTDE
   VGTRWAVLVAGSNGYENYRHQADVCHAYQLLIKGGLKEENIVVF
   MYDDIA
   WHELNPRPGVIINNPRGEDVYAGVPKDYTGEDVTAENLFAVILG
   DRSKVK
   GGSGKVINSKPEDRIFIFYSDHGGPGVLGMPNEQILYAMDFIDV
   LKKKHA
   SGGYREMVIYVEACESGSLFEGIMPKDLNVFVTTASNAQENSWG
   TYCPGT
   EPSPPPEYTTCLGDLYSVAWMEDSESHNLRRETVNQQYRSVKER
   TSNFKD
   YAMGSHVMQYGDTNITAEKLYLFQGFDPATVNLPPHNGRIEAKM
   EVVHQR

DAELLFMWQMYQRSNHLLGKKTHILKQIAETVKHRNHLDGSVEL
IGVLLY
GPGKGSPVLQSVRDPGLPLVDNWACLKSMVRVFESHCGSLTQYG
MKHMRA
FANICNSGVSESSMEEACMVACGGHDAGHLHPSKRGYIA

DNA sequence:

ATGGGCCACCATCATCATCATCATTCTTCTGGTGTAGATCTGGG
TACCGA
GAACCTGTACTTCCAATCCATGGCAAGGCTGAACCCACAGAAGG
AGTGGG
ATTCGGTTATTCGCTTACCAACTGAACCGGTAGACGCTGACACG
GATGAA
GTGGGAACACGATGGGCCGTTCTCGTCGCTGGTTCAAACGGCTA
TGAAAA
TTATAGGCATCAAGCCGATGTATGCCATGCATACCAGTTGTTGA
TAAAAG
GTGGATTAAAAGAAGAGAATATTGTGGTGTTTATGTACGATGAC
ATAGCA
TGGCACGAGTTGAATCCCAGGCCTGGAGTCATCATCAACAATCC
TCGGGG
GGAAGATGTGTATGCGGGTGTCCCTAAAGATTACACTGGTGAGG
ACGTGA
CAGCGGAGAACCTATTTGCAGTCATTCTTGGGGACAGGAGTAAA
GTGAAG
GGAGGAAGTGGCAAAGTGATCAACAGTAAACCTGAGGACAGGAT
ATTTAT
TTTTTACTCTGATCATGGAGGTCCCGGAGTTCTTGGGATGCCGA
ACGAGC
AAATCCTTTACGCCATGGATTTTATTGATGTTTTGAAGAAGAAA
CATGCT
TCAGGAGGGTACAGGGAAATGGTTATATACGTGGAAGCTTGTGA
AAGTGG
GAGCCTCTTTGAGGGTATCATGCCCAAGGATCTGAATGTTTTTG
TCACAA
CTGCATCAAACGCACAAGAGAATAGCTGGGGAACTTATTGTCCT
GGGACG
GAGCCTTCTCCACCACCAGAGTACACCACTTGCTTGGGTGATTT
GTACAG
CGTTGCTTGGATGGAAGACAGTGAGAGTCACAATTTGAGAAGGG
AAACGG
TGAACCAACAATACCGCTCGGTAAAGGAACGGACTTCAAATTTT
AAAGAC
TATGCAATGGGATCTCATGTGATGCAATACGGTGACACTAACAT
CACAGC
TGAAAAGCTTTACTTATTCCAAGGTTTTGATCCCGCCACAGTGA
ATTTAC
CTCCACACAACGGCAGGATAGAAGCTAAAATGGAAGTTGTTCAC
CAGAGA

GATGCAGAACTTCTCTTCATGTGGCAAATGTATCAGAGATCAAA
CCATCT
ACTAGGAAAGAAGACACACATCCTTAAGCAAATTGCAGAGACAG
TGAAGC
ATAGGAATCACTTAGATGGTAGCGTGGAACTGATTGGAGTTTTA
CTGTAT
GGACCAGGGAAAGGTTCTCCAGTTCTACAATCCGTGAGGGATCC
TGGTCT
GCCCCTTGTTGACAACTGGGCATGTTTAAAATCAATGGTTCGGG
TATTCG
AGTCTCACTGTGGGTCACTGACTCAGTATGGTATGAAACACATG
CGAGCA
TTCGCCAACATATGCAACAGTGGTGTTTCCGAGTCCTCAATGGA
AGAGGC
TTGTATGGTAGCATGTGGTGGCCATGATGCTGGACATCTACATC
CATCCA
AGAGAGGCTATATTGCTTGA

2. The DNA sequence encoding butelase 2 was amplified by forward and reverse primers that also contain sequences for ligation independent cloning (LIC) (underlined).

   Forward primer:
   TACTTCCAATCCATGGCAAGGCTGAACCCACAGA

   Reverse primer:
   TATCCACCTTTACTGTCAAGCAATATAGCCTCTCTTGGA

   Perform PCR according to the manufacturer's instructions using Phusion high fidelity polymerase.

3. Digest pFB-LIC-Bse vector with BseRI restriction enzyme according to the manufacturer's instructions. Treat PCR product and the pFB-LIC-Bse vector with T4 DNA polymerase in the presence of dCTP and dGTP, respectively. Mix PCR product and vector and incubate at room temperature for 30 min, then inactivate at 70 °C for 2 min. Proceed to transform competent Mach 1 cells with mix through heat shock [8]. Plate the transformation mix and screen for colonies carrying correctly subcloned vectors with insert either through colony PCR or Sanger sequencing.

4. Inoculate 1 mL lysogeny broth (containing 50 μg/mL ampicillin) with a single colony and incubate at 37 °C overnight with agitation. Isolate the plasmids and transform chemically competent DH10Bac cells to produce the recombinant baculovirus genome (bacmid) according to Shrestha et al. [7]. Briefly, transformation to freshly prepared DH10Bac competent cells is performed using heat shock approach. The transformed cells are incubated at 37 degree, 700 rpm for 5 h to allow transposition to take place. After shaking incubation, transfer cell culture to the blue/white selection plate containing Bluo-gal (100 ug/

mL), IPTG (40 ug/mL), gentamycin (70 ug/mL), tetracyclin (100 ug/mL), and kanamycin (50 ug/mL), incubate at 37 degree. After 48 h, pick white colonies for bacmid production.

*3.2.2 Transfection and Infection*

1. Plate 1 mL of 2 × $10^5$ mid-log phase Sf9 cells per mL of SFM900 III medium supplemented with 2% FBS in a 24-well tissue culture plate. Incubate at 27 °C for 1 h.
2. In a separate flat-bottomed 96-well microtiter plate, mix 50 μL Grace's insect medium and 10 μL bacmid DNA. Please refer to Shrestha et al. [7] for the bacmid preparation. Add a mixture of 50 μL Grace's insect medium and 3 μL Cellfectin® into the bacmid mixture and incubate for 45 min at room temperature inside a biosafety cabinet.
3. Wash the Sf9 cells with 1 mL of Grace's insect medium. Dilute the bacmid Cellfectin® mixture with 100 μL Grace's insect medium and add into the Sf9 cells. Add additional 0.2 mL Grace's insect medium into the cells and incubate for 5 h at 27 °C.
4. Remove the transfection mixture, and add 0.7 mL SF900 III medium containing 2% FBS and antibiotics (50 U penicillin and 50 μg streptomycin per mL medium). Collect the virus when there is sign of infection (around 96 h post transfection or more). Store this P0 virus at 4 °C and away from light. Add 500 μL of P0 virus to 35 mL SF9 cells (2 × $10^6$ cells/mL) and incubate in a 27 °C shaker incubator, 120 rpm for 3 days to make P1 virus.
5. Add 10 mL of P1 virus to 90 mL SF9 cells (2 × $10^6$ cells/mL) and incubate for 3 days in 27 °C shaker incubator to make P2 virus.
6. Add 25 mL of P2 virus to infect 1 L cells (2 × $10^6$ cells/mL of SF900 III supplemented with 1% FBS) and grow for 48–56 h for protein production. Harvest the cells by centrifugation at 3000 × *g* for 10 min.

*3.2.3 Protein Lysis and Purification*

1. Lyse 10 g of cell pellet using 50 mL lysis buffer with sonication. Clarify the cell lysate by centrifugation (4500 × *g*, 4 °C, 30 min).
2. Mix clarified extract with 1 mL Ni-NTA beads for 45 min at 4 °C and then load into column. Collect the flow through and load it into the second Ni-NTA column. Wash each column with 50 column volumes (CV) of wash 1 buffer and 50 CV of wash 2 buffer. Elute the protein from the column (2 CV for each elution until all the protein is eluted). Protein needs to be desalted before the next step.
3. Add Tobacco Etch Virus (TEV) protease (1:40 ratio w/w) to the desalted protein solution and incubate at 4 °C overnight to

remove the His6-TEV tag. Reload the solution into column packed with 1 mL Ni-NTA beads and wash with IMAC 2 wash buffer. Collect the flow through and wash solutions. Analyze the fractions by SDS-PAGE.

4. Take a small aliquot from each fractions containing butelase 2 precursor (54 kDa) to test activity. Auto-activate butelase 2 by acidifying the stock solution to pH 5 using 100 mM sodium acetate buffer and incubate at 37 °C for 2 h. Mix 0.1 μM enzyme with 50 μM Z-AAN-AMC in sodium phosphate buffer (pH 6.0) and incubate at 37 °C for 30 min. Measure relative fluorescence intensity at 460 nm. The increase of intensity by about tenfolds indicates the recombinant butelase 2 is functionally active.
5. Concentrate butelase 2 precursors by Amicon ultracentrifugal unit with a 10 kDa molecular-weight cutoff. Calculate the zymogen concentration by 205 nm absorbance (Conc/1 mg/mL = $A_{205}$/31). Store solution at −20 °C.

### 3.3 Extraction of RNA and Preparation of Transcriptome

Full transcriptome data of common plants can be obtained on NCBI SRA database or OneKP plant database. If the information is not available online, RNA can be extracted from fresh plant using Trizol reagent and send to a sequencing company for transcriptome analysis. Here we describe briefly the general procedure of extraction, sequencing, and assembly.

1. Extract RNA from fresh plant tissue using Trizol reagent.
2. Quality check for the RNA extracts using Agilent 2100 Bioanalyzer.
3. Enrich Poly(A) RNA using oligo(dT) magnetic beads and fragment as template for cDNA synthesis.
4. Purify short fragments and resolve with 10 mM Tris-Cl, pH 8.5 for end reparation and single nucleotide A (adenine) addition.
5. Select suitable fragments for PCR amplification.
6. Sequence the mixture using Illumina HiSeq 2000.
7. Analyze and assemble the RNA sequences using Trinity software to get transcriptome data.

### 3.4 Chemical and Enzymatic Derivatization of Cyclotides

#### 3.4.1 N-Terminal Acetylation of Linear Peptides (**See Note 4**).

1. Pre-chill E-II extract on ice and adjust pH to 3.0–3.3 using 0.2 M ammonium acetate buffer.
2. Add to E-II extract 1 M acetic anhydride solution and 1% acetic acid in a ratio of 20:3:7 (v/v).
3. Incubate on ice for 5–10 min.
4. Freeze the reaction mixture in liquid nitrogen.
5. Lyophilize the mixture to remove unreacted acetic anhydride obtaining freeze-dried E-III extract.

*3.4.2 Fractionation by Strong Cation Exchange Liquid Chromatography (SCX-LC) (**See Note 5**)*

1. Redissolve the lyophilized E-III extract into the concentration of 3.5 mg/mL in buffer SCX-A.
2. Determine MALDI MS profile of E-III to confirm the N-terminal acetylation of linear peptides (*see* **Note 6**).
3. Fractionate E-III extract by SCX-LC with the ion exchange column using a gradient of 10–60% buffer SCX-B. Collect fractions every minute and combine them according to peak distribution to obtain multiple E-IV fractions.
4. Desalt E-IV fractions using a reversed-phase C-18 cartridge. Elute peptides from the cartridge with 10 mL buffer RP-B.
5. Determine MALDI MS profiles for each E-IV fraction.
6. Freeze E-IV fractions in liquid nitrogen.
7. Lyophilize E-IV fractions.

*3.4.3 Conversion of Cys to Pseudo-Lys*

1. Redissolve lyophilized E-IV fractions in water to a final concentration of 3.5 mg/mL. This concentration is equivalent to 0.5–1.5 mM since cyclotides consist of 28–38 residues and their molecular mass ranges from 3 to 4 kDa. Fractions with low-abundant peptides can be concentrated to enhance MS signals.
2. Thaw DTT stock solution and BrEA stock solution on ice.
3. Add DTT, BrEA, and Tris–HCl buffer (pH 8.6) to a final concentration of 0.5 mM peptides, 30 mM DTT, 60 mM BrEA, and 0.2 M Tris–HCl buffer.
4. Perform one-pot reduction and alkylation of peptides incubating the E-IV fractions at 55 °C for 60 min (*see* **Note 7**).
5. Quench the reaction by adjusting pH to 6 using 1 N HCl.
6. Desalt the reduced and alkylated E-IV fractions using a reversed-phase C-18 cartridge. Elute peptides with 1 mL buffer RP-B obtaining reduced and alkylated E-V fractions.
7. Determine MALDI MS profiles for each E-V fraction.
8. Dry E-V fractions by SpeedVac (no heating).
9. Redissolve E-V fractions in water to a concentration of 3.5 mg/mL. Keep E-V fractions at −20 °C or use fresh in the next step.

*3.4.4 Linearization of Cyclic Backbone by Butelase 2*

1. Thaw E-V fractions on ice, if required.
2. Thaw butelase 2 stock solution on ice and bring to 25 °C.
3. Activate butelase 2 at pH 5 by mixing 8 μL 0.1 μM butelase 2 with 12 μL enzyme activation buffer and incubate at 37 °C for 2 h.
4. Add butelase 2 and digestion buffer to each E-V fraction to a final concentration of 2 μM butelase 2, 0.5 mM reduced and alkylated peptide mixture, and 5 mM EDTA in 0.1 M sodium phosphate buffer (pH 6.5).

5. Incubate reaction mixture at 42 °C for 30 min.
6. Quench butelase 2 reaction by adjusting pH to 4 using 1 N HCl obtaining digested E-VI fractions.
7. Store E-VI fractions at −20 °C or use fresh in the next step.

### 3.5 LC-MS/MS Analysis of CRP Mixtures

#### 3.5.1 Sample Preparation for LC-MS/MS Analysis

1. Redissolve E-VI fractions in 20 μL of 3% ACN, 0.1% FA in water.
2. Shake E-VI fractions for 20 min at high speed using a vortex mixer.
3. Sonicate E-VI fractions for 20 min in an ultrasonic bath.
4. Centrifuge E-VI fractions at 15,000 × *g* at 4 °C for 20 min.
5. Transfer supernatants into new HPLC vials (*see* **Note 8**).

#### 3.5.2 Liquid Chromatography

Separation of CPRs is performed with a UHPLC system using a reversed phase column with:

*Flow rate*: 0.3 μL/min.

*Column chamber temperature*: 35 °C.

*Injection volume:* ~ 1 μL (corresponding to ~ 1 μg protein). Injection volume can be adjusted depending on peptide concentration and signal strength.

*Gradient:* 3% eluent B for 1 min, 3–35% B over 47 min, 35–50% B over 4 min, 50–80% B over 6 s, 80% for 78 s and then reverted to the initial state over 6 s and maintained isocratic for 6.5 min.

It is recommended to optimize separation of peptides according to the peptide hydrophobicity by modifying the slope of the gradient when analyzing other CRP mixtures.

#### 3.5.3 Mass Spectrometry Analysis (**See Note 1**)

Analysis of CRPs is performed in an ESI mass spectrometer. Sample is sprayed using a nanospray ion source and peptides are analyzed in positive ion mode. Data acquisition is performed alternating between full Fourier Transform-MS (FT-MS) (350–3000 $m/z$, resolution 60.000, with 1 micro-scan per spectrum) and FT-MS/MS (150–2000 $m/z$, resolution 30.000, with 2 micro-scan averaged per MS/MS spectrum) using electron transfer dissociation (ETD) activation mode at 65, 80, and 95 ms of activation times with an ETD reagent automatic gain control (AGC) of $5 \times 10^5$. FT-MS/MS is performed for the three most intense precursors with charge >2+. The precursor ions are isolated using a 2 Da isolation window. The automatic gain control of both full FT-MS and FT-MS/MS is set to $1 \times 10^6$.

### 3.6 Data Analysis

Data analysis is performed by PEAKS studio software using 10 ppm MS and 0.05 Da MS/MS tolerances. A false discovery rate < 1% is recommended. Posttranslational modifications are searched using PEAKS PTM algorithm [9] (*see* **Notes 9** and **10**).

## 4 Notes

1. Instruments mentioned in this protocol may be substituted by others with similar capabilities although conditions should be optimized to get similar results.
2. MALDI-TOF MS analysis of CRPs is performed in reflectron acquisition mode using 1000–6000 Da mass range with a focusing mass of 3000 Da. A C-18 Ziptip is used to desalt the sample before spotting. Each MALDI spot contains 0.5 μL of sample and 0.5 μL of saturated α-cyano-4-hydroxycinnamic acid (CHCA) MALDI matrix prepared in 75% ACN with 0.1% TFA.
3. Fresh CRP extract provides the most abundant data. After prolonged storage, the signals of low-abundant noncyclic fragments are reduced or missing.
4. N-terminal acetylation has to be performed on ice and during limited reaction time to prevent undesired side-chain acetylation.
5. Fractionation of E-III extract can be performed using other stationary and mobile phases depending on chemical properties of the extract. Based on the transcriptome data we can obtain a broad image of CRP sequences. If the population is mainly negatively charged, anion-exchange chromatography is recommended.
6. Linear peptides are generally present at low abundances, thus appreciation of the mass shift generated by the acetylation addition could be not obvious.
7. Do not use long incubation times for the one-pot reduction and alkylation in order to avoid polymerization of the alkylating agent. Similarly, use mild temperature for the reaction.
8. While transferring the supernatant, avoid touching the bottom of the tube with the tip. Do not disturb the pelleted material that may be generated during centrifugation.
9. CRPs extracted from fresh plant tissues and old tissues have different posttranslational modification profiles, indicating aging-related changes.
10. Other search engines or software can be used for the same purpose.

### References

1. Silverstein KAT, Graham MA, Paape TD, VandenBosch KA (2005) Genome organization of more than 300 defensin-like genes in arabidopsis. Plant Physiol 138(2):600–610. https://doi.org/10.1104/pp.105.060079
2. Tam J, Wang S, Wong K, Tan W (2015) Antimicrobial peptides from plants. Pharmaceuticals 8(4):711
3. Craik D, Daly N, Bond T, Waine C (1999) Plant cyclotides: a unique family of cyclic

and knotted proteins that defines the cyclic cystine knot structural motif. J Mol Biol 294:1327–1336
4. Wong C, Rowlands D, Wong C-H, Lo T, Nguyen G, Li H-Y, Tam J (2012) Orally active peptidic bradykinin B1 receptor antagonists engineered from a cyclotide scaffold for inflammatory pain treatment. Angew Chem Int Edit 51(23):5620–5624. https://doi.org/10.1002/anie.201200984
5. Serra A, Hemu X, Nguyen GK, Nguyen NT, Sze SK, Tam JP (2016) A high-throughput peptidomic strategy to decipher the molecular diversity of cyclic cysteine-rich peptides. Sci Rep 6:23005. https://doi.org/10.1038/srep23005
6. Ma B, Zhang K, Hendrie C, Liang C, Li M, Doherty-Kirby A, Lajoie G (2003) PEAKS: powerful software for peptide de novo sequencing by tandem mass spectrometry. Rapid Commun Mass Spectrom 17(20):2337–2342. https://doi.org/10.1002/rcm.1196
7. Shrestha B, Smee C, Gileadi O (2008) Baculovirus expression vector system: an emerging host for high-throughput eukaryotic protein expression. In: Starkey M, Elaswarapu R (eds) Genomics protocols, Methods in molecular biology™, vol 439. Humana Press, Totowa, NJ, pp 269–289. https://doi.org/10.1007/978-1-59745-188-8_19
8. Gräslund S, Nordlund P, Weigelt J, Hallberg BM, Bray J, Gileadi O, Knapp S, Oppermann U, Arrowsmith C, Hui R, Ming J, dhe-Paganon S, Park H-w, Savchenko A, Yee A, Edwards A, Vincentelli R, Cambillau C, Kim R, Kim S-H, Rao Z, Shi Y, Terwilliger TC, Kim C-Y, Hung L-W, Waldo GS, Peleg Y, Albeck S, Unger T, Dym O, Prilusky J, Sussman JL, Stevens RC, Lesley SA, Wilson IA, Joachimiak A, Collart F, Dementieva I, Donnelly MI, Eschenfeldt WH, Kim Y, Stols L, Wu R, Zhou M, Burley SK, Emtage JS, Sauder JM, Thompson D, Bain K, Luz J, Gheyi T, Zhang F, Atwell S, Almo SC, Bonanno JB, Fiser A, Swaminathan S, Studier FW, Chance MR, Sali A, Acton TB, Xiao R, Zhao L, Ma LC, Hunt JF, Tong L, Cunningham K, Inouye M, Anderson S, Janjua H, Shastry R, Ho CK, Wang D, Wang H, Jiang M, Montelione GT, Stuart DI, Owens RJ, Daenke S, Schütz A, Heinemann U, Yokoyama S, Büssow K, Gunsalus KC (2008) Protein production and purification. Nat Meth 5(2):135–146. http://www.nature.com/nmeth/journal/v5/n2/suppinfo/nmeth.f.202_S1.html.
9. Han X, He L, Xin L, Shan B, Ma B (2011) PeaksPTM: mass spectrometry-based identification of peptides with unspecified modifications. J Proteome Res 10(7):2930–2936. https://doi.org/10.1021/pr200153k

# Chapter 27

# Analysis of Endogenous Peptide Pools of *Physcomitrella patens* Moss

**Igor Fesenko, Regina Khazigaleeva, Vadim Govorun, and Vadim Ivanov**

## Abstract

Here, we report our approach to peptidomic analysis of the plant object which led to structure elucidation of the title peptides. *P. patens* samples were extracted under conditions preventing proteolytic digestion by endogenous proteases. The extracts were fractionated on size exclusion columns and the peptide fractions subjected to LC-MS/MS analysis. Mass spectra datasets were analyzed for the presence of peptides derived from the proteins encoded by the moss genome. Experimental details are given for each step, selected chromatograms and mass-spectra are presented in figures.

**Key words** *Physcomitrella patens*, Peptidomics, Size exclusion chromatography, Mass spectrometry

## 1 Introduction

In the past decade, participation of endogenous peptides in plant growth and development has been demonstrated [1]. Peptides were shown to affect the defense reaction of plants, cell division, nodulation, reproduction, and other functions [2, 3]. In each case, the peptides were found during the search of the active signaling molecule triggering the given biological phenomenon. However, comprehensive study of the peptide fraction of plant origin remained beyond the scope of available analytical techniques. Rapid development of high resolution mass-spectrometry coupled with growing availability of genomic information opened new perspective for systemic study of protein and peptide structures of plant origin.

In our laboratory, we carried out such study of the model plant object *P. patens* [4]. There are two life forms of moss—filamentous protonemata representing the juvenile stage and leafy gametophores that represent the adult stage [5]. Also, protonema filaments are a source of protoplasts, which provide an ideal system for the study of plant development mechanisms due to its ability to form intact plants [6]. Both plant tissues and protoplast cells were studied in

Michael Schrader and Lloyd Fricker (eds.), *Peptidomics: Methods and Strategies*, Methods in Molecular Biology, vol. 1719, https://doi.org/10.1007/978-1-4939-7537-2_27, 

the abovementioned work. Generally, peptidomic research requires considerable care on the stage of sample preparation. Degradation of the protein content by endogenous proteases might lead to formation of artifact peptides. On the other hand, peptides might escape detection as a result of efficient binding with endogenous proteins. In several cases (e.g., analysis of blood samples [7]) brief heating is applied to inactivate proteases and dissociate peptide–protein complexes. In our work we suppressed proteolysis by a specific cocktail of inhibitors. No measures were taken to prevent peptide–protein association. Judging from the abundance of detected peptides such binding did not introduce any major distortion into the overall picture.

In this communication we provide a detailed description of the respective workflow which led to discovery of over 20,000 endogenous peptides. We consider the respective list of sequences as a first presentation of a plant peptidome.

## 2 Materials

1. Plant material: The protonemata and gametophores of the moss *P. patens* subsp. *patens* Gransden 2004.
2. Plastic dishes for the growth of protonemata and gametophores.
3. Knop medium with 1.5% agar.
4. Plant Growth Incubator (such as Sanyo MLR-352H, Panasonic, Osaka, Japan) with a photon flux of 61 μmol/$m^2$•s during a 16-h photoperiod at 24 °C.
5. 500 mg/L ammonium tartrate.
6. Extraction buffer: 1 M acetic acid in 10% acetonitrile (*see* **Note 1**).
7. Protease Inhibitor Cocktail (Sigma-Aldrich, USA).
8. DMSO (*see* **Note 2**).
9. Grinder, such as the ball mill Mikro-Dismembrator S (Sartorius, Goettingen, Germany) (*see* **Note 3**).
10. Shaking flasks made of PTFE, volume approx. 25 mL.
11. Glass beads of three sizes: $d$ = 0.1 mm, $d$ = 0.3 mm, and $d$ = 1 mm.
12. Size exclusion chromatography system, such as the ÄKTA pure 25 L (GE Healthcare Life Science).
13. UV monitor for detection at 280 nm.
14. Sample pump that can provide flow rates up to 50 mL/min.
15. Fraction collector.

16. Size exclusion column, such as the XK 26/40 column (GE Healthcare Life Science, USA), prepacked with Sephadex G25 Superfine (*see* **Note 4**).
17. Mobile phase: 0.1 M aqueous acetic acid.
18. Standard protein: 3 mg/mL Aprotinin (Mr = 6500 Da) in 50 mM phosphate buffer, 150 mM NaCl, pH (6–8) (*see* **Note 5**).
19. Freeze Dry system, such as the FreeZone 2.5 Liter 4 (Labconco, Kansas City, USA).
20. Reversed-phase C18 microcolumns prepared in 200 µL tips for an automatic pipette with two layers of Empore™ extraction disk reversed-phase C18 membrane (Supelco, Bellefonte, PA, USA) 1.6 mm in diameter [8].
21. Dissolve solution: 50% acetonitrile/49.9% $H_2O$/0.1% trifluoroacetic acid (TFA).
22. Conditioning solution: Methanol.
23. Equilibration solution: 0.1% TFA in $H_2O$.
24. Elution solution: 80% acetonitrile/19.9% $H_2O$/0.1% TFA.
25. Dilution solution: 5% acetonitrile/94.9 $H_2O$/0.1% TFA.
26. Vacuum concentrator, such as the Speedvac concentrator from Savant (Thermo Fisher scientific, Waltham, MA, USA).
27. Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF-MS) such as the Ultraflex™ TOF/TOF (Bruker Daltonics, Bremen, Germany).
28. Matrix solvent: 50% acetonitrile, 50% proteomics grade water, and 0.1% TFA.
29. Matrix: 20 mg/mL 2,5-dihydroxybenzoic acid (DHB) in matrix solvent.
30. NanoLC-MS/MS analysis system, such as the TripleTOF 5600 + mass-spectrometer with NanoSpray III ion source (ABSciex, Framingham, USA) coupled with a NanoLC ultra 2D+ nano-HPLC system (Eksigent, Dublin, CA, USA).
31. Trap column Chrom XP C18 (3 µm, 120 Å, 350 µm × 0.5 mm; Eksigent).
32. Separation column 3C18-CL-120 (3 µm, 120 Å, 75 µm × 150 mm; Eksigent).
33. Buffer A: 1% methanol, and 0.1% formic acid (FA) in $H_2O$.
34. Buffer B: 0.1% FA in acetonitrile.
35. Data processing software: ProteinPilot software 4.5 (ABSciex, Framingham, USA).

## 3 Methods

The general scheme of experiment is shown in Fig. 1. To avoid nonspecific degradation of proteins and peptides during extraction, we used an acid extraction buffer containing a mixture of plant protease inhibitors, and all steps were performed on ice.

### *3.1 Calibration of Size Exclusion Chromatography Column*

Use standard proteins (e.g., Aprotinin) for evaluating the separation efficiency.

1. Set up size exclusion chromatography system: wash the tubes and the injector, set the flow rate to 0.5 mL/min, and check the pressure.
2. Set flow rate to 2 mL/min by Unikorn 7 software and equilibrate size exclusion column with 2 V of elution buffer (*see* **Note 6**).
3. Add 4 mL pre-cold Extraction buffer in tube 15 mL.
4. Add 3 mg/mL Aprotinin (6.5 kDa) in Extraction buffer.
5. Fill the syringe with standard solution prepared in **step 4**. Connect the syringe to the column "drop to drop" through sample inject valve.
6. Determine the elution volume (Ve) for Aprotinin, the example is shown in Fig. 2a (*see* **Note 7**).

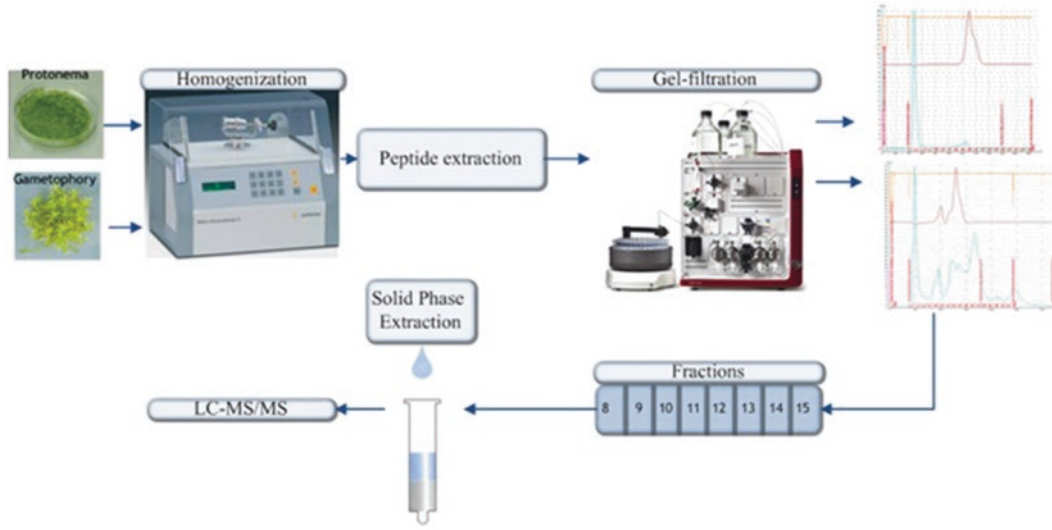



**Fig. 1** Overview of pipeline steps

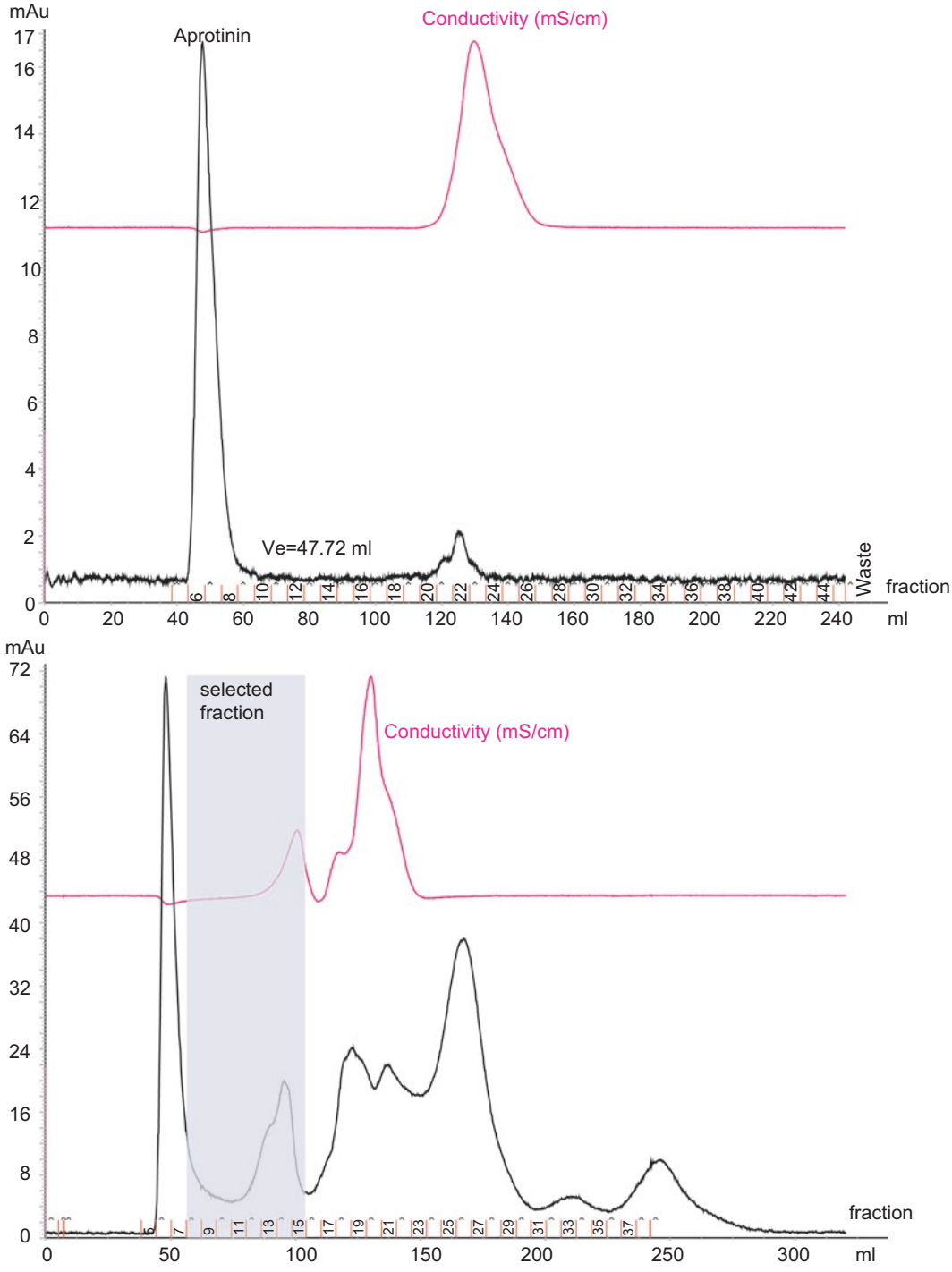



**Fig. 2** Examples of chromatographic separation. Chromatograms of standard protein Aprotinin (**a**) and protonema sample (**b**)

### 3.2 Sample Preparation (see Note 8)

1. Harvest protonemata/gametophores from the surface of the agar medium.
2. Place immediately into a porcelain mortar with liquid nitrogen and ground to fine dust with a pestle pre-cooled to −70 °C.
3. Place three sizes of glass beads (0.1, 0.3, and 1 mm) into shaking flask and put on ice.
4. Prepare 4 mL of extraction buffer in tube and place it on ice.
5. Add 67 μL Protease Inhibitor Cocktail Protease Inhibitor in extraction buffer (*see* **Note 9**).
6. Place frozen plant tissue, 2 g for protonemata or 0.3 g for gametophores, into pre-cooled shaking flask with extraction buffer containing proteinase inhibitors.
7. Homogenize using a ball mill at 2600 $min^{-1}$ for 2 min.
8. Transfer the homogenate from shaking flask in centrifuge tube.
9. Centrifuge at 10000 × *g* at 4 °C for 5 min.
10. Transfer supernatant to a new 15 mL centrifuge tube, place on ice, and use for size exclusion chromatography.

### 3.3 Size Exclusion Chromatography of the Peptide Samples

1. Set flow rate of the eluent to 2 mL/min as in **step 2**, Subheading 3.1, pressure 0.5 MPa.
2. Equilibrate chromatographic column.
3. UV detection should be set to 280 nm.
4. Fill the syringe with supernatant from **step 10**, Subheading 3.2. Connect the syringe to the column "drop to drop" to avoid introducing air into the column (*see* **Note 10**).
5. Elute with two volumes of elution buffer.
6. Collect the peptide containing fractions. An example of the chromatogram is shown in Fig. 2b (*see* **Note 11**).
7. Transfer the collected fractions in 15 mL centrifuge tubes and freeze in liquid nitrogen.
8. Freeze-dry the collected fractions.

### 3.4 Redissolving and Desalting Peptide Samples

1. Dissolve each lyophilized fraction in Dissolve solution (500 μL).
2. Centrifuge samples at 10,000 × *g*, at room temperature for 5 min.
3. Transfer each supernatant into individual centrifuge tube.
4. Concentrate samples to 30 μL using a concentrator/SpeedVac. Check the progress every 1–2 min.
5. Dilute samples to 100 μL by adding Dilution solution and mix thoroughly.

6. Centrifuge at 15,000 × *g* for 10 min.
7. Accurately pick out supernatants in individual centrifuge tubes.
8. Wet the reversed phase C18 microcolumns by passing 30 μL Conditioning solution (*see* **Note 12**).
9. Add 30 μL equilibration solution to the microcolumn and pass the solution through (*see* **Note 13**).
10. Load the sample (from **step 7**) onto the microcolumn and pass through the tip.
11. Add 30 μL Equilibration solution to the microcolumn and pass the liquid through the tip.
12. Add 30 μL Elution solution and collect the eluate in new tube.
13. Concentrate samples to 5 μL using a concentrator/SpeedVac. Check the progress every 1–2 min.
14. Add 10 μL Dilution solution.

### 3.5 MALDI-TOF Mass Spectrometry

1. Mix 1 μL of Matrix (20 mg/mL DHB) with 1 μL of the peptide sample.
2. Deposit 0.5 μL of the mixture on the MALDI target plate and air dry.
3. Perform MALDI analysis for different fractions. Examples of mass spectra are shown in Fig. 3.

### 3.6 LC-Coupled Tandem Mass Spectrometry and Data Analysis

1. Set up the nano-HPLC system and configure it in a trap–elute mode.
2. Load 3 μL peptide sample on a trap column at 3 μL/min flow rate for 10 min.
3. Elute through the separation column into the ESI-QTOF mass spectrometer at a flow rate of 300 nL/min using a linear mobile phase gradient (A/B): from 5 to 40% buffer B over 120 min.
4. Collect tandem mass spectra (MS2) in an information-dependent acquisition mode (*see* **Note 14**).
5. Regenerate the column and the precolumn between runs with 95% buffer B for 7 min. Equilibrate the column with 5% buffer B for 25 min.
6. Analyze tandem mass spectrometry data with the ProteinPilot Data processing software using the search algorithm Paragon and the default parameter set for protein identification with the following adjustments: no enzyme, uniref100_Physco_35213 protein sequence database, no Cys alkylation, TripleTOF 5600 equipment, organism type not specified, search effort—thorough ID, detection protein threshold—unused protein score 0.05.

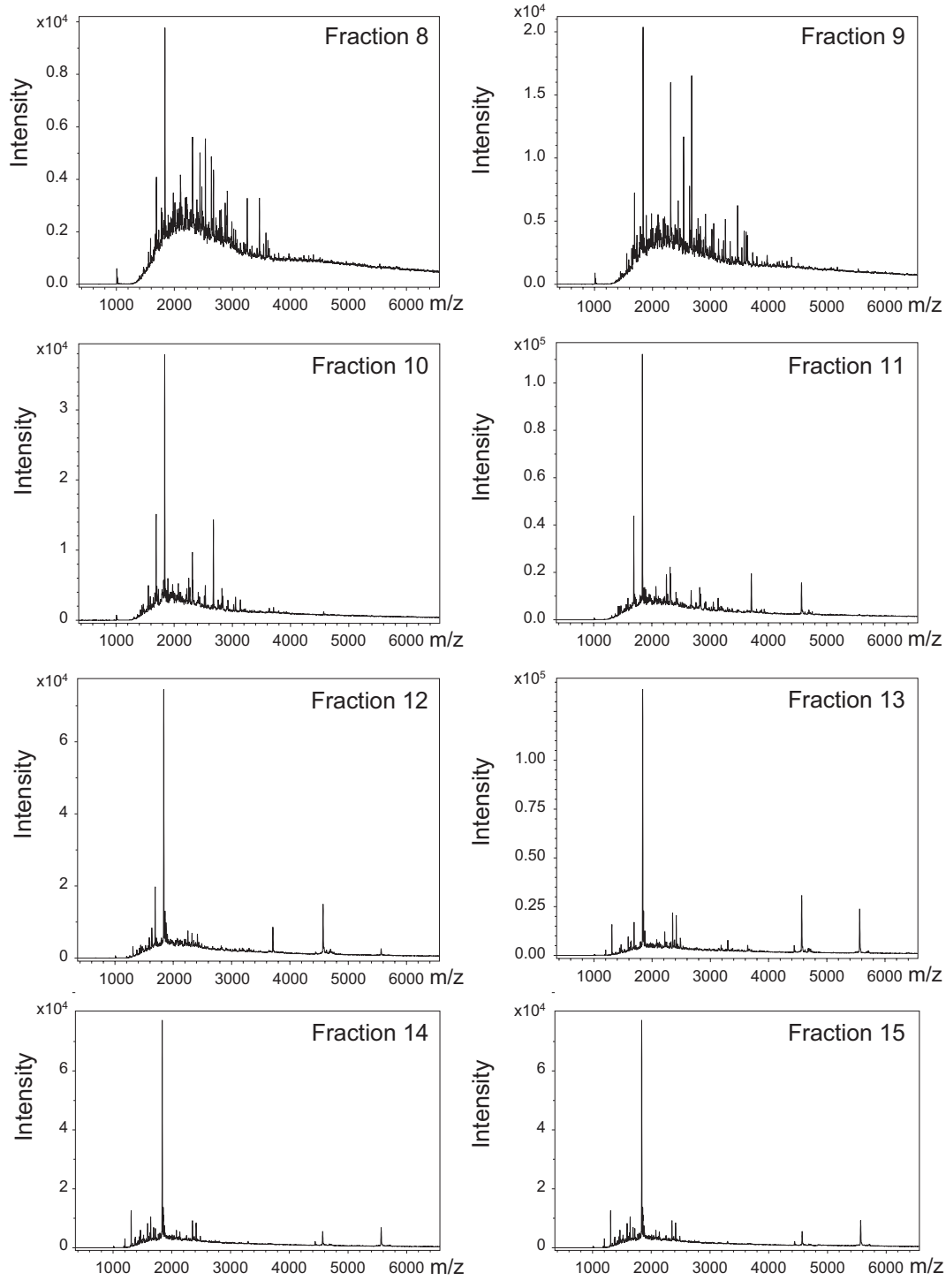



**Fig. 3** MALDI-TOF MS analysis of peptide containing fractions

## 4 Notes

1. To minimize artifacts during peptide extraction, we use an acid extraction buffer with a mixture of plant protease inhibitors.
2. The Protease Inhibitor Cocktail (Sigma-Aldrich, USA) inhibits serine-, cysteine-, and aspartic proteases, metalloproteases, and aminopeptidases. This mixture contains AEBSF, 1,10-phenanthroline, Pepstatin A, Leupeptin, Bestatin, and E-64. AEBSF inhibits serine proteases; bestatin inhibits aminopeptidases; E-64 acts against cysteine proteases; Leupeptin acts against both serine and cysteine proteases. Pepstatin A inhibits acid proteases; 10-phenanthroline acts against metalloproteases.
3. There are two main strategies to disrupt cell wall in plant: (a) apply mechanical forces; (b) enzymatical degradation. Ball mills allow effective and quick mechanical homogenization of plant tissue, thereby minimizing plant protease action. In addition, using cellulolytic enzymes may interfere with subsequent analysis.
4. There are different size exclusion chromatography columns with different ranges of molecular mass separations. Sephadex G-25 is a well established gel filtration medium that allows for separation of peptides (from 1000 to 5000 Da) from globular proteins.
5. The calibration procedure should be performed to select peptide containing fractions. Aprotinin (Mr = 6500 Da) in 50 mM phosphate buffer can be used for this procedure.
6. The flow rates and pressure limits depend on the column and specific configuration of the chromatographic system. Column equilibration is a time-consuming process and you can run it overnight at a low rate.
7. The elution volume (Ve) is the volume of eluent collected from the start of loading the sample to the point of its maximal elution.
8. Extraction is the crucial step of peptidomic analysis and, therefore, all procedures should be carried out without unnecessary delays, at 4 °C, to minimize protein degradation.
9. According to the manufacturer's instruction, 1 mL of solution is recommended for inhibition of protease activity in 100 mL of cell lysate from 30 g of plant tissue.
10. The sample volume should not exceed 3% of the total volume.
11. The peptide containing fractions should be collected after the protein peak. According to column calibration we collect eight fractions, about 6 mL of each, as shown in Fig. 2 of Chapter 23.

12. Conditioning activates the reversed-phases to provide consistent interaction between the analyte and the sorbent functional groups. Reversed-phase sorbent is usually conditioned with 1 tube volume of a water miscible solvent such as methanol or acetonitrile.
13. Reversed-phase peptide separation is usually performed under acidic conditions. At those pH values below 3, practically all basic amino acid residues and the free N-termini are positively charged enhancing thereby retention of peptides. TFA reduces the pH and should be added to mobile phase at a concentration of 0.1%. This concentration produces good peak shapes with most reversed-phase columns.
14. An information-dependent mass-spectrometer (MS) experiment includes one survey MS1 scan followed by 50 dependent MS2 scans. MS1 acquisition parameters are 300–1250 *m/z* mass range for analysis and subsequent ion selection for MS2 analysis and 250 ms signal accumulation time. Ions for MS2 analysis are selected on the basis of intensity with a threshold of 400 cps and a charge state from 2 to 5. MS2 acquisition parameters: resolution of quadrupole set to UNIT (0.7 Da), measurement mass range 200–1800 *m/z*, optimization of ion beam focus to obtain maximal sensitivity, and signal accumulation time of 50 ms for each parent ion. Collision activated dissociation is performed with nitrogen gas with collision energy ramping from 25 to 55 V within the 50 ms signal accumulation time. Analyzed parent ions are sent to a dynamic exclusion list for 15 s to get an MS2 spectrum at the chromatographic peak apex (minimum peak width throughout the gradient was about 30 s).

## Acknowledgments

This work was supported by the Russian Science Foundation (project No. 14-50-00131).

### References

1. Tavormina P, De Coninck B, Nikonorova N, De Smet I, Cammue BP (2015) The plant peptidome: an expanding repertoire of structural features and biological functions. Plant Cell 27(8):2095–2118. https://doi.org/10.1105/tpc.15.00440
2. Marmiroli N, Maestri E (2014) Plant peptides in defense and signaling. Peptides 56:30–44. https://doi.org/10.1016/j.peptides.2014.03.013
3. Durgo H, Klement E, Hunyadi-Gulyas E, Szucs A, Kereszt A, Medzihradszky KF, Kondorosi E (2015) Identification of nodule-specific cysteine-rich plant peptides in endosymbiotic bacteria. Proteomics 15(13):2291–2295. https://doi.org/10.1002/pmic.201400385
4. Fesenko IA, Arapidi GP, Skripnikov AY, Alexeev DG, Kostryukova ES, Manolov AI, Altukhov IA, Khazigaleeva RA, Seredina AV, Kovalchuk SI, Ziganshin RH, Zgoda VG, Novikova SE,

Semashko TA, Slizhikova DK, Ptushenko VV, Gorbachev AY, Govorun VM, Ivanov VT (2015) Specific pools of endogenous peptides are present in gametophore, protonema, and protoplast cells of the moss *Physcomitrella patens*. BMC Plant Biol 15:87. https://doi.org/10.1186/s12870-015-0468-7

5. Cove DJ, Perroud PF, Charron AJ, McDaniel SF, Khandelwal A, Quatrano RS (2009) Culturing the moss *Physcomitrella patens*. Cold Spring Harb Protoc 2:pdb.prot5136. https://doi.org/10.1101/pdb.prot5136
6. Wang X, Qi M, Li J, Ji Z, Hu Y, Bao F, Mahalingam R, He Y (2014) The phosphoproteome in regenerating protoplasts from *Physcomitrella patens* protonemata shows changes paralleling postembryonic development in higher plants. J Exp Bot 65(8):2093–2106. https://doi.org/10.1093/jxb/eru082
7. Ziganshin R, Arapidi G, Azarkin I, Zaryadieva E, Alexeev D, Govorun V, Ivanov V (2011) New method for peptide desorption from abundant blood proteins for plasma/serum peptidome analyses by mass spectrometry. J Proteome 74(5):595–606. https://doi.org/10.1016/j.jprot.2011.01.014
8. Rappsilber J, Mann M, Ishihama Y (2007) Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. Nat Protoc 2(8):1896–1906. https://doi.org/10.1038/nprot.2007.261

# Chapter 28

# The Bright Future of Peptidomics

Peter D.E.M. Verhaert

## Abstract

In this final chapter I project my personal perspective on the future of peptidomics. A bird's eye view is shed on the discipline and a bid is made to frame it in the broader arena of the life sciences of tomorrow. Inferring from its present state-of-the-art and from the general direction of some evolutionary trends which are to be discerned, a case is made that peptidomics enjoys full ripeness as a young branch of science today, from which a bright future for the discipline can be predicted.

**Key words** Peptidomics, Present status, Future perspectives

## 1 Introduction

The field of peptidomics is nearly two decades old. Appropriate timing and a great opportunity to write down my personal vision on how I imagine it may develop in the near future, extrapolating from its present status (such as nicely documented in Chapters 1 and 8 of this book). At such occasion it is tempting to jot down a set of provocative statements which come to mind when inferring from the recent evolution and developments observed in the peptidomics arena itself and in related disciplines.

I would like to refer the interested reader to my personal historic view on the initial years of the domain [1], which was published a decade ago. At that time, I also made some predictions into the future (see also [2]), and it is interesting to note to date that some of them were proven to be rather accurate. This includes our prophecy that with the enhanced analytical power (sensitivity and specificity) of mass spectrometry (MS) technology in particular, many more neuropeptides were still to be discovered. Whereas in 2005 it was stated that, with a mere 15 peptides being fully sequenced, the neuropeptidome of a specific species of insect was largely known [3], the very same research group published 5 times more peptides of that same species barely 5 years later [4], whereas as of today the number of different peptides detectable in this

Michael Schrader and Lloyd Fricker (eds.), *Peptidomics: Methods and Strategies*, Methods in Molecular Biology, vol. 1719, https://doi.org/10.1007/978-1-4939-7537-2_28, 

species exceeds 850 (S. Neupert and R. Predel, personal communication). I do not find it a surprising discovery that a multicellular organism as sophisticated as a cockroach makes use of many more than a few handful of peptides in its intercellular communication. Indeed the multitude of physiological functions requiring fine regulation and harmonization, is very comparable to that of *Homo sapiens*, a well-studied species with many hundreds of endogenous peptides documented [5].

## 2 The 2017 Sanibel Meeting on Peptidomics and the Current State-of-the-Art of the Field

Whereas the present book provides an excellent cross section on peptidomics as a mature scientific specialty today, its publication is very timely. In December 2016, the 29th edition of the annual international *Lake Louise Tandem MS Workshop* [6] featured a thematic session on "MSMS of Endogenous Peptides" in its main program, for the very first time. Shortly thereafter (January 2017), "Peptidomics" featured as focal theme of the 29th American Society for Mass Spectrometry Sanibel Conference on Mass Spectrometry [7]. Both meetings, but in particular the latter, nicely documented how blooming (and booming) the subject of peptidomics currently is.

Previously peptidomics scientists were to be spotted at technology (predominantly MS) meetings to learn about the latest developments in methods to analyze endogenous peptides. Today the reverse trend is to be observed, with MS technology developers attending true peptidomics meetings, to learn about the wide variety of life science problems (biological as well as medical research questions) which peptide MS can solve. This was finely exemplified by the 2017 Sanibel Conference [7]. Although being a typical MS focused conference, rather than being organized by analytical technology driven *mass spectrometrists*, the 2017 organizers all shared a principal background in academic life sciences research: P. Andrén (pharmaceutical neurobiology; Uppsala University), L. Li (invertebrate (crustacean) physiology; University of Wisconsin), and J. Sweedler (comparative neurobiology, University of Illinois). It was remarkable to see how a meeting concentrating on biological issues attracted expert MS scientists from all over the globe as they are gradually understanding the high relevance of the topic.

Thus, the Sanibel peptidomics meeting was honored to have prominent names in MS(/MS) fundamentals as well as instrumentation among its active participants (R.M. Caprioli, R.M.A. Heeren, J. Henion, D.F. Hunt, M. Vestal, to name a few). This was unquestionably very appropriate because tandem MS has been one of the elemental technologies driving the field forward, ever since its conception two decades ago [8]. Illustrative of the vigor and fitness of

present day peptidomics was the healthy mix at the meeting of numerous dynamic young researchers with a great core of highly active peptidomics pioneers, including the Sanibel organizers, along with world renowned researchers in comparative biology. The program featured some of the early adopters of MS for the analysis of natural peptides (D.M. Desiderio, L.D. Fricker, R. Kennedy, A.T. Lebedev, S. Rubakhin, A.O. Stretton, and B. Ueberheide). To ensure that the complete field of peptidomics was well covered, it was supplemented with experts in relevant subdisciplines. These included endogenous peptide bioinformatics (N. Bandeira [5]), biochemical peptide processing (V. Hook [9]), and disease related posttranslational modifications such as glycans (R.R. Drake [10]).

The evident shift of today's emphasis in peptidomics from fundamental, technological developments into more and more applications reflects the field's maturity. These comprise both general biological and medical applications, including pharmaceutical and food science. It is my vision that in the years to come this trend will continue with an ever-increasing implementation of peptidomics approaches in the life sciences (medicine and health care in particular), and with a consequent escalating impact on society.

## 3 Overlap with Other "-Omics," and Resulting Trends

The tag line of the above discussed Sanibel Meeting was "Bridging the Gap between Proteomics and Metabolomics by Mass Spectrometry." It is clear that peptidomics falls almost literally in between these disciplines. The molecular size of the analytes is between that of typical proteins and that of metabolites and other small molecules. In the past this has led to the neglecting of peptides in both adjacent disciplines. Peptides were "too small" for (bottom-up) shotgun proteomics and "too big" for conventional metabolomics. Some peptides are direct gene translation products, hence fulfilling the definition of a "protein," while many others are products of enzymatically catalyzed metabolic reactions, and therefore qualify as "metabolites" per se.

Although many of the same general techniques are employed in both proteomics and metabolomics, the historic limitations and strengths of these analytical technologies have led to diverging analytical workflows for both fields. These go beyond the mere settings of the instrumental hardware (first and foremost mass spectrometers) which understandably are markedly different (especially mass range) enabling *proteomicists* or *metabolomicists* to focus on their respective types of analytes.

Sample preparation is obviously tailored toward the biomolecules that are the focus of the study. Different physicochemical extraction methods have been boosted for metabolites and proteins and

various bioanalytical protocols optimize the detection and quantification of the analytes under investigation. In metabolomics, this frequently requires derivatization of the small molecule to better ionizable chemical variants. In proteomics, on the other hand, the focus has been on generating "proteotypic" fragments representative of each of the proteins present in a biological sample mix, the so-called bottom-up approach. Indeed, the intact proteins themselves were often found to be too large precursor ions for efficient (top-down) collision induced dissociation-based tandem MS fragmentation. Trypsin appeared to be the ideal sequence specific protease to generate smaller size and well-ionizable representatives of proteins, even in highly complex samples, rendering them analyzable with the required mass accuracy and resolution to enable the identification of the original proteins present.

These divergent workflows evidently entail subsequent different "post acquisition" data processing and interpretation strategies/algorithms which have further developed in two fundamentally different directions. In metabolomics, much analytical attention has gone to the extreme mass accuracy one can relatively easily obtain on biomolecules of the small size typical for the average metabolite. From an accurate mass of a metabolite, its elemental composition can be directly deduced, and together with an additional tandem MS signature, this will be sufficient to positively identify its chemical structure. Compiling this information in well annotated and as complete as possible databases (spectral libraries), has absorbed a lot of energy in the latest metabolomics advancements (see, e.g., [11, 12]). It will be only a matter of time before a similar effort will be initiated to establish an endogenous peptide centered data repository encompassing high resolution single stage MS data and (partial) de novo MS/MS spectra.

The dominating bottom-up, a.k.a. shotgun, approach in proteomics has not been a good thing for *peptidomicists*. On the contrary, peptidomics has really suffered from this. The favorite analytes of the peptidomicist (i.e., the low abundant endogenous peptides) are virtually rendered inaccessible by the brute force bottom-up proteomics approaches, in which they get completely snowed under by a haystack of easily ionizable tryptic peptides. This is why peptidomicists warmly embrace the current flourishing of top-down proteomics [13]. Thanks to the development of novel fragmentation techniques, and their implementation in more and more universally available high performance MS instruments, the necessity to digest proteins prior to a successful analyte identification is becoming less of an absolute requirement. This is a true blessing for the peptidomics scientists. I would dare to state here that, as invaluable as they are, bottom-up proteomics tactics are a thing of the past, having been mainly driven by the technological challenge to attempt a complete proteome characterization of a biological system, rather than addressing a specific biological ques-

tion in it. As such, the focus on the biologically relevant protein isoforms, the so-called **proteoforms,** has often been lost. I would argue that the endogenous peptides may well be among the most important proteoforms in any biological system, which hence deserve encouraged attention in the future. Top-down proteomics tactics will make endogenous peptides much more readily accessible.

From the above it shall be clear that the future of peptidomics will undoubtedly benefit from developments in both metabolomics and proteomics. Nonetheless, I believe that more technologies/tools should be developed that are specific for peptidomics itself.

## 4 Peptidomics Specific Future Trends

Top-down proteomics (thus *a fortiori* peptidomics) requires, besides a mass measurement as accurate as possible, additional information for analyte sequence elucidation, preferably tandem MS fragmentation data, which allow de novo sequencing. To increase the sequence coverage, the outcome of multiple MS/MS fragmentation strategies (CID/HCD, ECD/ETD, UVPD) may be combined. And ultimately all this information should be compiled in (tandem mass) spectral libraries, which need to be set up. Furthermore software will need to be developed to facilitate the analysis of isotopically resolved data and PTMs.

It is encouraging to see that institutes which provide public search engines such as the Swiss Institute of Bioinformatics (Geneva, Switzerland) as well as companies like BioInformatics Solutions (Waterloo, ON, Canada) have expressed their interest in a renewed analytical focus on endogenous, non-tryptic, peptides. Hence the development of novel bioinformatics search procedures, as well as of innovative de novo sequencing algorithms and workflows, can be expected. I imagine that these will also include negative ionization strategies, in particular for highly acidic, as well as other endogenous peptides which are poorly ionizable in positive mode. Until today these are still very much unexplored *terra incognita*.

In addition, a continual updating and an integration of the currently available peptidomics databases is timely. In general, the field will profit from an up-to-date integrated endogenous peptide database completed with peptide spectral libraries (incorporating both MS and $MS^n$ spectra from different charge states in a variety of modern instruments).

Over the years various peptidomics databases have been created, including SwePep [14], NeuroPedia [5], NeuroPep [15], neuropeptides.nl [16], although not all have been updated frequently. In analogy to UniProt as successor of SwissProt one could picture “UniPep” as successor of the above peptidomics databases.

This should ideally integrate all non-redundant information collected in the separate slightly differently flavored peptide collections into a fully searchable database complete with spectral libraries and the like, including PTMs, variants, and enzymatic processing. In analogy to NextProt, which is focused on *Homo sapiens*, and which is still very much (bottom-up) proteomics oriented, one could visualize a peptidomics equivalent "NextPep," which should focus on the fully processed endogenous human peptides and their natural posttranslationally modified variants ("*peptiforms*"), including L/D conversions, truncations, disulfide bridges, and the typical PTMs such as phosphorylation, sulfation, acetylation, methylation, amidation, glycation, lipidation, as well as typical chemical modifications like oxidation and deamidation. This would also combine different levels of data (genomic, transcriptomic, proteomic, tissue/disease specificity, etc). As such, it would be great to have a link with the Human Protein Atlas, which could accommodate a Human Endogenous Peptide Sub Atlas, which would structure all the information on the tissue and cellular specificity of the NextPep information, including *Mass Spectrometry Histochemistry* images of each endogenous peptide for all human reference tissues. By itself, out of the *Homo sapiens* "UniPep" lexicon, the "vocabulary" of each and every tissue will be instituted. It shall be clear that this is a tremendous endeavor, as it is highly likely that the peptide messages sent out by cells will differ both in health and disease. Crosslinks with related databases like SATPdb (a database of structurally annotated therapeutic peptides [17]) are much desired.

In an ideal world, a unified peptidomics database will be a subdatabase of a global top-down proteomics database. For this I strongly advocate teaming up the peptidomics and top-down proteomics communities/consortia to agree on the best strategy to integrate their respective data. Indeed, keeping both databases separate on the basis of the size of the analyte as only discriminating factor between a proteoform (protein) and a peptiform (peptide) would be very arbitrary and artificial with very little if any biological meaning or relevance whatsoever.

I am looking forward to the next generation SwePep/NeuroPedia/NeuroPep, ideally integrated with one another and with increased user friendliness. In my vision this means that over time for each systematic taxon in the tree of life, a lexicon will be compiled which will culminate in a complete endogenous peptide dictionary per species.

## 5 Other Promising Evolutions and Developments

The extensive work by Andren and coworkers clearly showed that the analysis of endogenous (neuro)peptides, with their often very labile PTMs is very challenging [18]. A lot of benefit is achieved by

analyte stabilization strategies, including physical ones such as rapid freezing, or preferably, heat stabilization [19], the latter being irreversible, in contrast to the former. I anticipate that these as well as alternative stabilization tactics, including chemical ones will still be optimized in the future and/or find their way into the field.

Our personal successful neuropeptide imaging (Fig. 1) as well as tandem MS sequencing in insect neuronal tissues which were formalin fixed over 30 years ago [20] suggest that such chemical fixation may be an effective sample conservation step in (neuro) peptidomics, at least for a subset of relatively short peptides. In combination with the ever-increasing performance of the modern mass spectrometer systems (more sensitive ion sources, linked to more selective analyzers and more sensitive detectors), the de facto "filtering" of the MS analyzable molecular species by chemical fixation, such as formaldehyde cross linking, maybe a employed as a very beneficial sample "enrichment" step. The combination with the obtained spatial information of the peptides makes this a most valuable addition to other analytical methods available to help and fish out the endogenous peptide "needles" from the "haystacks" of proteins which are orders of magnitude more abundant.

In this respect the recently more intensely studied extracellular vesicle mediated intercellular communication [21] will deserve more interest in the future. Undeniably this nonconventional way of extracellular communication is a very interesting way of cells to increase their peptidergic "vocabulary," as in this way from one signaling cell/tissue/organ to a recepient cell/tissue/organ peptides can be transferred which do not have a typical signal sequence and hence are not typically released through the conventional secretory pathway via fusion of a secretory vesicle with the plasma membrane of the peptide synthesizing and secreting cell.

With secretory peptides being an important, if not the most important, means for cells to communicate, it is clear that also other types of intercellular communication exist. In order to fully understand the biomolecular languages of the living cell, future integration of peptidomic signaling with classical small molecular neurotransmitter signaling as well as with others will be necessary. As such I personally envisage big promise in assimilating peptidomics data in other global biomolecular characterization projects such as MetaSpace [22].

For intracellular functions, an important area will be the study of a promising subset of endogenous peptides: small open reading frame encoded (poly)peptides, which potentially function as intracellular messengers [23, 24]. These peptides require other analytical strategies to address (particularly to enable extraction from their intracellular compartments), and I trust that in the coming years these will be developed, as initiated recently [25]. In addition, non-ribosomal peptides, have remarkable pharmacological potential, and hence are a class of "endogenous" peptides which will be heavily studied in the years to come [26].

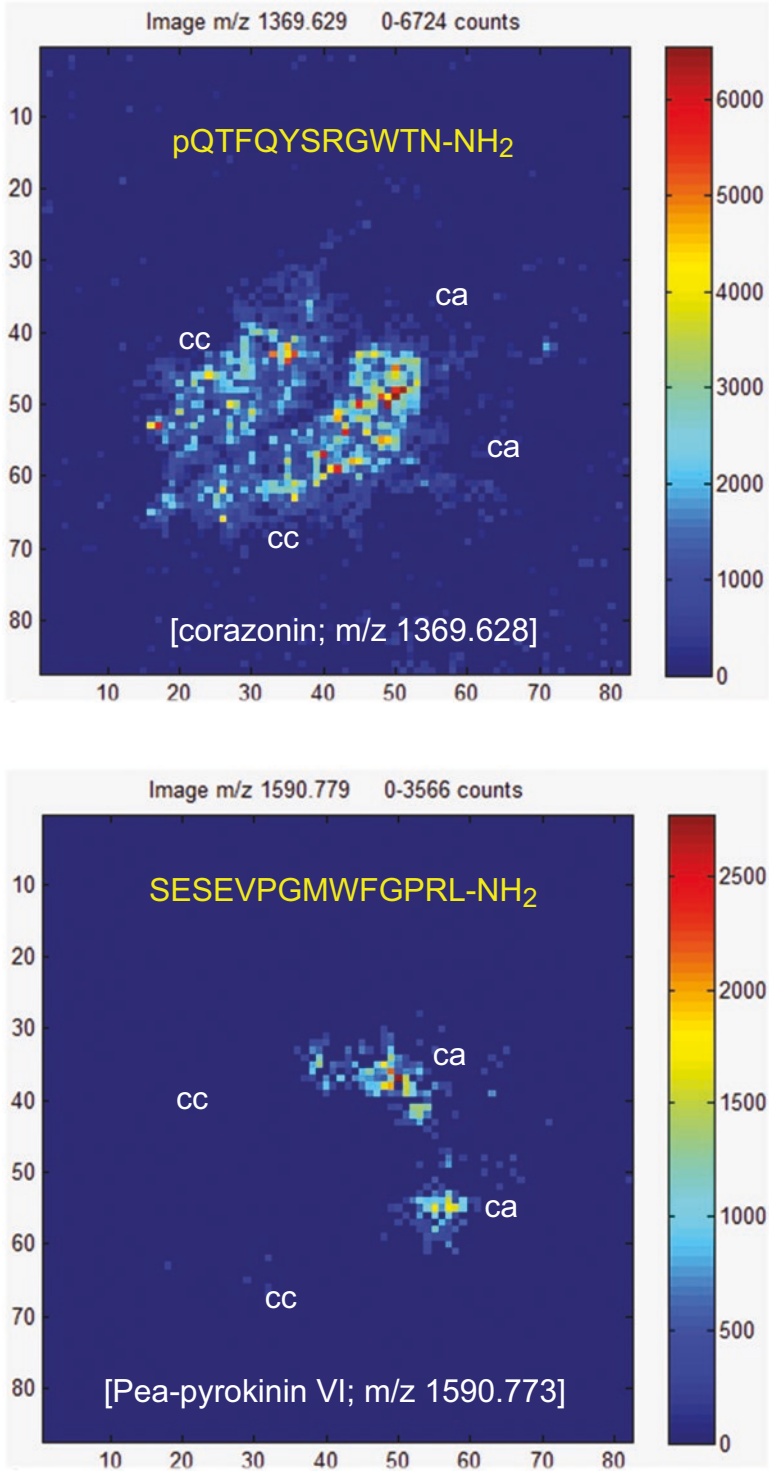



**Fig. 1** Neuropeptide mass spectrometry histochemistry (MSHC) on FFPE tissue. Section through cockroach (*Periplaneta americana*) retrocerebral neurohemal organ (pituitary equivalent). (**a**) MSHC image of "corazonin," with distribution primarily in the corpora cardiaca (cc), (**b**) MSHC image of "Pea pyrokinin VI," with localization restricted to the corpora allata (ca) region

Personally, I see a highly promising contribution for peptidomics in trying to disclose the secrets hidden in the many thousands and thousands of clinically stored biobanked materials. More effort should be devoted to the development and large scale employment of analytical tools to translate the wealth of clinical knowledge buried in medically defined biobanked material into biomolecular, especially peptidomic signatures/profiles. I predict that peptidomic mining of freshly frozen (cryosections), but unexpectedly also of formalin-fixed paraffin-embedded (FFPE) tissues, will become accessible for peptidomics analysis, including mass spectrometry imaging, i.e., mass spectrometry histochemistry. I am hopeful that the current situation in which MS imaging has drifted away from endogenous peptides, in favor of lipids, metabolites, and small molecules, which are more easily identified "on the fly," is about to change.

I anticipate that such peptidomics analysis will promote the increasing appearance of peptides in the drug industry (diagnostics, "prognostics," therapeutics), as peptidomics data will uncover a lot of the biology of peptides from studying examples out of nature, not only on the multitude of functions of peptides, but also on ways of increasing their drugability [27].

Future technological developments (with ever-increasing sensitivity), will unquestionably make the peptidome analysis of biological samples even less invasive than it can already be today. This will remove some of the existing bioethical barriers for certain types of experimental peptidomics. As such, the need for the employment of model organisms in order to try and understand human biology/physiology/pathology/disease etiology will further diminish. Concurrently, peptidomics will get a more profound embedding in the study of human biology, pharmacology, medicine, and health care.

## Acknowledgments

The author is most grateful to Prof. P. Andren for his stimulating comments and expert advice on the outline and early versions of the manuscript. Figure 1 was produced from data acquired through a very much-appreciated collaboration with Dr. M. Paine, Dr. S. Ellis and Prof. R. Heeren. Equally many thanks to the editors of this volume, Profs. L. Fricker and M. Schrader, for their invaluable help with the final shaping of the paper.

### References

1. Verhaert PDEM, Pinkse MWH, Prieto-Conaway MC, Kellmann M (2008) A short history of insect (neuro)peptidomics—a personal story of the birth and youth of an excellent model for studying peptidome biology. In: Soloviev M, Andren P, Shaw C (eds) Peptidomics methods and applications. Wiley, Hoboken, pp 25–54
2. Shaw C, Verhaert P (2008) Peptidomics and biology. Two disciplines driving each other. In: Soloviev M, Andren P, Shaw C (eds) Peptidomics methods and applications. Wiley, Hoboken, pp 389–396

3. Predel R, Gaede G (2005) Peptidomics of neurohemal organs from species of the cockroach family Blattidae: how do neuropeptides of closely related species differ? Peptides 26:3–9
4. Neupert S, Fusca D, Schachtner J, Kloppenburg P, Predel R (2012) Toward a single-cell-based analysis of neuropeptide expression in *Periplaneta americana* antennal lobe neurons. J Comp Neurol 520:694–716
5. Kim Y, Bark S, Hook V, Bandeira N (2011) NeuroPedia: neuropeptide database and spectral library. Bioinformatics 27:2772–2773
6. www.lakelouisemsms.org
7. www.ASMS.org
8. Manguy J, Jehl P, Dillon ET, Davey NE, Shields DC, Holton TA (2017) Peptigram: a web-based application for peptidomics data visualization. J Proteome Res 16:712–719
9. Hook V, Bandeira N (2015) Neuropeptidomics mass spectrometry reveals signaling networks generated by distinct protease pathways in human systems. J Am Soc Mass Spectrom 26:1970–1980
10. Drake RR, Powers TW, Jones EE, Bruner E, Mehta AS, Angel PM (2017) MALDI mass spectrometry imaging of N-linked Glycans in cancer tissues. Adv Cancer Res 134:85–116
11. https://omictools.com/metabolite-libraries-category
12. metabolomicssociety.org/resources/metabolomics-databases)
13. www.topdownproteomics.org
14. Fälth M, Sköld K, Norrman M, Svensson M, Fenyö D, Andren PE (2006) SwePep, a database designed for endogenous peptides and mass spectrometry. Mol Cell Proteomics 5:998–1005
15. Wang Y, Wang M, Yin S, Jang R, Wang J, Xue Z, Xu T (2015) NeuroPep: a comprehensive resource of neuropeptides. Database (Oxford) 2015:1–9. https://doi.org/10.1093/database/bav038
16. www.neuropeptides.nl
17. Singh S, Chaudhary K, Dhanda S, Bhalla S, Usmani S, Gautam A, Tuknait A, Agrawal P, Mathur D, Raghava G (2016) SATPdb: a database of structurally annotated therapeutic peptides. Nucleic Acids Res 44:D1119–D1126
18. Sköld K, Svensson M, Norrman M, Sjögren B, Svenningsson P, Andrén PE (2007) The significance of biochemical and molecular sample integrity in brain proteomics and peptidomics: stathmin 2-20 and peptides as sample quality indicators. Proteomics 7:4445–4456
19. Secher A, Kelstrup CD, Conde-Frieboes KW, Pyke C, Raun K, Wulff BS, Olsen JV (2016) Analytic framework for peptidomics applied to large-scale neuropeptide identification. Nat Comm 7:11436. https://doi.org/10.1038/ncomms11436
20. Verhaert PD, Ellis SR, Paine MRL, Maloney D, Heeren RMA (2017) Top-down neuropeptide MALDI imaging MS on FFPE sections with high mass resolution and MS/MS capabilities: towards true "Mass Spectrometry Histochemistry". Proceedings of the 65th ASMS Conference on Mass Spectrometry and Allied Topics, Indianapolis, IN, USA, June 4–8, 2017
21. Yáñez-Mó M, Siljander PR-M, Andreu Z, Zavec AB, Borràs FE, Buzas EI, Buzas K, Casal E, Cappello F, Carvalho J, Colás E, Cordeiro-da Silva A, Fais S, Falcon-Perez JM, Ghobrial IM, Giebel B, Gimona M, Graner M, Gursel I, Gursel M, Heegaard NHH, Hendrix A, Kierulf P, Kokubun K, Kosanovic M, Kralj-Iglic V, Krämer-Albers E-M, Laitinen S, Lässer C, Lener T, Ligeti E, Linē A, Lipps G, Llorente A, Lötvall J, Manček-Keber M, Marcilla A, Mittelbrunn M, Nazarenko I, Nolte-'t Hoen ENM, Nyman TA, O'Driscoll L, Olivan M, Oliveira C, Pállinger É, Del Portillo HA, Reventós J, Rigau M, Rohde E, Sammar M, Sánchez-Madrid F, Santarém N, Schallmoser K, Stampe Ostenfeld M, Stoorvogel W, Stukelj R, Van der Grein SG, Vasconcelos MH, Wauben MHM, De Wever O (2015) Biological properties of extracellular vesicles and their physiological functions. J Extracell Ves 4:27066
22. Palmer A, Trede D, Alexandrov T (2016) Where imaging mass spectrometry stands: here are the numbers. Metabolomics 12:107. https://doi.org/10.1007/s11306-016-1047-0
23. Crappé J, Van Criekinge W, Menschaert G (2014) Little things make big things happen: a summary of micropeptide encoding genes. EuPA Open Proteom 3:128–137
24. Mouilleron H, Delcourt V, Roucou X (2016) Death of a dogma: eukaryotic mRNAs can code for more than one protein. Nucl Acids Res 44:14–23
25. Ma J, Diedrich JK, Jungreis I, Donaldson C, Vaughan J, Kellis M, Yates JR, Saghatelian A (2016) Improved identification and analysis of small open reading frame encoded polypeptides. Anal Chem 88:3967–3975
26. Sable R, Parajuli P, Jois S (2017) Peptides, peptidomimetics, and polypeptides from marine sources: a wealth of natural sources for pharmaceutical applications. Mar Drugs 15:124. https://doi.org/10.3390/md15040124
27. Uhlig T, Kyprianou T, Martinelli FG, Oppici CA, Heiligers D, Hills D, Ribes Calvo X, Verhaert P (2014) The emergence of peptides in the pharmaceutical business: from exploration to exploitation. EuPA Open Proteom 4:58–69

# INDEX


## A

Acetylation
N-terminal.....102, 218, 230, 364, 380, 389, 390, 392
Acid precipitation.....227
Affinity
chromatography.....128, 129, 203, 211, 382
column.....17, 202, 213, 214
Algorithms.....20, 22, 23, 67, 69, 83, 143, 151, 155, 156, 183, 210, 294, 296, 308, 401, 410, 411
Alignments.....74, 75, 78, 79, 84–86, 88, 92, 189, 218, 296, 361, 364
Alkylation.....52, 54, 68, 264, 337, 380, 390, 392, 401
Amide
C-terminal.....7, 16, 57, 200
Amino acid sequence.....7, 11, 14, 84, 104, 195, 235, 273, 308, 328, 331, 359, 370, 372, 385, 386
Aminopeptidases.....108–110, 403
Anhydrotrypsin.....127, 202–206
Animal toxins.....15, 16, 335
Antibacterial peptides.....202
Antimicrobial
activity.....67, 223, 236, 319, 328
peptide.....15, 16, 24, 68, 70, 224, 236
*Astatotilapia burtoni*.....73, 76–78, 80, 81, 83, 85, 86, 88, 90
Autolysis.....298, 308
Automated gain control (AGC).....63–65, 217, 245, 391
Autosampler.....299, 313, 314, 316, 338, 339, 363

## B

Basic residues.....128, 129, 200–202
Best Linear Alignment Software Tool (BLAST™).....74, 77, 80, 248, 355
Biochemical assays.....16, 17
Bioinformatics.....10, 22, 67, 71–94, 109, 224, 225, 234, 245, 273, 317, 338, 361, 363, 364, 384, 409, 411
Biological activity.....3, 7, 107, 127, 236, 249, 335, 349
Biomarker
discovery.....17, 18, 21, 187, 312
peptides.....6, 16–18, 23, 187, 312
Body fluid.....4, 12, 13, 15, 18, 19, 23, 312
*Bothrops*.....350, 352, 355
Bottom-up proteomics.....4, 8, 22, 142, 237, 409, 410, 412
Bradford assay.....228
Brain tissue.....41–44, 206, 207
Butelase.....380, 382, 385–391

## C

*Caenorhabditis elegans*.....156, 202, 271–283, 285, 288, 289
Calibration standards.....327
*Cancer borealis*.....248, 253
Capillary electrophoresis (CE).....19, 107, 263, 264, 312
Carbamidomethylation.....364
Carboxypeptidase
carboxypeptidase D (CPD).....200, 201
carboxypeptidase E (CPE).....17, 128, 129, 137, 201, 202, 207
carboxypeptidase E-deficient mice (Cpe$^{fat}$).....201, 203
*C. elegans*, *see Caenorhabditis elegans*
Cell culture.....52, 55, 99, 100, 103, 122, 126, 134, 321–324, 382
Cell dissection.....371, 373–375
Cell isolation.....369
Centrifugal filter device.....43, 45
Cerebrospinal fluid (CSF).....11, 15, 18, 23, 193, 194, 302, 303
Charge states.....64–66, 69, 129, 168, 217, 228, 338, 342, 353, 404, 411
Cheese.....23, 224, 226, 227, 237
Chirality.....108, 112, 113
Circulating fluid.....248
Cleavage
enzymatic.....11, 71, 223, 366
prediction.....68, 88, 94, 234
site.....13, 22, 51, 54, 56, 68, 72, 88–90, 93, 94, 168, 201, 234, 294, 296–298, 301, 303
*Clitoria ternatea*.....380, 381, 384
Cloning.....385–388
Clustal Omega™.....74, 79, 86, 88
Collision energy (CE).....63, 64, 116, 154, 155, 217, 228, 230, 245, 308, 340, 342, 353, 354, 363, 404
Collision-induced dissociation (CID).....10, 20, 63–65, 67, 155, 237, 245, 264, 353, 354, 357, 359, 363, 410, 411
Computational analysis.....218
Conditioned medium.....52, 53, 55



Michael Schrader and Lloyd Fricker (eds.), *Peptidomics: Methods and Strategies*, Methods in Molecular Biology, vol. 1719, https://doi.org/10.1007/978-1-4939-7537-2, © Springer Science+Business Media, LLC 2018

CONSTANd™ ..... 143, 145–147, 150–153, 156, 157
Cpe^fat, *see* Carboxypeptidase
Crab ..... 250, 252, 253, 256, 260–262
Cross-linking ..... 214, 219, 413
Crustacean ..... 242, 408
Cryostat ..... 257
C-terminus ..... 16, 17, 58, 201
Cultured cells ..... 53, 56, 134, 169
Culture medium ..... 53, 55, 323
Cyanoborohydride ..... 130, 161, 162, 164, 169–171
Cyclotides ..... 24, 379, 380, 382, 383, 389–391
Cysteines ..... 16, 68, 116, 142, 155, 336, 340, 359, 363, 379, 403

## D

Dairy product ..... 23, 224, 226, 238
D-amino acid ..... 14, 117, 320
Danio rerio, *see* Zebrafish
Data
  analysis ..... 22, 63, 65, 109, 110, 125, 132, 134, 154, 163, 173, 188, 189, 193–195, 225, 249, 251, 254, 256, 259, 260, 273, 312, 313, 315, 317, 338, 340–342, 350, 352, 391, 401
  conversion ..... 188, 189
  mass spectrometric ..... 188–191, 193–195
  mzML ..... 188, 189, 364
  preprocessing ..... 187–195
  processing ..... 21–23, 66, 188, 191, 251, 267, 285, 313, 315, 316, 397, 401, 410
  reduction ..... 188, 189, 191
  visualization ..... 23, 52, 188, 189, 191, 193–195
Database
  gene ..... 75, 77, 78, 80
  genomic ..... 11, 83, 242, 248, 265
  NCBI ..... 74, 77, 78, 245, 352, 353, 389
  peptide ..... 22, 63, 68, 102, 111, 125, 133, 150, 156, 205, 229, 236, 248, 249, 296, 308, 355, 366
  peptidomics ..... 11, 22, 125, 157, 229, 248, 411, 412
  protein ..... 11, 12, 22, 75, 77, 79–81, 83, 85, 86, 102, 157, 229, 232, 297
  search ..... 9, 14, 22, 63, 77, 79–81, 83, 111, 125, 168, 205, 218, 229, 236, 237, 245, 265, 287, 352, 353, 355, 357, 360, 364, 366, 412
  sequence ..... 11, 22, 74, 77, 86, 102, 242, 248, 296, 353, 355, 356, 366, 372, 401
  Uniprot™ ..... 74, 313, 335, 336, 352, 353
Data-dependent
  acquisition ..... 152, 245, 300, 356, 363
  mode ..... 170, 228, 363
Defensin ..... 16, 379
Degradation
  enzymatic ..... 16, 24, 311, 403
  post-mortem ..... 44
  proteolytic ..... 4, 11, 14, 42, 44
Degradomics ..... 294, 356
Delipidation ..... 244, 286, 289
Denaturation ..... 5, 43, 44, 57, 219, 307, 366
De novo analysis ..... 355, 356, 364
De novo sequencing ..... 14, 20, 22, 132, 251, 256, 264, 265, 287, 312, 317, 336, 340, 342, 350, 355, 357, 360, 370, 376, 411
Derivatization ..... 10–12, 108, 112, 116, 264, 373, 376, 378, 389–391, 410
Desalting ..... 55, 104, 115, 147, 148, 150, 155, 166, 170, 226, 228, 244, 253, 254, 257, 266, 316, 365, 384, 388, 390, 392, 400, 401
Detergent ..... 55, 68, 204, 207, 226, 307, 322, 326
Deuterated reagent ..... 162
Differential peptidomics ..... 141, 142, 148–152, 154
Digestion ..... 4, 11, 42, 63, 97, 108, 115, 142, 223, 224, 226, 227, 237, 246, 262, 265, 294, 359, 360, 362–364, 366, 382, 387, 390, 410
Dimethylpimelimidate ..... 211, 214, 219
Disulfide bridge ..... 16, 24, 68, 336, 337, 341, 342, 412
Dithiothreitol (DTT) ..... 60, 61, 68, 361, 363, 380, 390
*Drosophila melanogaster* ..... 374

## E

Edman degradation ..... 273, 320, 328
Electron transfer dissociation (ETD) ..... 9, 20, 61, 63–67, 264, 357, 391
Electrospray ionization (ESI)
  mass spectrometry (ESI-MS) ..... 8–12, 20, 23, 99, 102, 178, 182, 183, 203, 255, 256, 351, 353–355, 357, 401
  quadrupole time-of-flight (ESI-Q-TOF) ..... 154, 351, 354
Endopeptidase ..... 17, 72, 200, 201, 238, 380
EndoProteoFASP, *see* Sample preparation
ESI-TOF, *see* Electrospray mass spectrometry
ETD, *see* Electron transfer dissociation (ETD)
Ethyl acetate ..... 61, 62, 243, 244
Extraction ..... 3, 44, 97, 108, 143, 165, 175, 187, 199, 210, 224, 242, 251, 284, 299, 312, 330, 350, 360, 371, 381, 382, 396

## F

False discovery rate (FDR) ..... 67, 144, 218, 246, 297, 315, 391
Filter-aided sample preparation (FASP), *see* Sample preparation
FASTA format ..... 68, 229, 235, 236
Fluorescamine ..... 100, 104, 162, 163, 165, 166, 169
Fluorescence
  detection ..... 169
  dye ..... 371, 373, 374, 377
Folch liquid-liquid extraction ..... 227
Food peptidomics ..... 23

Formaldehyde ..... 130, 131, 134–136, 161–164, 166, 167, 169–171, 249, 250, 263, 264, 413
Fourier transform ion cyclotron resonance (FT-ICR) ..... 9, 317
Fourier transform-mass spectrometry (FT-MS) ..... 391
Fractionation ..... 133, 217, 263, 323, 337–339, 351, 353, 354, 365, 381, 382, 390, 392
Fragmentation ..... 10, 14, 16, 20–22, 59, 63–65, 67, 125, 131–133, 138, 143, 146, 217, 228, 245, 264, 287, 338, 340, 342, 343, 353, 354, 356, 357, 359, 370, 375, 376, 378, 410, 411
Fragmentation pattern ..... 14, 16, 59, 113, 125
Frog skin ..... 15
FT-ICR, *see* Fourier transform ion cyclotron resonance (FT-ICR)
FT-MS, *see* Fourier transform-mass spectrometry (FT-MS)

## G

Gel filtration ..... 54, 98, 382, 403
Gel permeation ..... 52, 54, 56, 294
Gene ontology (GO) ..... 67, 296, 308, 315, 317
Gene prediction ..... 72, 74, 80, 91, 93
GeneWise™ ..... 74, 80, 82, 84, 86, 93
Genomics ..... 4, 6, 16, 24, 71–73, 75–81, 83, 85, 86, 88, 92, 93, 210, 218, 272, 273, 387, 395, 412
Gradient ..... 62, 69, 104, 110, 112, 137, 154, 170, 228, 237, 255, 265, 285, 287, 300, 314, 321, 330, 338, 342, 343, 353, 363, 366, 390, 391, 401, 404

## H

Half-live ..... 6, 14, 104, 320, 356
Heat stabilization ..... 43–47, 206, 261, 266, 413
Hemolymph ..... 248, 249, 252, 253, 255, 261, 262
Higher-energy collisional dissociation/ high-energy collisional dissociation (HCD) ..... 20, 63, 64, 154, 217, 237, 287, 340, 343, 411
High performance liquid chromatography (HPLC)
  column ..... 99, 100, 131, 162, 320, 321, 323, 330, 337
  purification ..... 110
High-throughput analysis ..... 59, 242, 244
HLA binding peptides (HLAp) ..... 209, 216, 217
Homology
  search ..... 77, 79–81, 83, 225, 236, 248, 355, 361, 364
Hormone ..... 6–8, 10, 12, 13, 15–17, 21, 22, 41, 77, 78, 179, 180, 184, 199, 201, 241, 247, 248, 311

## I

Immobilized metal ion affinity chromatography (IMAC) ..... 382, 389
Immunoaffinity ..... 212
Immunopeptidomics ..... 11, 24, 209–220
Insect
  cell culture ..... 382
  hemolymph ..... 175
Insulin secretion ..... 320, 330
Iodoacetamide ..... 52, 55, 60, 212, 261, 361
Isobaric labeling ..... 249
Isotope shift (NIS) ..... 340
Isotopic label ..... 9, 15, 20, 122, 126, 135

## L

Label
  -based ..... 141
  -free ..... 122, 124–126, 133, 135, 141, 205, 224, 230, 231, 233, 249, 265, 315, 337
  -free relative quantification ..... 125, 224
Labeling
  fruit fly ..... 177, 180, 181
  isobaric ..... 126, 133, 135, 143, 249
  isotopic ..... 9, 15, 20, 122, 126, 130, 131, 133–135, 142, 161, 162, 167, 217, 249
  metabolic ..... 175–184
  protocol ..... 155, 162, 163
  yeast ..... 176, 177, 179, 180
Laser energy ..... 255, 259, 264, 378
LC/ESI-MS, *see* Liquid chromatography (LC)
LC/MS, *see* Liquid chromatography (LC)
LC-MS/MS
  analysis ..... 54, 114, 147, 150, 353, 356, 363, 391
  fragmentation ..... 16, 52, 113, 125, 133, 142, 154, 228, 356, 411
  *See also* Liquid chromatography (LC)
Limited digestion ..... 97
Linear gradient ..... 69, 102, 182, 183, 217, 245, 314, 323, 339, 353
Linear Trap Quadrupole (LTQ) ..... 61, 183, 243–245, 251, 255, 256, 258–260, 305, 313, 315, 338, 361
Liquid chromatography (LC)
  electrospray ionization mass spectrometry (LC/ESI-MS) ..... 41, 99, 203
  mass spectrometry (LC/MS) ..... 163, 205
  nano (nanoLC/nano-HPLC) ..... 9, 41, 44, 46, 47, 109, 243, 305, 312–314, 317, 351, 397, 401
  strong cation exchange (SCX-LC) ..... 390
  tandem mass spectrometry (LC-MS/MS) ..... 41, 170, 202, 300, 312, 351, 381, 401
  ultra-high performance (UHPLC, UPLC) ..... 61, 109, 251, 284, 383
LysC ..... 142

## M

MALDI, *see* Matrix-assisted laser desorption/ionization (MALDI)
MALDIQuant™, *see* Matrix-assisted laser desorption/ ionization (MALDI)
MALDI-TOF/TOF, *see* Matrix-assisted laser desorption/ ionization (MALDI)
Marfey's reagent ..... 108, 109, 112, 116

Mascot™ ..... 52, 65, 102, 147, 151, 155, 156, 168, 171, 173, 203, 205, 243, 245, 285, 287, 289, 295, 297, 300, 352, 353, 357, 370
Mass accuracy ..... 8, 9, 20, 315, 322, 331, 337, 350, 356, 370, 410
Mass defect ..... 132, 336, 340
MassLynx™ ..... 102, 167, 352, 354, 355
Mass spectrometric technology ..... 14, 20, 21
Mass spectrometry (MS)
  data ..... 52, 63, 66, 125, 135, 136, 138, 189, 193, 194, 210, 224, 245, 285, 287, 297, 337, 353, 361, 364, 401, 410
  imaging ..... 21, 249, 257–260, 415
MATLAB™ ..... 338, 340, 342, 347
Matrix-assisted laser desorption/ionization (MALDI)
  imaging ..... 250, 258
  ion trap-orbitrap mass spectrometer ..... 251, 255
  MALDIQuant™ ..... 189, 190
  mass spectrometry ..... 251, 328, 375, 397, 401
  mass spectrometry imaging ..... 21
  matrix ..... 9, 250, 295, 299, 322, 326, 382, 392
  MS/MS ..... 202, 352–354
  tandem TOF mass spectrometry (MALDI-TOF/TOF) ..... 115, 178, 179, 182–184, 203, 264, 293, 295, 322
  target ..... 182, 372, 375, 401
MaxQuant™ ..... 210, 212, 218, 219, 251, 313, 315–317
Metabolic labeling ..... 176, 177, 179, 180, 184
Metabolomics ..... 409–411
Metalloproteinases ..... 350
Methylation ..... 130, 131, 134, 135, 161–172, 412
Microcolumn ..... 61, 397, 401
Microdialysis ..... 248, 249, 253, 262, 263
Microwave
  irradiation ..... 42, 43, 169, 207
  oven ..... 98, 99, 102, 203, 206
  reactor ..... 109, 111
Milk ..... 23, 223–238, 355, 361, 362
Model system ..... 241, 273
Molecular diversity ..... 380
Monoisotopic mass ..... 58, 167, 168, 173, 328, 336, 338, 342, 343, 346, 354
Moss ..... 24, 398, 402, 404
Mouse ..... 16, 17, 42, 44, 72, 127–129, 137, 202, 206
MS/MS, *see* Tandem mass spectrometry (MS/MS)
Multiple reaction monitoring (MRM) ..... 108, 109, 112, 113, 116
Multiplexing ..... 132, 142, 153, 263
Multivariate analysis ..... 188, 192, 193

## N

Nano-HPLC, *see* Liquid chromatography (LC)
NanoLC, *see* Liquid chromatography (LC)
Nanospray ..... 212, 226, 313, 391, 397
Nervous system ..... 175, 182, 247–249
Neuroendocrine cell ..... 248
Neuronal network ..... 54, 248, 369
Neuropeptide
  crustacean ..... 249, 261, 264, 265
  discovery ..... 11, 78, 84, 87, 94
  endogenous ..... 272, 274–283, 289
  homolog ..... 273
  hormone ..... 248
  identification ..... 71, 72, 90
  imaging ..... 249, 413
  intermediate ..... 201
  like protein (nlp) ..... 273, 277–280
  location of ..... 249
  precursor ..... 12, 68, 201, 202, 207, 289
  putative ..... 71, 72, 75, 370
Neuropeptidergic signaling ..... 273
Neuropeptidomics ..... 15, 41, 43, 44
NeuroPred ..... 72, 74, 88, 90, 91, 94
Neurotransmitters ..... 7, 41, 241, 311, 319, 413
Non-targeted method ..... 108
Non-targeted peptidomics ..... 187–195
N-terminal acetylation ..... 218, 230, 364, 380, 389, 390, 392
N-terminal derivatization ..... 370
N-terminus ..... 72, 108, 114, 115, 117, 127, 142, 168

## O

Oligopeptidase ..... 97–104
Orbitrap™ ..... 9, 20, 61, 63, 132, 133, 154, 183, 224, 226, 229, 231, 243–245, 251, 255, 256, 258, 260, 261, 265, 266, 284, 305, 337, 338, 356, 361, 383
Ovarian cancer ..... 214, 304–306

## P

Parallel computing ..... 188
Parathyroid hormone (PTH) ..... 76, 86
Partial least squares discriminant analysis (PLS-DA) ..... 315
Patch clamp ..... 320, 373, 374
Peak detection ..... 189, 190, 192, 193
PEAKS Studio™ ..... 109, 156, 352, 357, 361, 364, 384, 391
Peptidase ..... 4, 8, 13, 17, 21, 41, 68, 108, 169, 200, 245, 289, 305, 308, 320
Peptide
  bioactive ..... 4, 6–8, 13, 15–17, 23, 24, 42, 53, 55, 57, 86, 101, 107, 183, 200, 202, 223, 236, 320, 331, 336, 379
  biologically active ..... 223, 242, 273, 335, 355
  bradykinin potentiating (BPP) ..... 349, 350
  C-terminally amidated ..... 16, 54, 57
  cysteine containing ..... 335–347

cysteine-rich (CRP).....16, 341, 359, 380, 392
D-amino acid-containing.....107–117
derivatization ..... 10, 12, 264, 337, 370, 371, 373, 375, 376, 378, 382, 383, 389–391, 410
diastereomers.....107
digestion .....97, 109–111, 114, 115, 124, 223, 224, 262, 265, 294, 309, 360–363, 390
disulfide-rich .....12
endogenous.....3, 51, 59, 60, 108, 113, 114, 142, 210, 242, 272, 274–283, 289, 303, 312, 317, 336, 395–404, 408–413, 415
extraction .....8, 10, 12, 19, 25, 43–46, 57, 61, 109, 113, 115, 126, 155, 169, 177, 178, 180, 203, 204, 227, 228, 243, 245, 266, 286, 289, 312, 314, 337, 373, 374, 377, 398, 403
FMRFamide-like (*flp*) .....273–276
hormone .....6–8, 10–13, 15–17, 21, 22, 51, 57, 175, 199–202, 241, 247, 248, 311
identification.....10, 11, 14, 43, 60, 63, 67, 72, 109, 125, 142, 143, 210, 224, 225, 229, 242, 271–289, 294, 295, 308, 312, 315, 317, 331, 336, 353, 359–366, 380, 392
isomerization .....14, 107
labeled.....16, 126–132, 136, 137, 142–144, 148, 150, 154, 162, 163, 166, 167, 170, 172, 179, 256, 373
native .....3, 21, 42, 311, 359, 360, 362–364, 366
precursor.....10, 12, 16, 17, 24, 42, 51, 57, 59, 102, 124, 137, 142, 151, 152, 157, 171, 200–202, 217, 234, 242, 350, 354
processing .....13, 17, 22, 23, 51, 65, 67, 135, 136, 141, 157, 183, 200–202, 242, 285, 312, 350, 409
profiling.....10–12, 176, 248, 249
purification .....7, 8, 11, 61, 102, 163, 165, 166, 177, 178, 199, 201, 204, 205, 211, 214–216, 321, 330, 331
salivary..... 293, 294, 296
secretory.....53, 55, 56, 72, 171, 200, 201, 349, 413
sequence.....10, 11, 20, 55, 56, 72, 90, 115, 117, 234, 236, 242, 248, 274–283, 349, 354, 366, 370
signaling ..... 6, 175, 176, 199, 200, 247, 395, 413
sulfur containing.....336
venom.....339, 342, 350, 355, 360, 362, 363, 365, 366
Peptidergic signaling .....175, 242
Peptidome
cerebrospinal fluid (CSF) ..... 12, 23, 301, 304
crustacean ..... 248, 249, 260, 267
human.....18, 24, 65, 172, 293, 412
mouse ..... 16, 127, 205
plant.....21, 24, 379–392, 396
-protease profile.....297, 298, 302, 303, 306, 309
salivary.....293, 294
snake venom .....349–357
urine .....311–317
zebrafish .....241–246
*Periplaneta americana*..... 370, 373, 414
Phosphorylation ..... 14, 43, 121, 171, 194, 230, 232, 412
Phylogenetic tree .....88, 89
PicoTip™ emitter.....313, 361
Plant
cysteine-rich peptide (CRP)..... 16, 380, 392
peptide, 16, 23, 24, 379, 395, 398, 403
Post-mortem degradation.....204, 261
Post-source decay (PSD) .....20
Post-translational modification (PTM)..6, 13, 14, 16, 72, 88, 107, 156, 157, 256, 265, 312, 317, 336, 361, 364, 380, 391, 411, 412
Precursor
ion.....57, 217, 238, 245, 287, 300, 315, 353–356, 363, 391, 410
protein .....17, 24, 51, 58, 71, 194, 201, 349, 350
single-charged .....201, 245
Preproproteins .....242, 273
Probability scores.....68, 70
Processing intermediates .....17, 200–202
Progenesis™ LC-MS .....147, 150
Prohormone.....200, 201
convertase 1 (PC1/3)..... 137, 200, 201
convertase 2 (PC2) .....137, 200
convertase 3 (*see* convertase 1 (PC1/3))
family.....76, 77, 79, 81, 86–88, 93
gene .....72, 75–80, 88, 89, 91–93
sequence..... 72, 74–76, 79, 84, 86, 88, 93
Protease .....352
endogenous.....56, 294, 308, 312, 395
inhibition.....403
inhibitor cocktail..... 226, 261, 381, 382, 396, 398, 403
inhibitors .....4, 56, 212, 226, 227, 237, 261, 306, 308, 352, 355, 356, 359, 366, 382, 396, 398, 403
natural.....312
salivary.....293–309
Proteasix™ .....294, 296, 298, 300–306, 308
Proteasome fragment.....201
Proteinase, *see* Protease
Protein lysis, *see* Proteolysis
Proteolysis .....23, 51, 54, 303, 355, 396
Proteolytic cleavage .....13, 42, 124, 142, 200
Proteolytic enzymes.....22, 43, 142, 237, 238
Proteolytic process..... 13, 22, 59, 312
Proteome Discoverer™ .....65, 66, 69, 146, 156, 224, 231, 233, 238, 243, 265
Proteomics
bottom-up ..... 4, 8, 22, 142, 237, 409, 410, 412
methods.....311
shot-gun .....146
top-down.....311, 410–412
ProteoWizard™ ..... 188, 229, 285, 287
Purification.....7, 8, 11, 61, 98, 199–207, 210, 212–217, 320, 328, 336, 370, 382, 388, 389

Pyroglutamate ..... 114, 142
Pyroglutamic acid ..... 246, 287, 350, 353

## Q

Quadrupole-time-of-flight (QTOF)
mass spectrometer ..... 9, 109, 130, 133, 154, 170, 224, 351, 352, 354–356, 361, 363, 401
Quantitation, quantification ..... 9, 11–15, 18, 20, 25, 48, 122, 124, 125, 127, 130–135, 137, 142, 143, 145, 147, 156, 162, 176, 183, 224, 230, 231, 233, 235, 237, 238, 247–267, 313, 315, 316, 360, 362, 365, 410
Quantitative peptidomics ..... 121–138, 141–157, 173

## R

Radioimmunoassay ..... 121, 320, 322, 325
Rat ..... 12, 16, 17, 21, 47, 55, 72, 319, 321, 331
Rebinning ..... 191
RefSeq™ ..... 74, 85
Regulator
autocrine ..... 41
paracrine, 41
Reporter ion ..... 142, 143, 145, 150, 154
Resolution ... 8, 9, 19, 20, 64, 80, 93, 132, 143, 191, 217, 228, 245, 249, 262, 264, 267, 287, 315, 322, 323, 327, 331, 336–340, 350, 363, 366, 391, 395, 404, 410
Retention time ..... 11, 102, 108, 112–114, 151, 189, 224, 231, 347, 356, 361, 364
Retrograde labeling ..... 372, 373, 377
Reversed-phase (RP)
HPLC (RP-HPLC) ..... 8, 16, 99, 102, 131, 162, 178, 321, 328, 330, 351, 353, 354, 356
LC (RP-LC) ..... 8, 122, 129, 130
Reverse phase, *see* Reversed-phase (RP)
Reversed-phase HPLC (RP-HPLC), *see* Reversed-phase (RP)
Ropls package ..... 313, 315
R™ (statistical computing packages) ..... 147, 150, 188, 190, 192, 315

## S

Saliva ..... 15, 18, 293–309
Sample preparation
EndoProteo-filter-aided (EndoProteoFASP) ..... 294, 298
filter-aided (FASP) ..... 294
SCX-HPLC, *see* Strong-cation exchange HPLC (SCX-HPLC)
Secretory granules ..... 53, 55, 56
Secretory vesicles ..... 200, 201, 413
Separation
chromatographic ..... 15, 19, 20, 300, 307, 365, 399
technology ..... 19
Sequence alignment ..... 74, 78, 86, 88, 92, 93
Sequence annotation ..... 72, 230, 345
Sequence overlap alignment ..... 360, 364, 366
Sequencing
Edman ..... 4, 9, 11, 21
mass spectrometry ..... 11, 14, 21, 72, 264, 384, 413
N-terminal ..... 7, 10, 11, 93
Serum ..... 18, 53, 55, 56, 98, 103, 193, 322, 381, 382
Signaling pathway ..... 242, 248
Signal peptide ..... 72, 75, 83, 88, 93, 94, 200, 349
Signal-to-noise ..... 191, 326, 327, 375
Single cell peptidomics ..... 369–378
Size exclusion
chromatography (SEC) ..... 314, 396, 398, 400, 403
column ..... 284, 286, 289, 313, 398
Snake venoms ..... 15, 349, 350, 353, 355, 357
Software
data analysis ..... 109, 110, 313, 317, 338, 340, 391
de-novo sequencing ..... 251, 256, 264, 265
mass spectrometry ..... 12, 20, 133, 361, 363, 372
package ..... 147, 154, 312, 313, 372
Solid-phase extraction ..... 52, 53, 55, 57, 109, 115, 147, 148, 155, 226, 228, 244, 284, 286, 351, 352, 360, 362, 363, 365
Sonicator ..... 98, 99, 203, 204, 207, 243, 250, 261, 284, 295
Spectral library ..... 156, 230, 410–412
Spider venom ..... 10, 12, 359, 360, 362, 365, 366
Spin column ..... 109, 115, 133, 165–167, 243, 244, 313
Standard
peptide ..... 8, 11, 100, 104, 112, 114, 124–126, 132, 327
protein ..... 8, 362, 365, 397–399
Statistical data analysis ..... 315
Statistics ..... 16, 18, 20, 21, 23, 122, 123, 136, 151, 152, 191, 194, 195, 312, 313, 315, 355
Stress
biological ..... 249, 261
Strong-cation exchange HPLC (SCX-HPLC) 337–339, 342
Substrate capture ..... 101, 104
Sulfophenyl isothiocyanate (SPITC) ..... 370, 371, 373, 375, 376, 378
Supernatant ..... 46, 47, 53–56, 99, 103, 179, 182, 204, 215, 219, 227, 244, 251, 252, 262, 285, 296, 306, 313, 323–325, 339, 352, 362, 384, 391, 392, 400
Surface enhanced laser desorption/ionization (SELDI) ..... 18, 312
Syringe pump ..... 250, 253, 258, 354

## T

Tags
isobaric ..... 126, 129, 131, 132, 135, 142, 144, 149, 157
isotopic ..... 122–124, 126, 128, 129, 134, 136, 137, 142, 171, 337
Tandem mass spectrometry (MS/MS)
analysis ..... 312, 381
fragmentation ..... 10, 16, 20, 22, 154, 353, 410, 411
Tandem mass tag (TMT)
labeling ..... 134, 146, 147, 149, 150

Tissue
dissection ..... 177, 178, 181, 182, 206, 207, 251, 257, 265, 337, 338, 372
fresh ..... 43, 44, 47, 48, 207, 212, 215, 384, 389, 392
frozen ..... 43–45, 48, 182, 215, 400, 415
sample ..... 21, 44, 182, 210, 211, 213, 214, 216, 220, 373
TMAB, *see* Trimethylammoniumbutyryl (TMAB)
TMT, *see* Tandem mass tag (TMT)
Top-down proteomics ..... 311, 410–412
Toxins ..... 15, 349, 359, 360
Transcriptome
analysis ..... 381, 384, 389
Transcriptomic data ..... 23, 73, 248
Transfection ..... 381, 388
Trap column ..... 61, 178, 182, 183, 226, 338, 339, 351, 353, 361, 363, 365, 366, 397, 401
Trimethylammoniumbutyryl (TMAB) ..... 128–130, 134–136, 142–144, 146–155
Truncated forms ..... 51
Trypsin ..... 22, 63, 124, 142, 206, 237, 261, 308, 311, 322, 324, 361, 363, 410
Tumor cells ..... 16, 209

## U

Ultrafiltration ..... 166, 314
Ultrafiltration device ..... 165, 166
Ultra-high performance liquid chromatography (UHPLC, UPLC) ..... 61, 62, 217, 226, 251, 284, 287, 288, 351, 361, 363, 383, 384, 391
Urine ..... 11, 12, 16, 18, 59, 65, 67, 317
UV absorbance ..... 115, 323, 354

## V

Vacuum centrifuge ..... 52, 53, 100, 133, 165–167, 203, 212, 216, 226, 243, 244, 250, 252–254, 284, 321, 352, 354, 383
Venom
milking ..... 355, 361, 362
peptidomics ..... 350, 355, 357
snake, 15 ..... 350, 355, 357
spider ..... 10, 12, 360, 362, 366
Visual data analysis ..... 193–195

## X

X! Tandem™ ..... 224, 229, 230, 233, 238, 352, 357

## Z

Zebrafish ..... 80, 86, 246
ZipTip™ ..... 133, 147, 250, 263, 284, 384, 392
Zymography ..... 294, 306