2ND INTERNATIONAL TOP-DOWN PROTEOMICS SYMPOSIUM
October 3-5, 2023
Northwestern University, Chicago, IL
TDP2023.ORG
WELCOME ADDRESS

Dear Colleagues,

We would like to welcome you to the Second International Symposium on Top-Down Proteomics. The first Symposium held in Paris in 2019 was a resounding success, and we hope that you will find this meeting as productive and enjoyable.

The fields of top-down proteomics and proteoform biology continue to evolve at a rapid pace with advances in separations, analyzers, and data methods moving us into exciting application areas such as single cell and single molecule analysis, native top-down of complexes, and proteoform-specific imaging and clinical diagnostics. In addition, the advent of solutions for chip-based and non-MS analysis will help usher in a new era for rapid, broad-based research in proteoform discovery and screening.

This Symposium will host world leaders presenting their latest findings in the development of technology, as well as applications across the spectrum of life science research – from basic proteoform biology to biopharmaceutical development and clinical diagnostics. Additionally, we will host a series of round table discussions to engage all the participants. These discussions will examine the art and science of proteoform discovery, next-generation proteomics and proteoform sequencing, and the clinical value of proteoforms.

There will also be a round table on the Human Proteoform Project where the community can lend its ideas and help advance this exciting endeavor. We would like to thank the Scientific Committee that has put together this wonderful program as well as the Organizing Committee that has made all of this possible. And finally, we thank all of our sponsors and exhibitors who are an essential part of this community and an important engine in the growth and success of top-down proteomics.

We hope you enjoy the meeting and thank you for your participation.

Sincerely yours,

Paul Davis, PhD
CEO, Consortium for Top-Down Proteomics

Neil Kelleher, PhD
President, Consortium for Top-Down Proteomics Board of Directors
Director, Chemistry of Life Processes Institute and Northwestern Proteomics, Northwestern University
SCIENTIFIC PROGRAM
TUESDAY, OCTOBER 3

REGISTRATION OPENS

9:00 am  Northwestern Medicine’s Prentice Women’s Hospital | 3rd Floor Harris Atrium | 250 E Superior St., Chicago IL

10:30 am  Early Career Researchers Session (Prentice Hospital, 3rd Floor, Room M)
This in-person and online session is targeted to early career researchers and requires pre-registration. The session will include discussion and Q&A with Luca Fornelli, PhD, University of Oklahoma; Kyowon Jeong, PhD, University of Tuebingen; and Neil L. Kelleher, PhD, Northwestern University, followed by lunch (included) from 11:30 am-12:30 pm.

INSTALLATION OF POSTERS & EXHIBITOR TABLES – Harris Atrium

12:00 pm

WELCOME AND OPENING – Room L

1:00 pm  Milan Mrksich, PhD, Northwestern University
Neil L. Kelleher, PhD, Northwestern University

SESSION 1: NEW FRONTIERS IN PROTEOMICS – ENTERING THE PROTEOFORM ERA

Session Chair: Luca Fornelli, PhD, University of Oklahoma

1:10 pm  New Frontiers in Proteomics: Proteoforms, Proteoform Families, and the Human Proteoform Project (1)
Lloyd M. Smith, PhD, University of Wisconsin – Madison

1:45 pm  Ultrasensitive Methods for Hypotheses-Driven Protein Discovery and Validation (2)
David R. Walt, PhD, Harvard Medical School
2:20 pm  *A Systems Biology Approach to Discovery of Clinically Actionable Protein Isoforms* (3)
*Gloria M. Sheynkman, PhD*, University of Virginia School of Medicine

2:45 pm  *A Word from the Consortium for Top-Down Proteomics* (4)
*Paul O. Danis, PhD*, Consortium for Top-Down Proteomics

2:50 pm  Coffee Break

**SESSION 2: ADVANCING MASS SPECTROMETRY TECHNOLOGIES FOR PROTEOFORM ANALYSIS**

Session Chair: *Si Wu, PhD*, University of Oklahoma

3:15 pm  *Recent Advances in Mass Spectrometry of Proteoforms* (5)
*Joseph A. Loo, PhD*, University of California, Los Angeles

3:50 pm  *Lessons from Navigating the Ion Activation Network of the Omnitrap Platform Applied to Top-Down Mass Spectrometry* (6)
*Dimitris Papanastasiou, PhD*, Fasmatech

4:15 pm  *Capillary Electrophoresis-Mass Spectrometry for Proteoforms and Protein Complexes* (7)
*Liangliang Sun, PhD*, Michigan State University

4:40 pm  *Top Double-Down and Middle-Down Characterization of Proteoforms Using TIMS–UVPD–TIMS–Q–ECD–TOF MS/MS (PROMOTED TALK)* (8)
*Francisco A. Fernandez-Lima, PhD*, Florida International University

5:00 pm  And now a word from our sponsors!
*Bruker*
*Thermo Fisher Scientific*
ROUND TABLE 1: THE ART AND SCIENCE OF PROTEOFORM DISCOVERY

5:20 pm  Moderator: Hartmut Schlüter, PhD, University Medical Center Hamburg-Eppendorf

Panelists:
Luca Fornelli, PhD, The University of Oklahoma
Rohan Thakur, PhD, Bruker
Si Wu, PhD, The University of Oklahoma
Vlad Zabrousakov, PhD, Thermo Fisher Scientific

POSTER SESSION & COCKTAILS/HORS D’OEUVRES – Harris Atrium

7:00 pm  Harris Atrium

WEDNESDAY, OCTOBER 4

SESSION 3: EMERGING PROTEOMIC PLATFORMS – Room L

Session Chair: Kyowon Jeong, PhD, University of Tuebingen

9:00 am  Towards Long-Read Single-Molecule Protein Sequencing on an Array of Unfoldase-Coupled Nanopores (9)
Jeffrey Nivala, PhD, University of Washington

9:35 am  Miniaturized Quantitative Low-Input (Nano)-Top-Down Proteomics Reveals Information About Proteoforms Not Accessible Via Bottom-Up Proteomics (PROMOTED TALK) (10)
Andreas Tholey, PhD, Kiel University

9:55 am  And now a word from our sponsors!
Nautilus
Quantum-Si

10:15 am  Coffee Break
10:45 am  *Sampling the Proteome by Mass Spectrometry Versus Emerging Single Molecule Counting Methods* (11)
*Michael MacCoss, PhD*, University of Washington

**ROUND TABLE 2: NEXT-GENERATION PROTEOMICS AND SINGLE-MOLECULE PROTEOFORM SEQUENCING**

11:15 am  Moderator: *Lloyd Smith, PhD*, University of Wisconsin - Madison

Panelists:
*Michael MacCoss, PhD*, University of Washington  
*Parag Mallick, PhD*, Nautilus Biotechnology  
*Gloria Sheynkman, PhD*, University of Virginia  
*Kenneth Skinner, PhD*, QuantumSi

**LUNCH & POSTER SESSION** – Harris Atrium

12:00 pm

**SESSION 4: MAPPING SPATIAL AND SINGLE-CELL PROTEOFORM LANDSCAPES**

Session Chair: *Julea Vlassakis, PhD*, Rice University

1:30 pm  *Single-cell omics: Precision microanalytical tools designed to profile proteoforms* (12)
*Amy E. Herr, PhD*, University of California, Berkeley

2:05 pm  *Toward Single-cell Proteoforms and Proteoform Imaging* (13)
*Ljiljana Pasa-Tolic, PhD*, Pacific Northwest National Laboratory

2:35 pm  *Digitizing Proteoform Biology with Single Molecule Mass Spectrometry* (14)
*Neil Kelleher, PhD*, Northwestern University

3:05 pm  Coffee Break
SESSION 5: DEFINING FUNCTIONS OF PROTEOFORMS AND THEIR COMPLEXES

Session Chair: Bryan Drown, PhD, Purdue University

3:35 pm  Functionalizing Proteoforms (15)
Marc Vidal, PhD, Harvard Medical School

4:10 pm  Elevated Pressure CID in Tandem-Trapped Ion Mobility Spectrometry (Tandem-TIMS-CID) for Structural Characterization of Protein Complexes (PROMOTED TALK) (16)
Fanny C Liu, PhD, Florida State University

4:30 pm  Chemical proteomic strategies to discovery proteoform-specific small-molecule probes (17)
Ben Cravatt, PhD, The Scripps Research Institute

4:55 pm  Direct Determination Of Membrane Protein Complexes From Cellular Membranes Through Native Top-Down MS (PROMOTED TALK) (18)
Wonhyeuk Jung, PhD, Yale University

ROUND TABLE 3: THE HUMAN PROTEOFORM PROJECT

5:15 pm  Moderator: Neil L. Kelleher, PhD, Northwestern University

Panelists:
Alexandra Naba, PhD, University of Illinois at Chicago
Douglas Sheeley, ScD, Office of Strategic Coordination - the Common Fund, the National Institutes of Health
Marta Vilaseca, PhD, Institute for Research in Biomedicine
David R. Walt, PhD, Harvard Medical School

COCKTAILS AND DINNER AT WOODWIND RESTAURANT

7:00 – 11:00 pm  Registration includes cocktails and dinner at Woodwind Restaurant located at 259 E. Erie St., 18th Floor, just a few minutes walk from Prentice.
THURSDAY, OCTOBER 5

SESSION 6: PROTEOFORMS AS DRIVERS OF CLINICAL RESEARCH – Room L

Session Chair: Marta Vilaseca, PhD, IRB Barcelona

9:00 am  Top-down Proteomics for Cardiac Precision Medicine and Clinical Diagnosis (19)
Ying Ge, PhD, University of Wisconsin-Madison

9:35 am  The Importance of Proteoform-Level Knowledge in Biomedical Research (20)
Julia Chamot-Rooke, PhD, Institute Pasteur

10:00 am  Top-Down Protein Analysis for Accurate Identification Of Hemoglobin Variants Using Capillary Electrophoresis-High-Resolution Mass Spectrometry(PROMOTED TALK) (21)
Ruben Luo, PhD, Stanford University

10:20 am  Coffee Break

10:45 am  Accelerating disease marker identification and drug discovery by clinical proteoformics (22)
Hartmut Schlüter, PhD, University Medical Center Hamburg- Eppendorf

CLOSING DAY LECTURE

11:05 am  Cleavage of Histone H2A During Embryonic Stem Cell Differentiation Destabilizes Nucleosomes to Counteract Gene Activation (23)
Benjamin A. Garcia, PhD, Washington University
ROUND TABLE 4: THE CLINICAL VALUE OF PROTEOFORMS

11:40 am  Moderator: Ying Ge, PhD, University of Wisconsin-Madison

Panelists:
Josh Levitsky, MD, Northwestern Medicine
Wendy Sandoval, PhD, Genentech
Chris Shuford, PhD, LabCorp
Julea Vlassakis, PhD, Rice University

FINAL REMARKS

12:25 pm  Neil L. Kelleher, PhD, Northwestern University

12:30 pm  Symposium Close
SPEAKER AND POSTER ABSTRACTS

Please use the QR code below to view speaker and poster abstracts.
ORGANIZATIONAL COMMITTEE

- Michael Caldwell, PhD, Northwestern University
- Julia Chamot-Rooke, PhD, Institut Pasteur
- Paul Danis, PhD, Consortium for Top-Down Proteomics
- Sheila M. Judge, PhD, Northwestern University
- Neil L Kelleher, PhD, Northwestern University (Chair)
- Lisa La Vallee, MS, Northwestern University (Lead Organizer)

SCIENTIFIC PROGRAMMING COMMITTEE

- Michael Caldwell, PhD, Northwestern University
- Julia Chamot-Rooke, PhD, Institut Pasteur
- Paul Danis, PhD, Consortium for Top-Down Proteomics
- Ying Ge, PhD, University of Wisconsin-Madison
- Neil L Kelleher, PhD, Northwestern University (Chair)
- Joseph A. Loo, PhD, University of California, Los Angeles
- Hartmut Schlüter, PhD, Universität Hamburg
- Gloria Sheynkman, PhD, University of Virginia
- Lloyd M. Smith, PhD, University of Wisconsin-Madison
- Ljiljana Pasa-Tolic, PhD, Pacific Northwestern National Laboratory
Symposium Venue:
Northwestern Medicine Prentice Women's Hospital
250 E Superior St, 3rd Floor, Harris Atrium & Room L, Chicago, IL

Omni Hotel
676 Michigan Ave, Chicago, IL
The scientific program will take place in L North and L South. The Early Career Session will be held in Room M. Refreshments will be served in Harris Family Atrium where poster sessions and sponsor tables are located.
GENERAL INFORMATION AND INSTRUCTIONS

CHECK-IN
When you enter Northwestern’s Women’s Prentice Hospital, please take the elevator up to the third floor and proceed to the check-in table at the entrance of Harris Atrium. Your name tag will serve as proof of payment and enable access to the symposium, refreshments, lunch, cocktail party, and dinner at Woodwind Restaurant.

EARLY CAREER RESEARCHER SESSION
This session will take place in Room M located on the third floor of Prentice near the Harris Atrium.

SYMPOSIUM
All invited promoted talks, roundtable discussions and sponsor presentations will take place in Room L adjacent to the Harris Atrium.

POSTER SESSIONS
The poster sessions will take place in the Harris Atrium and take place during the Oct. 3 cocktail party at 7:00 p.m. and the Oct. 4 lunch period at 12 p.m. Set up begins at 11 a.m. on Oct. 3. Please check the matching number on the board to display your poster in the right place. Pins are available at the welcome table to fix your poster.

MEALS
Coffee breaks, lunch, and the cocktail party will be served in Harris Atrium. Dinner on Wednesday night is at Woodwind Restaurant located on Northwestern’s downtown campus (see map). Access is limited to registered participants.

EXHIBITORS
Exhibitor tables are located in a designated area inside Harris Atrium. Set up begins at 11 a.m. on Oct 3. Please check for your company’s name sign to find your table’s location.

PRESENTERS
Speakers, please email a copy of your slide deck to Mike Caldwell, Michael.Caldwell@northwestern.edu, at least 24 hours in advance of your talk.
Cocktails and Dinner at Woodwind Restaurant
October 4, 7:00 p.m. – 11:00 p.m.

Registration includes cocktails and dinner at Woodwind restaurant located just a few blocks from Prentice Hospital. The restaurant can be accessed via inside walkways (see yellow path on the provided map), or outdoors by taking the elevators down to the main entrance on Fairbanks Court, exit right (south), Walk 2.5 short blocks to Erie St. and turn right. Enter the Northwestern Lavin Family Pavilion on Erie Street and take the elevator up to the 18th floor.
CONTACTS

General Inquiries:
Lisa La Vallee
lisa.lavallee@northwestern.edu
773.319.8011

Catering, Evening Program:
Penelope Johnson*
penelope.johnson@northwestern.edu
847.528.2631

Early Career Researcher Session:
Ryan Fellers
ryan.fellers@northwestern.edu
217.273.1331

Poster Session Inquiries:
Jody Hirsh
jkhirsh@northwestern.edu
847.703.2825

Sponsor Inquiries:
Sheila Judge*
s-judge@northwestern.edu
708.309.1942

Speaker Inquiries:
Mike Caldwell
michael.caldwell@northwestern.edu
caldwell@northwestern.edu
443.915.3704

*Can also provide answers to general inquiries.
Scale your science for protein and biopharmaceutical characterization. Increased throughput, versatility, and ease-of-use, the Thermo Scientific™ Orbitrap™ Ascend Tribrid mass spectrometer accelerates proteomics and native protein analyses.

Learn more at thermofisher.com/OrbitrapAscend

For Research Use Only. Not for use in diagnostic procedures. © 2022 Thermo Fisher Scientific Inc. All rights reserved. All trademarks are the property of Thermo Fisher Scientific and its subsidiaries unless otherwise specified. AD000645-EN 0722M
Session 2: Advancing Mass Spectrometry Technologies for Proteoform Analysis on October 3 at 3:50 pm

Lessons From Navigating the Ion Activation Network Of The Omnitrap Platform Applied To Top-Down Mass Spectrometry

Dimitris Papanastasiou Fasmatech Science & Technology – A Bruker Company

Session 2: Advancing Mass Spectrometry Technologies for Proteoform Analysis on October 3 at 5:00 pm

High Throughput Top-Down Sequencing via Trapped Ion Mobility

Michael Greig, Bruker
Unleash the Proteome

NAUTILUS™
BIOTECHNOLOGY

WWW.NAUTILUS.BIO
Control More of Your Protein Research

Introducing Platinum™ – The World’s First Next-Generation Protein Sequencer

The power of protein sequencing is now in your hands! Sequence proteins right in your lab with Platinum™, the NEW benchtop solution from Quantum-Si.

Our first-of-its-kind platform gives you the power to take more control of your protein research by conveniently delivering simple, affordable protein sequencing capabilities right to your bench, without the need for bioinformatics expertise or extensive infrastructure. Now you can get deeper insights into protein biology to complement your existing proteomics approaches...right in your lab!

- Conduct proteomics experiments in your lab at your bench
- Interrogate protein variants and modifications, and correlate with biological function
- Achieve deeper proteomics insights faster
- Perform analytics with no bioinformatics expertise required

Quantum-Si.com/TDPControl

The Protein Sequencing Company™
Transforming Labs that Transform the World

As a global leader in life science, diagnostics and analytical laboratory technologies, Agilent creates innovative instruments, software, services and solutions that provide trusted answers to our customers’ most critical questions and challenges.

www.agilent.com

MASS SPECTROMETRY & CAPILLARY ELECTROPHORESIS SOLUTIONS

Uncover previously inaccessible structural information

- Fully tunable fragmentation for top and middle down analysis
- Detect and quantify lower abundance proteoforms with increased sensitivity
- Maintain and characterize PTMs with minimal effort or artifact introduction
Optimize Protein Extraction & Processing
For Any Sample Type or Analysis Method

Enabled by Adaptive Focused Acoustics® (AFA®) Technology
Thank you to our sponsors!
2023 INTERNATIONAL TOP-DOWN PROTEOMICS SYMPOSIUM
SPEAKER & POSTER ABSTRACTS
NEW FRONTIERS IN PROTEOMICS - PROTEOFORMS, PROTEOFORM FAMILIES, AND THE HUMAN PROTEOFORM PROJECT

Lloyd M. Smith

1Department of Chemistry, University of Wisconsin - Madison

Proteins are the primary effectors of function in biology, and thus complete knowledge of their structure and behavior is needed to decipher function. However the richness of protein structure and function goes far beyond the linear amino acid sequence dictated by the genetic code. Multigene families, alternative splicing, coding polymorphisms, and post-translational modifications, work together to create a rich variety of proteoforms, whose chemical diversity is the foundation of the biological complexes and networks that control biology. "Proteoforms" are the specific molecular forms in which proteins are present in biological systems; only direct analysis of the proteoforms themselves can reveal their structures, dynamics, and localizations in biological systems.

Remarkably, the dominant paradigm of proteomics research, “bottom-up” proteomics, does not identify proteoforms – rather, proteins are enzymatically digested into peptides, whose identification then indicates the likely presence of their parent proteins in the sample. This strategy destroys the information as to what form of the protein the peptide represents, and thus the critical information needed to identify proteoforms is lost. The entire field of Biology is thus attempting to understand life in the absence of the ability to understand the molecules that define life. This limitation of today’s technology provides a “grand challenge” to the scientific community, to devise new strategies and approaches that are able to comprehensively and quantitatively reveal the full breadth of the proteome at the proteoform level2,3.

In this presentation I will provide an overview of this interesting problem, along with a variety of new tools and approaches that we and others are developing to address it. Developing the technology to decipher proteoforms, building a comprehensive atlas of proteoforms present in human systems, and eventually deciphering the functional roles they play in normal and disease biology, comprise central elements in the quest to understand human biology.

The development and broad dissemination of new technologies for comprehensive proteoform analysis is central to deciphering function in biological systems.

ULTRASENSITIVE METHODS FOR HYPOTHESIS-DRIVEN PROTEIN DISCOVERY AND VALIDATION

David R. Walt$^{1,2}$

$^1$Wyss Institute for Biologically Inspired Engineering at Harvard University, Boston, MA, USA; $^2$Department of Pathology, Brigham and Women’s Hospital, Harvard Medical School, Boston, MA, USA

Mass Spectrometry and several other existing (OLink, SOMAscan) and emerging methods (protein and peptide sequencing and some array-based approaches) have the potential to enable unbiased discovery of both proteomic and proteoform composition. In addition to these top-down approaches, additional methods are needed to detect proteins at even lower concentrations than these methods can access. These approaches will need to be hypothesis driven based on either transcriptomic data or mechanistic insights. A challenge with these ultrasensitive protein detection methodologies is they require affinity reagents that bind specifically to different proteins and proteoforms.

In addition to their use for discovery, ultrasensitive methods will be essential to validate proteins discovered through top-down methods. The tunable dynamic range of binding assay methods, coupled with their improved specificity over most top-down approaches will make such assays an essential component of the proteoform discovery and validation process.

The ability to bind and detect extremely low concentrations of proteoforms provides a unique capability to identify and quantify protein molecules that are inaccessible through most top-down approaches.

A SYSTEMS BIOLOGY APPROACH TO DISCOVERY OF CLINICALLY ACTIONABLE PROTEIN ISOFORMS

Gloria M. Sheynkman\textsuperscript{1,2,3,4}

\textsuperscript{1}Department of Molecular Physiology and Biological Physics, University of Virginia, Charlottesville, VA
\textsuperscript{2}Department of Biochemistry and Molecular Genetics, University of Virginia, Charlottesville, VA
\textsuperscript{3}Center for Public Health Genomics, University of Virginia, Charlottesville, VA
\textsuperscript{4}UVA Comprehensive Cancer Center, University of Virginia, Charlottesville, VA

Abstract

Diverse protein forms, or “proteoforms”, can arise from post-transcriptional mechanisms such as alternative splicing and/or post-translational mechanisms such as phosphorylation. Alternative splicing is a mechanism that leads hundreds of thousands of potential proteoforms that differ in their primary protein sequence, i.e., protein isoforms. The dysregulation of splicing underlies the spectrum of human pathophysiology, including cancer, cardiovascular diseases, and neurological disorders, its role underscored by systems genetics findings of thousands of splicing quantitative trait loci (sQTLs) colocalized with disease phenotypes. The capability to link sQTLs to their downstream risk and non-risk proteoforms (protein isoforms) represents a tremendous opportunity to systematically uncover and treat proteoform-driven diseases.

The goal of the Sheynkman lab is to accelerate the discovery of clinically actionable protein isoforms. We accomplish this through a team science, interdisciplinary approach that integrates cutting-edge analytical and computational approaches from systems genetics, proteogenomics, and network biology. This disease agnostic approach involves the interconnections of several methods 1) enhanced detection of protein isoforms, through integration of MS-based proteomics and long-read RNA sequencing technologies [1, 2], 2) discovering protein isoform drivers of disease, through integration of systems genetics and proteogenomics [3], and 3) dissect the mechanism by which protein isoforms drive phenotypic changes, through machine learning-based and experimental interactomic workflows.

Novel aspect
Integration of proteogenomic and systems genetics approaches to accelerate the discovery of clinically actionable protein isoforms.

A WORD FROM THE CONSORTIUM FOR TOP-DOWN PROTEOMICS

Paul O. Danis

Consortium for Top-Down Proteomics, Cambridge, MA USA

The Consortium for Top-Down Proteomics, founded in 2012, is focused on building the community of researchers and investigators, and enabling facile and extensive characterization of the various proteoforms that form the biological complexes and networks that control biology. We do this by both developing new information and methodology through our community initiatives as well as sharing best practices and latest development through our seminars and outreach programs.

The Consortium has recently undertaken several initiatives aimed at increasing the effectiveness of top-down methodologies as well as lowering the barriers to adoption. These initiatives have addressed the challenges of standardized proteoform notation\(^1\), intact protein separations\(^2\), proteoform identification metrics\(^3\), methods for antibody characterization\(^4\), capillary electrophoresis of proteoforms, standard methods for native analysis\(^5\), as well as the Human Proteoform Project\(^6\). The majority of these are completed, and their publications have been well received with over 500 citations in total. In addition, the original proteoform paper\(^7\) has been cited over 1000 times. Our outreach efforts have brought attendees from over 50 countries to our seminars and programs.

Going forward a major thrust of the Consortium is the Human Proteoform Project, where we will be engaging the community to bring ideas and energy to the development of the roadmap, connecting with other organizations to be sure the outputs are aligned with their related efforts, and seeking partners to help bring this project to fruition.

References

\(^1\) ProForma: A Standard Proteoform Notation

\(^2\) Best practices and benchmarks for intact protein analysis for top-down mass spectrometry

\(^3\) A five-level classification system for proteoform identifications

\(^4\) Interlaboratory Study for Characterizing Monoclonal Antibodies by Top-Down and Middle-Down Mass Spectrometry


RECENT ADVANCES IN MASS SPECTROMETRY OF PROTEOFORMS: ARE WE THERE YET?

Joseph A. Loo

University of California- Los Angeles, Department of Chemistry and Biology, UCLA Molecular Biology Institute, Los Angeles, CA

Mass spectrometry and top-down proteomics have advanced to provide exquisite definition of proteoforms that have and/or will be found to be important in biology and medicine. An Editorial in Nature Methods in 2008 highlighted “top-down mass spectrometry” as a “Method to Watch” for its ability “to sequence intact proteins - post-translational modifications” [1]. Yet, it pointed out deficiencies in the technologies available 15-years ago: “...larger sample quantities are required and the analysis time is longer than for a bottom-up experiment, precluding high-throughput analyses. New methods are needed for efficient protein separation, and robust computational tools for assigning protein identities and PTMs from top-down data are also lacking.” Certainly, there has been much progress in all of these areas, exemplified by recent examples of single ion/molecule measurements and even single cell top-down proteomics. This Symposium will feature many examples of cutting-edge mass spectrometry-based technologies that are ready to be deployed for capturing and profiling proteoforms. But can the field claim to have addressed the concluding statement of the Nature Methods editorial [1], “Perhaps one day this approach will be the method of choice for investigating the biological importance of combinatorial PTMs”? Where are the gaps, if any, that remain to be addressed by future technological advancements? For sure, we can do so much with today’s experimental platforms, but challenging questions in biology remain that even current methods can not address. What’s on the horizon?

Novel Aspect: Advances in mass spectrometry and top-down proteomics is the leading strategy for elucidating proteoforms

LESSONS FROM NAVIGATING THE ION ACTIVATION NETWORK OF THE OMNITRAP PLATFORM APPLIED TO TOP-DOWN MASS SPECTROMETRY

Dimitris Papanastasiou

\textsuperscript{1}Institution/Affiliation (Arial, Font size 10, italicized)

Fasmathec Science & Technology – A Bruker Company

Abstract Text (Arial, Font Size 11, 300-word limit)

The Omnitrap platform is a powerful linear ion trap supporting an extensive functionality landscape for processing ions in the gas phase [1]. Performance enhancements are enabled by incorporating variations of the entire range of ion activation methods into a single platform in a highly dynamic fashion. The ion activation arsenal comprises reagent ions, radical neutral species, photons, electrons and collisions, while the ability to apply these methods in parallel or in tandem offers unique opportunities for comprehensive characterization of intact proteins. The ion accumulation functionality increases the in-scan dynamic range and is a key feature for efficient multidimensional multiple-stage experiments, offering new-levels in protein characterization by top-down mass spectrometry.

Optimized MSn workflows involving collisional activation in Q2 and ECD reactions in Q5 of the Omnitrap platform are presented for the analysis of intact non-reduced mAbs and Fab units sprayed under both denatured and native conditions. Enhanced sequence coverage across the variable domain in mAbs is demonstrated by MS3 and MS4 workflows. Dissociation of inter-chain disulfide bonds is reported in MS2 CID and of intra-chain bonds in MS4 collisionally activated ECD. Top-down experiments of intact mAbs using fast hydrogen ion beams are also discussed and a mechanism for fragment ion formation is proposed. Hydrogen atom rearrangements are mapped for all primary fragment types, while the internal fragment enigma and the corresponding mass accuracy requirements for accurate spectral annotation are highlighted. A new data processing software with a sophisticated scoring system for annotating highly congested mass spectra is presented. Additional workflows combining MS2 CID and MS3 ECD developed for the structural characterization of glycans are also described [2]. A powerful experimental setup combining IR activation with ExD and UVPD is reported, and preliminary top-down data are presented [3].


CAPILLARY ELECTROPHORESIS-MASS SPECTROMETRY FOR PROTEOFORMS
AND PROTEIN COMPLEXES

Liangliang Sun

Department of Chemistry, Michigan State University, 578 S Shaw Lane, East Lansing, MI 48824

Capillary electrophoresis-mass spectrometry (CE-MS) has been recognized as a promising analytical tool for top-down characterization of proteoforms and protein complexes since 1980s. During the last decade, CE-MS has attracted more and more attention for global denaturing and native top-down proteomics (TDP), aiming to achieve complete pictures of proteoforms and protein complexes in complex biological systems. We recently showed several cases of applying advanced CE-MS techniques to the delineation of proteoforms and protein complexes. First, we performed the first TDP study of a pair of isogenic human nonmetastatic and metastatic colorectal cancer (CRC) cell lines (SW480 and SW620) using CE-MS/MS. [Sci Adv, 2022] We identified 23,622 proteoforms of over 2000 genes from the two cell lines, representing nearly fivefold improvement in the number of proteoform identifications compared to previous TDP datasets of human cancer cells. We revealed substantial transformation of CRC cells in proteoforms after metastasis.
Second, we developed a CE-ion mobility spectrometry (IMS)-MS/MS technique for online multi-dimensional separation of proteoforms for the first time and showed that the technique could substantially improve the identification of large proteoforms (>30 kDa) in complex samples. [Anal Chem, 2023] Third, we developed a native capillary isoelectric focusing (nIEF)-MS technique for high-resolution separation and accurate delineation of protein complexes (i.e., an interchain cysteine-linked antibody-drug conjugate). [Anal Chem, 2022] The nIEF-MS technique enabled precise measurements of isoelectric points (pIs) of protein complexes, allowing us to study how protein sequence variations/PTMs regulate the pIs of protein complexes.

This work demonstrates the novel technical development and biological applications of CE-MS/MS for proteoforms and protein complexes.

TOP DOUBLE-DOWN AND MIDDLE-DOWN CHARACTERIZATION OF PROTEOFORMS USING TIMS-UVPD-TIMS-Q-ECD-TOF MS/MS

Francisco A. Fernandez-Lima

1Department of Chemistry and Biochemistry, Florida International University

Post-translational modifications (PTMs) play a major role in regulating protein dynamics and influence major biological processes. The nature and position of each PTM is crucial to decipher how this information is translated into a biological response. In the present work, the potential of a novel tandem top-“double-down” approach –ultraviolet photodissociation followed by mobility and mass selected electron capture dissociation and mass spectrometry (UVPD-TIMS-q-ECD-ToF MS/MS)– is illustrated for the characterization of positional proteoforms. During double down, a 213 nm UV laser generates UVPD fragments prior to the ion mobility-mass precursor separation and ECD TOF MS. UVPD fragment of interest from different proteoforms (e.g., H4 Ac+Me2, 2Ac+Me2 and 3Ac+Me2) can be separated in the mobility and m/z domain prior ECD fragmentation and TOF MS analysis leading to a clear assignment of the PTM localization for each of the H4 proteoforms. In addition, the workflow is illustrated for the analysis of α-synuclein positional glycoforms (e.g., T72, T75, T81 and S87 modified with a single O-GlcNAc). Mobility profiles of the intact α-synuclein glycoforms exhibited large structural heterogeneity across the 8+-15+ charge state distribution; however, all the four α-synuclein glycoforms were only partially separated at the protein level in the mobility domain. Alternatively, a middle-down approach based on chymotrypsin digestion followed by tandem TIMS-q-ECD-MS/MS permitted the separation of the glycoforms using the Val40-Phe94 (55 residues) proteolytic product that contains all four glycoforms of interest. The ECD fragmentation of the ion mobility and m/z separated products allowed for the PTM assignment with a sequence coverage of ~80%.
TOWARDS LONG-READ SINGLE-MOLECULE PROTEIN SEQUENCING ON AN ARRAY OF UNFOLDASE-COUPLED NANOPORES

Keisuke Motone¹, Daphne Kontogiorgos-Heintz¹, Jeff Nivala¹,²

¹Paul. G. Allen School of Computer Science and Engineering, University of Washington, Seattle, WA;
²Molecular Engineering and Science Institute, University of Washington, Seattle, WA.

The ability to sequence single protein molecules in their native, full-length form would enable a more comprehensive understanding of proteomic diversity. Current technologies, however, are limited in achieving this goal. In response to this challenge, we are developing a method for long-range, single-molecule reading of intact protein strands on a commercial nanopore sensor array (Oxford Nanopore Technologies’ MinION device) [1]. By using the ClpX unfoldase to ratchet proteins through a protein nanopore sensor [2], we have demonstrated single-amino acid level sensitivity, the capability to resolve combinations of amino acid substitutions across long protein strands, and the capacity to process intact, folded protein domains for complete end-to-end analysis. These results provide proof-of-concept for a platform that has the potential to characterize full-length proteoforms at single-molecule resolution.


MINIATURIZED QUANTITATIVE LOW-INPUT (NANO)-TOP-DOWN PROTEOMICS REVEALS INFORMATION ABOUT PROTEOFORM NOT ACCESSIBLE VIA BOTTOM-UP PROTEOMICS

Jan Leipert, Philipp T Kaulich, Max K Steinbach, Konrad Winkels, Andreas Tholey

Proteomics & Bioanalytics – Institute for Experimental Medicine, Christian-Albrechts-Universität zu Kiel, Kiel, Germany

Abstract

While most nanoproteomics approaches for the analysis of low-input samples are based on bottom-up proteomics workflows, top-down approaches enabling proteoform characterization are still underrepresented.

Using mammalian cell proteomes, we established a facile one-pot sample preparation protocol based on protein aggregation on magnetic beads and intact proteoform elution. The method was then adapted for top-down proteomics sample preparation on a digital microfluidics (DMF) device, which is a tool enabling the manipulation (i.e. splitting, merging, mixing) of small droplets of liquids with volumes between ca. 0.5 to 2 µL on a chip by application of electric fields.

Single intact Caenorhabditis elegans nematodes, which consist of an invariant number of 959 cells, were lysed (with the additional challenge to crack the cuticule surrounding the organism), and the intact proteoforms were isolated and desalted on-chip, providing LC-MS ready proteome samples [1]. In combination with a recently developed multi-compensation voltage (cv) LC-FAIMS MS setup [2], the number of proteoform identifications compared to in-tube sample preparation was increased by 46% [3]. Label-free quantification of single nematodes grown under different conditions allowed us to identify changes in abundance of proteoforms not distinguishable by bottom-up proteomics [3, 4]. The presented workflow will facilitate proteoform-directed analysis on samples of limited availability

Novel Aspect

Development of a novel sample preparation platform for multiplexed low-input top-down proteomics.

References

SAMPLING THE PROTEOME BY MASS SPECTROMETRY VERSUS EMERGING SINGLE MOLECULE COUNTING METHODS

Michael J. MacCoss¹, Javier Antonio Alfaro²,³,⁴, Danielle A. Faivre¹, Christine C. Wu¹,
Meni Wanunu⁵ & Nikolai Slavov⁶,⁷

1) Department of Genome Sciences, University of Washington, Seattle, WA, USA.
2) International Centre for Cancer Vaccine Science, University of Gdańsk, Gdańsk, Poland.
3) Department of Biochemistry and Microbiology, University of Victoria, Victoria, British Columbia, Canada.
4) School of Informatics, University of Edinburgh, Edinburgh, UK.
5) Department of Physics, Northeastern University, Boston, MA, USA.
6) Departments of Bioengineering, Biology, Chemistry and Chemical Biology, Single Cell Proteomics Center and Barnett Institute, Northeastern University, Boston, MA, USA.
7) Parallel Squared Technology Institute, Watertown, MA, USA

Mammalian cells have about 30,000 times as many protein molecules as mRNA molecules, which has major complications in the adoption of methods successfully applied for the transcriptome towards the proteome. Liquid chromatography-tandem mass spectrometry (LC-MS/MS) has been the predominant strategy for the multiplex quantification of proteins in complex mixtures. More recently, novel methods for protein analysis are emerging. These new methods adapt flow-cell and nanopore methods analogous to those developed for nucleic acid analysis, for the identification and potentially even sequencing individual polypeptide molecules. The pros and cons of 1) counting molecules in solution versus as gas phase ions, 2) compressing molecules of the same type prior to measurement (e.g. as performed by LC-MS) versus counting molecules one at a time, 3) cost and transferability, and 4) sensitivity of the different strategies will be discussed. Strategies that are routine for counting billions of protein molecules by LC-MS/MS that can benefit emerging single-molecule methods, will be presented.

Novel Aspect: Discussion of the pros and cons of sampling the proteome by mass spectrometry versus emerging single molecule counting methods.

Single-cell omics: Precision microanalytical tools designed to profile proteoforms

Amy E. Herr¹,²

¹Department of Bioengineering, University of California, Berkeley, Berkeley, CA; ²Chan Zuckerberg Biohub San Francisco, San Francisco, CA

Both fundamental biosciences and applied biomedicine have been transformed by powerful single-cell resolution analysis tools, not least of which are sequencing tools. Yet, genomic and transcriptomic information is steps removed from cellular phenotype and function. To a large degree, proteins are phenotype. Proteins dictate function. Consequently, protein molecules are diverse. The natural diversity of proteoforms presents a tantalizing window into cellular phenotype and function, while also presenting measurement challenges.

To address this proteoform-measurement gap, our bioengineering tools-focused research addresses the design and introduction of single-cell analysis tools with the specificity to distinguish among proteoforms. Here, I will highlight our recent work that harnesses microfluidic design to resolve proteoforms from single cells using precision, single-cell electrophoresis modalities. Further, I will share our work to integrate proteoform information into single-cell omic analysis, with both analyses performed, importantly, on the same originating cell.

Our long-term vision is to create tools that allow researchers to query a unique originating cell for protein-level information, as informed by ex ante sequencing-based discovery.

Taken together, we strive to introduce tools uniquely equipped to measure both cellular and molecular heterogeneity as a means to more comprehensively understand cellular form and function along the many axes of cellular information flow.
TOWARD SINGLE-CELL PROTEOFORMS AND PROTEOFORM IMAGING

Ljiljana Paša-Tolić¹, Kevin J. Zemaitis⁴, Dušan Veličković¹, Mowei Zhou², James M. Fulcher⁴, Yen-Chen Liao⁴, Ying Zhu³, Sarah M. Williams⁴, David J. Degnan¹, Lisa M Bramer⁴, William R. Kew¹

¹Pacific Northwest National Laboratory, Richland, WA, United States
²Department of Chemistry, Zhejiang University, Hangzhou, Zhejiang, China
³Genentech, South San Francisco, CA, United States

The ability to measure proteins in a few cells or a single cell remains a major analytical objective and challenge. Current approaches for probing spatial distribution of the proteome typically rely on the use of antibodies, which limits multiplexing and requires a priori knowledge of protein targets. More recently, conventional bottom-up proteomics in nanoPOTS format has been demonstrated for proteome-wide analysis of small tissue sections; however, this approach cannot provide proteoform information and is therefore lacking as proteoforms drive cellular functions. To address this challenge, we integrated nanoPOTS top-down proteomics (TDP) and MALDI mass spectrometry imaging (MSI) and deployed this workflow for characterization of human, murine, plant, and microbial systems.

Fresh frozen tissue (or biofilm) sections were analyzed using MALDI (Spectroglyph LLC) coupled with a Q-Exactive HF Orbitrap MS upgraded with ultra-high mass range (UHMR) boards.[¹] LCM nanoPOTS TDP was accomplished in parallel with MALDI imaging. Laser capture microdissection (LCM) was used to cut tissue (or biofilm) microstructures from serial sections for TDP analyses. Confident TDP proteoform identifications served as a lookup table for MSI peak annotations using custom software (i.e., IsoMatchMS).[²] Furthermore, we have implemented UVPD to enable TDP directly from tissue for (targeted) proteoform assignment validation.[³] MALDI MSI was accomplished in tandem with autofluorescence imaging and/or histological staining allowing for segmentation and classification to put molecular findings into a histomorphological context. Our ultimate goal is the creation of a 3D proteoform images and their integration with other maps (e.g., metabolites, lipids, transcripts) to provide a deeper understanding of biology.


Digitizing Proteoform Biology with Single Molecule Mass Spectrometry
Neil Kelleher, PhD, Northwestern University
FUNCTIONALIZING PROTEOFORMS

Marc Vidal$^{1,2}$

$^1$Department of Genetics, Blavatnik Institute, Harvard Medical School, Boston, MA; $^2$Center for Cancer Systems Biology (CCSB), Dana-Farber Cancer Institute, Northwestern University, Boston, MA

The functional importance of proteoform expression remains to be established at proteome-scale. As mentioned in one of our publications addressing this question [1], “compiling a comprehensive catalog of different proteoforms and subsequently studying their distinct functions will be necessary for a full understanding of normal cellular biology, as well as disease pathogenesis at the systems level.”

In this talk, I will describe our most recent results in our quest of “functionalizing proteoforms” at proteome-scale.

ELEVATED PRESSURE CID IN TANDEM-TRAPPED ION MOBILITY SPECTROMETRY (TANDEM-TIMS-CID) FOR STRUCTURAL CHARACTERIZATION OF PROTEIN COMPLEXES

Fanny C. Liu, Tyler C. Cropley, Valentina Rangel-Angarita, Christian Bleiholder

Department of Chemistry and Biochemistry, Florida State University, Tallahassee, FL

Protein-protein interaction networks are at the core of complex diseases such as cancer, diabetes, or hypertension [1]. Mapping of protein interaction network requires identification of protein subunits and their interactions in multiprotein complexes. Characterization of intact protein complexes can be achieved by top-down proteomics, which utilizes intact protein mass and fragmentation patterns generated by MS/MS approach.

Here, we employed the top-down approach via collisional-induced dissociation (CID) at elevated pressure in our recently developed tandem-trapped ion mobility spectrometer/mass spectrometer (tandem-TIMS/MS) [2-5] to determine interactions between protein chains of protein complexes. We then investigated the three-dimensional structures of the protein precursor and generated subunits by our computational method, the structure relaxation approximation (SRA). SRA predicts ion mobility spectra for proteins and protein complexes with specified charge states, thereby providing detailed structural models for the analyte ions.

Tandem-TIMS-CID was performed at an elevated pressure of ~1-2 mbar on the following homotetrameric protein complexes: streptavidin (53 kDa), neutavidin (60 kDa), avidin (64 kDa), and concanavalin A (110 kDa). Our results showed that the overall native topology of the protein complex precursors was retained during the ion mobility / mass spectrometry measurements in the absence of solvent. We further observed compact monomers, dimers, and trimers at distinct activation voltages in each respective spectrum. The cross sections measured were consistent with those reported for surface-induced dissociation (SID) and estimated for the native structures obtained from x-ray crystallography. Further, the activation voltages required to produce the maximum abundance of compact trimers and dimers reflected the relative strengths of the subunit interactions in the crystal structures. Our SRA calculations suggested that while the monomer subunits were found to exhibit essentially no memory of the native precursor structure, many aspects of the native structure can be retained in trimers or other higher-order subunits.

Novel Aspect:
Combination of tandem-TIMS-CID and SRA methods offer a powerful approach to rigorously investigate subunits of protein complexes.

Chemical proteomic strategies to discovery proteoform-specific small-molecule probes
Ben Cravatt, PhD, The Scripps Research Institute
DIRECT DETERMINATION OF MEMBRANE PROTEIN COMPLEXES FROM CELLULAR MEMBRANES THROUGH NATIVE TOP-DOWN MS

Wonhyeok Jung¹,²#; Aniruddha Panda¹,²#; Caroline Brown¹,²; Jeff Coleman¹,²;
James E. Rothman¹,²; Kallol Gupta¹,²

¹Nanobiology Institute, Yale University, West Campus ISTC, 850 West Campus Drive, West Haven, CT - 06516, U.S.A
²Department of Cell Biology, Yale School of Medicine, 333 Cedar Street, PO Box 208002, New Haven, CT – 06520, U.S.A

#Authors contributed equally

Cellular membrane plays an essential role in regulating hierarchical organizations of membrane proteins (MP) and lipids that drive downstream signaling cascades. While nativeMS has been at the forefront of detecting these complexes, studying MPs with nativeMS demands prior dissolution of the cellular membrane through either chemical or mechanical means. We recently developed an in vitro platform that enables nativeMS of MPs directly from tunable lipid-bilayers¹,². Here we expand this platform directly to native membranes. Taking different physiological membranes of prokaryotic and eukaryotic origin, we demonstrate our ability to detect MP-complexes directly from intact lipid membranes, without requiring any prior dissolution. We further coupled this to native top-down fragmentation to ID the MP-complexes. To achieve this, we leveraged our existing technological platform to cell-derived native vesicles. The integrity of the vesicles was rigorously checked using negative-stain EM imaging, DLS, and functional analysis. By subjecting these cell-derived vesicles to nativeMS analysis, we first demonstrate our ability to directly ablate out and detect intact multimeric MP complexes. Particularly, we show that these MP complexes can be ablated using pre-quadrupole front-end activation alone. This enables us to directly isolate a desired target MP complex and subject it to downstream MS/MS activation. Subsequently, using both complex-down and top-down fragmentation of the isolated proteins, we demonstrate our ability to determine the ID of the constituent proteins, as well as their oligomeric stoichiometry directly from native cell membranes. For top-down analysis, we make use of both collision cell HCD and ECD fragmentation, achieved by swapping the transfer multipole with an ExD-cell. We first demonstrate the broad applicability of our platform by identifying multiple bacterial multi-protein membrane complexes, as well as bound endogenous ligands. We then apply it to set of key neuronal MPs to discover a novel synaptic-vesicle multiprotein complex that regulate the timescale of neurotransmitter release.

A native MS platform that enables detection and subsequent complex-down and top-down ID of MP complexes directly from cellular membranes.

References


TOP-DOWN PROTEOMICS FOR CARDIAC PRECISION MEDICINE AND CLINICAL DIAGNOSIS

Ying Ge, PhD

Department of Cell and Regenerative Biology, Department of Chemistry, and Human Proteomics Program, University of Wisconsin-Madison

“Omics” technologies offer transformative insights to elucidate disease mechanisms and are the enabling force for precision medicine. In the post-genomic era, proteomics is the next frontier allowing an in-depth understanding of the function of cellular systems in diseases. Unlike the genome, the proteome is dynamic and highly complex due to alternative splicing and post-translational modifications (PTMs). Top-down mass spectrometry (MS)-based proteomics is the most powerful technology to comprehensively characterize proteoforms that arise from genetic variations, alternative splicing, and PTMs. We have made major advances in top-down proteomics for analysis of intact proteins directly purified from heart tissue, blood, and human pluripotent stem cell-derived cardiomyocytes (hPSC). Recently, we have developed ultra-high-sensitivity top-down proteomics method for analysis of single cells that captures single muscle cell heterogeneity in large proteoforms (>200 kDa). Importantly, we have linked altered cardiac proteoforms to contractile dysfunction in heart diseases using animal models and human clinical samples. Furthermore, we are harnessing the power of innovative top-down proteomics technologies with patient-specific hPSC-derived cardiomyocytes (CMs) in engineered cardiac tissue to understand proteoform biology in cardiac diseases for precision medicine.

Notably, we have employed cutting-edge top-down proteomics to analyze surgical human heart tissue samples from hypertrophic cardiomyopathy (HCM) patients and provided the direct evidence that proteoforms can better reflect patient’s disease phenotypes than their genotypes. This opens the door for the development of therapeutic interventions that target the HCM proteoform phenotype rather than individual genotypes. Moreover, we have developed a novel nanoproteomics method for the enrichment and comprehensive analysis of low abundance proteoforms directly from serum. This nanoproteomics method enabled the capture and analysis of cardiac troponin I (cTnI), a gold-standard cardiac biomarker from serum, providing high-resolution proteoform-resolved molecular fingerprints of diverse cTnI proteoforms. We are now in the process of developing a comprehensive proteoform-based cTnI assay to be translated into the clinic to for precise diagnosis of various types of heart diseases.

Novel Aspect: The development and application of novel top-down proteomics methods to decipher proteoform biology for cardiac precision medicine and clinical diagnostics.

References

THE IMPORTANCE OF PROTEOFORM-LEVEL KNOWLEDGE IN BIOMEDICAL RESEARCH

Julia Chamot-Rooke

*Institut Pasteur, Paris, France*

In a recent paper, the Consortium for Top-Down Proteomics highlighted five important areas illustrating the critical role of proteoforms in disease and health: neurodegeneration, cardiovascular health, infectious disease, cancer and immunology. In this talk, the importance of proteoform-level knowledge in biomedical research will be highlighted through two main examples.

The first one lies in the field of clinical microbiology. The current technique used for microbial identification in hospitals is MALDI-TOF MS. However, it suffers from important limitations, in particular for the identification of closely-related species. We therefore set up a top-down proteomics platform to characterize bacteria at the proteoform level. After optimizing this platform, from sample preparation to data analysis, using *E. coli* as a model, we show that it can be used for the discrimination of enterobacterial pathogens undistinguishable by MALDI-TOF, qualifying top-down proteomics as a promising tool in clinical microbiology.

The second example concerns multiple myeloma. In some forms of the disease, monoclonal immunoglobulin light chains (LCs) are abundantly produced, leading in some cases, to the formation of deposits in organs such as kidney or heart. The factors driving the solubility of LCs are poorly understood, but it can be hypothesized that their sequence is important. We established a novel de novo sequencing workflow based on the combination of bottom-up and top-down proteomics without database search. This workflow was then used for the complete de novo sequencing of LC proteoforms extracted from urine of 10 patients. We demonstrate that top-down proteomics is required to achieve complete sequence coverage and allows the characterization of an unexpected range of modifications. Recent results obtained on a new instrumental platform allowing improved top-down proteomics (Q-Exactive HF modified with an Omnitrap and equipped with a FTMS Booster) will also be presented.

Novel Aspect
Proteoforms as potential biomarkers for clinical applications

Top-Down Protein Analysis for Accurate Identification of Hemoglobin Variants Using Capillary Electrophoresis-High-Resolution Mass Spectrometry

Ruben Y. Luo (Presenter)
Stanford University

Introduction

Structural characterization of hemoglobin (Hb) variants, particularly the mutant forms of α- and β-subunits, is of significant value in the clinical diagnosis of hemoglobinopathy. The conventional methods for identification of Hb variants in clinical laboratories can be inadequate due to the lack of detailed structural characterization when it goes to the analysis of those Hb variants with similar sizes and charge states.

High-resolution mass spectrometry (HR-MS) has been a central technology for structural characterization of proteins. As a novel approach for HR-MS protein analysis, top-down workflow analyzes proteins in intact state without prior enzymatic digestion, and it is capable of identifying unique proteoforms. Capillary electrophoresis (CE), as a powerful separation technology for proteins, has demonstrated excellent separation efficiency in the analysis of intact Hb forms and Hb subunits. CE can be coupled with HR-MS to form a CE-HR-MS system. By this means the CE separation is able to enhance the analytical power of HR-MS to allow for superior analytical performance in Hb analysis. In addition, when neutral coating of capillary is employed to suppress electroosmotic flow, CE separates analytes only by their individual electrophoretic mobilities, which enhances separation efficiency by maximizing the electrophoretic difference between the analytes. Thus, we established a neutral-coating CE-HR-MS method for accurate identification of Hb variants with top-down protein analysis strategy.

Method

An Orbitrap Q-Exactive Plus mass spectrometer was coupled with an ECE-001 CE unit through an EMASS-II ion source. A PS1 neutral-coating capillary was used for CE separation. The electrolyte for CE was 20% acetonitrile in water with 2% formic acid. CE voltage was set at 30 kV, and ESI voltage was controlled at 2.2 kV. Samples of red blood cells were lysed in water and diluted in 10 mM ammonium formate buffer for analysis.

Results

In the neutral-coating CE, since denaturing conditions were used, intact Hb forms were dissociated in background electrolyte, and baseline separation of individual Hb subunits was observed. Hb subunits in ion electropherograms followed the order of α-, β-, δ-, γ(1)-, γ(2)-subunit.

The identification of Hb variants using CE-HR-MS is a two-step process: (1) intact-protein analysis that preliminarily identifies Hb subunits by precursor ions, and (2) top-down analysis that characterizes the primary structures and confirms the identification of Hb subunits by fragments produced from particular precursor ions. In intact-protein analysis, multiple charge states of each Hb subunit were observed in the mass spectra. At each charge state of a Hb subunit, a cluster of isotopic MS peaks were observed, corresponding to the isotopes of the molecule. All the MS peaks in a mass spectrum can be deconvoluted and those resulted from one analyte can be merged to display a single MS peak at its accurate monoisotopic mass, which can be matched to the theoretical mass of a known Hb subunit or variant. In top-down protein analysis, fragments from a precursor ion were acquired after HCD fragmentation. The MS
peaks can be deconvoluted and those resulted from one fragment can be merged to display a single MS peak at its monoisotopic mass. The monoisotopic masses of fragments can be used to characterize the primary structure of the analyte by matching them to the theoretical masses of possible fragments from the structure of a known Hb subunit.

The CE-HR-MS method was applied to the analysis of normal Hb forms as well as Hb variants from adults and neonates. The structures of Hb subunit variants β-S, β-C, β-E, β-Riyadh, β-Koln, β-New York, β-G-Accra, β-G-Siriraj, β-Khartoum, β-D-Punjab, β-D-Iran, β-Deer Lodge, α-Tarrant, γ1-Kuala Lumpur, and P-Nilotic β-δ hybrid have been successfully identified and characterized with >30% amino acid residue coverage and >95% matched fragment coverage. More importantly, post-translational modifications of certain Hb variants have been identified by the CE-HR-MS method, such as N-terminal acetylation of β-Raleigh, which could not be solved by the conventional test methods or next-generation gene sequencing.

Conclusion

We have utilized the neutral-coating CE-HR-MS method for accurate identification of hemoglobin variants. With the use of CE, only a simple dilute-and-shoot sample preparation procedure is required. Baseline separation of Hb subunits can be achieved to enhance HR-MS data quality. With these advantages, the CE-HR-MS method is compatible with clinical laboratories. In addition, CE-HR-MS can be used for effective characterization of a wide range of protein markers of clinical diagnostic value, such as monoclonal immunoglobulins in monoclonal gammopathies.
Clinical proteomics on the level of proteoforms (proteoformics) is using differential relative quantitation comparing numbers of individual proteoforms of patients with those of healthy controls. Bottom-up- and antibody-based- clinical proteomics are only quantifying the sum of proteoforms of a protein and ignoring posttranslational modifications thereby excluding many species which may be associated with diseases. Clinical proteoformics is including the quantitation of intact proteoforms and their precise identification. Both analytical aspects are more difficult than in bottom-up proteomics, when using liquid chromatography coupled to tandem mass spectrometry (LC-MSMS) including an electrospray ionization (ESI) source. ESI is producing multiple charge states thereby requiring a deconvolution of the data of the MS1 spectrum for obtaining signals usable for getting extracted ion chromatograms being basis of quantitation [1]. A second challenge is associated with the problem that comparable proteoforms of different individuals may differ in their molecular weights because of mutations in the sequence or differences in the posttranslational modifications fixed to proteoforms sharing the same amino acid sequence or both cases. This problem in addition is complicating quantitation. However, there are already success stories describing the identification of disease markers using top-down mass spectrometry based on LC-MSMS published e.g., by Yin Ge and her group [2]. A further success story highlighting benefits of screening proteoform markers is MALDI-MS identification of microorganisms based on signal patterns of proteoforms [3]. MALDI-MS is not associated with above-mentioned problems but is restricted to detection of a very small number of proteoforms out of a very many. In the future clinical proteoformics will require new technologies overcoming the above-mentioned problems and giving the opportunity to quantify and identify ideally all proteoforms present in a sample, including proteoforms differing in their conformation only. Applying tools for clinical proteoformics will yield a deeper understanding of molecular mechanisms of diseases.

Novel Aspect: Clinical proteoformics in contrast to clinical proteomics is screening the complete universe of proteins increasing successful identification of new markers.

References

CLEAVAGE OF HISTONE H2A DURING EMBRYONIC STEM CELL DIFFERENTIATION DESTABILIZES NUCLEOSOMES TO COUNTERACT GENE ACTIVATION

Mariel Coradin¹ and Benjamin A. Garcia²

¹University of Pennsylvania School of Medicine, Philadelphia PA 19104
²Washington University School of Medicine, St. Louis, MO 63110

Histone proteolysis is a poorly understood phenomenon in which the N-terminal tails of histones are irreversibly cleaved by intracellular proteases. During development, histone post-translational modifications are known to orchestrate gene expression patterns that ultimately drive cell fate decisions. Therefore, deciphering the mechanisms of histone proteolysis is necessary to enhance the understanding of cellular differentiation. Here we show that H2A is cleaved by the lysosomal protease Cathepsin L during ESCs differentiation. Using top down mass spectrometry, we identified L23 to be the primary cleavage site that gives rise to the clipped form of H2A (cH2A), which reaches a maximum level of ~1% of total H2A after four days of differentiation. Using ChIP-seq, we found that preventing proteolysis leads to an increase in acetylated H2A at promoter regions in differentiated ES cells. We also identified novel readers of different acetylated forms of H2A in pluripotent ES cells, such as members of the PBAF remodeling complex. Finally, we showed that H2A proteolysis abolishes this recognition. Altogether, our data suggests that proteolysis serves as an efficient mechanism to silence pluripotency genes and destabilize the nucleosome core particle.
CONNECTING PROTEOFORMS TO STRUCTURE WITH CLINIC

Melanie Cheung See Kit, Ankita B. Gurav, Ashlyn N. Dollar, Erika Renbarger, Ian K Webb
Department of Chemistry and Chemical Biology, Indiana University – Purdue University Indianapolis; Center for Computational Biology and Bioinformatics, Indiana University School of Medicine

Intrinsically disordered proteins (IDPs) and intrinsically disordered regions (IDRs) have many critically important cellular functions. In fact, a single IDP/IDR is often multifunctional due to its structural flexibility. NMR, SAXS, and single molecule (sm) FRET are useful in the structural characterization of isolated IDPs/IDRs. While extremely useful structural tools, SAXS and NMR report on bulk averaged properties. Though sm-FRET can probe multiple coexisting conformers, the current state of the art in solution biophysical methods cannot detect the heterogeneous mixtures of conformers, proteoforms, and protein/protein interactions (PPIs) found in the cellular environment.

Native ion mobility mass spectrometry (n-IM/MS) can characterize the entire conformational space of heterogeneous IDP/IDR populations, measuring structural changes caused by mixtures of post translational modifications (PTMs) and PPIs. The ability of n-IM/MS to characterize three-dimensional structure is limited to providing overall size/shape information. As a result, many different structures can be matched to a single mobility and mass measurement. Furthermore, unlike NMR or sm-FRET, n-IM/MS does not provide distance restraints on the angstrom to nanometer scale, which are critical for accurate structural characterization of IDPs/IDRs. Due to these shortcomings, a method for providing distance restraints and localized structural information (e.g., solvent accessibility, binding site location) that is fully integrated within the IM/MS method is needed.

We are solving this problem by developing Chemical crossLinking Identified under Native or Intact Conditions (CLINIC). In this method, proteins are crosslinked (XL), without denaturation or proteolytic digestion before proteoforms and XL sites are identified via tandem MS of the intact XL proteins via top-down proteomics. This presentation will demonstrate the application of CLINIC to determining the effects of phosphorylation on the structural ensembles of the ID phosphoproteins αsynuclein, 4E-BP2 (Eukaryotic Translation Initiation Factor 4E Binding Protein 2), and ETS1 (erythroblast transformation specific-1).

This work shows the development of methods for proteoform-specific structural ensemble characterization enabled by crosslinking identified through top-down proteomics.
SYSTEMATIC OPTIMIZATION OF ELECTRON-ACTIVATED DISSOCIATION FOR TOP-DOWN TARGETED PROTEIN SEQUENCING

Jason Causon, David Colquhoun, and Ihor Batruch

SCIEX

Recent developments in mass spectrometry hardware and software have significantly improved top-down and middle-down analysis of proteins and protein sub-units in the case of monoclonal antibodies (mAbs). Acquisition methodology, chromatography separation and data processing are all factors in maximizing performance and sequence coverage.

In this work, we detail the systematic optimization of electron activated dissociation (EAD) with MRMHR for top-down / middle-down protein sequence confirmation, and we will investigate the dependencies between acquisition and processing strategies.

This work uses a novel fragmentation technique for top-down targeted protein sequencing.

4 SCIEX technical note, MKT-26997-A.
5 SCIEX technical note, MKT-27223-A.
6 SCIEX technical note, MKT-27427-A.
QUANTUM-SI PLATINUM™ PROTEIN SEQUENCING REVEALS PEPTIDES MISSED IN DEFAULT MASS SPECTROMETRY DATABASE SEARCHES

Kenneth Skinner, Katie Croce, Haidong Huang, Brian Reed

1Quantum-Si, 29 Business Park Drive, Branford, CT 06405

The identity and position of amino acid residues dictate protein folding and function. Thus, protein sequencing is critical to uncover novel biological insights. While Edman degradation and liquid chromatography-tandem mass spectrometry (LC-MS/MS) are the primary methods for determining the sequence of proteins, modified peptides may escape detection with these techniques. For example, N-terminal modifications often impede Edman chemistry and bottom-up strategies often produce peptides less than six amino acids in length that may not generate informative fragmentation ladders. Default LC-MS/MS parameters may restrict coverage of peptides absent from databases, confounding non-targeted identification of peptides and proteoforms. To overcome these challenges, we used Quantum-Si’s Platinum single-molecule protein sequencing platform. In this workflow, peptides are conjugated to macromolecular linkers and immobilized via C-termini in nanoscale reaction chambers. During sequencing, dye-labeled N-terminal amino acid (NAA) recognizers reversibly bind to their cognate NAAs, generating distinct binding patterns. Aminopeptidases sequentially cleave NAAs to expose subsequent amino acids for recognition. The order of recognizer binding and kinetic properties of recognition segments are analyzed to determine peptide sequence and post-translational modifications (PTMs). Here, we demonstrate Platinum’s ability to sequence interleukin 6 (IL-6) and fibroblast growth factor 2 (FGF2) to directly detect not only short peptides but also peptides that would undergo chemical modifications under LC-MS conditions. While non-targeted LC-MS/MS mapped nine peptides to IL-6, the peptide QIRYILDGISALRK was not detected due to N-terminal pyroglutamate modification. To facilitate LC-MS/MS detection of this peptide, a customized search for PTMs in Proteome Discoverer was needed. Further, non-targeted LC-MS/MS mapped 10 peptides to identify FGF2, but short peptides were initially missed. During the post-run analysis, a standard minimum peptide length filter was set to six amino acids, precluding detection of the RLYCK peptide of FGF2. In contrast, protein sequencing identified RLYCK and QIRYILDGISALRK peptides that escaped initial MS database matching.

This work demonstrates direct sequencing of oncogenic proteins, without peptide filters and customized searches, to detect short peptides and PTMs.

FALSE DISCOVERY RATE ESTIMATION IN SPECTRAL DECONVOLUTION IN TOP-DOWN PROTEOMICS

Kyoung Jeong¹,², Ayesha Feroz¹, Konstantin Nagornov³, Jiyoung Kim¹,², Timo Sachsenberg¹,², Yury O. Tsybin³, and Oliver Kohlbach¹,²,⁴

¹Applied Bioinformatics, Department for Computer Science, University of Tübingen, Sand 14, 72076 Tübingen, Germany
²Institute for Bioinformatics and Medical Informatics, University of Tübingen, Sand 14, 72076 Tübingen, Germany
³Spectroswiss, 1015 Lausanne, Switzerland
⁴Translational Bioinformatics, University Hospital Tübingen, Hoppe-Seyler-Str. 9, 72076 Tübingen, Germany

The complex signal structure of proteoforms and their product ions pose major challenges in top-down proteomics (TDP) data analysis. Spectral deconvolution (finding monoisotopic masses of the ions) is an early, but critical, preprocessing step for TDP. And yet, not enough research has been done on accurate estimation of false discovery rate (FDR) in deconvolution (the ratio of falsely deconvolved masses to all deconvolved masses). Here, we present a method to estimate the FDR of spectral deconvolution.

For the FDR estimation, we adopt the idea of the decay-free FDR estimation methods used in peptide identifications [1]. We attempt to find the second-best masses that would have been reported without the best masses (called “dummy masses”), and use a mixture distribution fitting method to fit the score distributions of dummy masses to the observed score distribution. From the fitted distribution that simulates the false positive distribution, FDR is estimated. This method is implemented in FLASHDeconv [2].

To evaluate our FDR estimation, “FTMS Simulator” software (Spectroswiss) was employed to generate in silico TDP datasets with varying resolutions (35k, 70k, and 140k) and noise levels (from 0 to 1e5). Each dataset contains 27 MS1 spectra from 534 distinct true masses (from 1k to 77k Da). On average, 38 masses are present per spectrum. The datasets were deconvolved by FLASHDeconv with FDR estimation activated. While different datasets showed highly variable false positive distributions, the fitted distribution followed them very closely. When the true and estimated FDRs were compared, the mean difference was less than 1.3 folds, with the maximum difference of only two folds. The results clearly demonstrate that our FDR estimation is accurate for in silico datasets. The method will be evaluated with MS/MS in silico and well-annotated experimental spectra. The estimation can be implemented with a simple modification in any deconvolution method.

Novel Aspect

A method to estimate false discovery rate in spectral deconvolution in top-down proteomics

References (Arial, Font Size 10)

[1] Peng et al., Bioinformatics, 36(26), i745-i753 (2020)
PIONEERING NATIVE LIQUID MALDI (NALIM) MASS SPECTROMETRY (MS) TO CHARACTERIZE MEMBRANE PROTEIN COMPLEXES AND OLIGOMER DISTRIBUTIONS

Zhamungui E¹, Orelle C², Jault JM², Jaubert J¹, Biliskov M¹,², Arizala JD³, Poget SF³, Boudvillain M¹, Suskiewisc M¹, Beaufour M¹, Cadene M¹,²

¹CBM CNRS UPR4301, affiliated with Université d’Orléans, Orléans, France. ²Molecular Microbiology & Structural Biochemistry (MMSB) UMR 5086, CNRS/University of Lyon, France. ³University of Zagreb, Zagreb, Croatia. ⁴Department of Chemistry, College of Staten Island, City University of New York, 2800 Victory Boulevard, Staten Island, New York 10314, United States. ⁵MO2VING platform, CBM CNRS UPR4301, Orléans, France

A challenge for the future of biology and medicine is to define the molecular mechanisms that underlie the biological function membrane proteins, which represent 2/3rds of potential drug targets. Gaining insight into their complexes in their native state is a challenge for most analytical approaches due to interference by lipophilic additives and difficulty of expression. Here native mass spectrometry (nMS) methods play an increasing role in support of high-resolution structure determination methods. ESI-based methods for nMS have reached the status of powerful, mature tools to dissect complexes while preserving the 3D structure and quaternary molecular assemblies of proteins [1]. However, detergents and other additives used for protein solubilization promote ion suppression, adduction, and degraded instrument performance. We previously leveraged MALDI’s greater tolerance to contaminants to conceive NALIM (Native Liquid MALDI), a new nMS method based on liquid matrix [2]. Here we show development of NALIM on a MALDI-TOF/TOF instrument for different types of membrane proteins, and for straightforward access to large oligomer distributions.

Instrumental parameters and sample preparation play a key role in the detection and preservation of membrane protein complexes. As the role of oligomers in biology seems to gain interest and prominence, NALIM was used to characterize oligomer distributions of different proteins, showing its promise as a tool to understand their dynamics. The Tx7335 actitoxin binding site on KcsA potassium channel [3] was found to differ from pore blockers’ site. Furthermore, NALIM was used to demonstrate that the dimer of the ABC family BmrA multidrug resistance efflux pump is stabilized through ligand–binding. Finally, NALIM successfully ejected BmrA from membrane vesicles (MVs), proving that, like its more evolved ESI cousin, analysis can be performed from the protein’s natural lipid environment, raising the exciting possibility that NALIM may someday detect membrane proteins from live cells.

Novel Aspect: NALIM as a new nMS tool to characterize membrane proteins in increasingly in vivo-like conditions.

References:
PATIENT-SPECIFIC ISOFORM CHARACTERIZATION VIA SPLICE- AND VARIANT-AWARE LONG-READ PROTEOGENOMICS

David Wissel¹, Madison Mehlerber², Marina Panteli¹, Mayank Murali², Saikat Bandyopadhyay², Roger Volden³, Elizabeth Tseng³, Gloria Sheynkman², Mark D. Robinson¹

¹Department of Molecular Life Sciences, University of Zurich, Zurich CH, 8057
²Department of Molecular Physiology and Biological Physics, University of Virginia, Charlottesville VA, 22903
³Pacific Biosciences, Menlo Park CA, 94025

An accurate prediction of all primary protein isoform sequences—including how all genetic- and splice-variation converge to produce such products—is critical to support proteoform discovery efforts and personalized medicine applications. Recently, construction of sample-specific protein databases derived from long-read RNA-seq has been shown to enhance characterization of the human isoform landscape [1], however; genetic variants were not captured [1]. Here, we present a pipeline that concurrently extracts and resolves both genetic and splicing information directly from high accuracy, long-read RNA sequencing data for proteomics analysis. For pipeline development, we collected Iso-seq long-read RNA-sequencing data and deep coverage bottom-up MS from a human iPSC line. We optimized the DeepVariant [2, 3] variant caller to extract variant information directly from long reads, and phased them using WhatsHap [4]. We determined full-length splice isoform sequences using Iso-Seq3 and SQANTI3 [5]. Integration of such data enabled construction of a full-length, allele-phased, isoform-resolved database that was used for MS searching with MetaMorpheus [6, 7]. We called 46,210 variants that passed our QC filters, of which 5,740 were indels and 40,470 SNVs. After phasing, we were left with 44,648 variants, 19,312 homozygous, and 25,336 heterozygous, respectively.

Our modified protein database contained an additional 2,804 protein isoforms due to genetic variation, relative to a database constructed only with splice variations. After MS search, we identified (< 1% FDR) 10,182 genes and 67,722 peptides, of which 250 were novel (54 and 196 were novel due to splice- and genetic variation, respectively). Lastly, we identified 15,175 peptides uniquely mapping to 2,408 phased protein isoforms, giving potential evidence of allele-specific isoform expression. Our method should extend the applicability of long-read-derived protein models for personalized medicine applications from cancer immunotherapy to human disease genetics.

This work describes a pipeline that builds patient-specific protein databases containing splice- and genetic variants from sample long-read RNA-seq.


ACHIEVING HIGHER PRODUCTIVITY IN CELL LINE OPTIMIZATION BY A STREAMLINED MIDDLE-DOWN WORKFLOW

Michael Poltash¹, Zoe Zhang², Haichuan Liu², Andy Mahan¹, Hirsh Nanda¹, Partha Chowdhury³

¹Janssen Pharmaceuticals R&D, Spring House, PA, ²SCIEX Redwood City, CA

Biotherapeutics hold incredible promise in alleviating life altering illnesses; however, the cost of manufacturing and delivery to patients is prohibitive. To lower production costs, media formulations, intensified bioreactor processes and advancements in CHO cell protein expression have all contributed to increases in manufacturing productivity. Changing culture conditions, however, induces cellular stress, adversely impacting product quality profile and creating additional challenges in manufacturing.

Biotherapeutic development and their respective characterization by mass spectrometry (MS) has matured greatly over the last few decades. Peptide mapping provides detailed information about protein sequence and modifications, but a considerable amount of effort is required for analysis. Advances in enzyme technology coupled with higher performance MS have enabled highly efficient middle-down fragmentation that reduce the risk of introducing artificial PTMs as induced by peptide mapping. Here, we explore applications of a streamlined middle-down workflow for biotherapeutics in an intensified high-titer culture process.

Two enzymes, GlySERIAS and Fabricator (Genovis), were used to digest a multi-specific antibody to generate consistent subunits before being analyzed. LC-MS data were acquired using a QToF MS equipped with EAD and data was analyzed using commercially available software. Higher than 70% sequence coverage was achieved for all subunits using just a single injection. Post translational modifications such as glycosylation were successfully characterized through middle-down fragmentation. Results were validated against traditional methods of LC-MS characterization. Moreover, structural confirmation of engineered disulfides in a multi-specific molecule was interrogated, something that was challenging to access by peptide mapping. Middle-down analysis provides a new approach for biotherapeutics characterization with fewer processing steps reducing the chance of introducing PTM artifacts or disulfide scrambling. It provides a unique tool to help provide a comprehensive molecular profile in biopharmaceutical manufacturing.

Novel Aspects: Adoption of a streamlined single injection middle-down workflow in real-world cell line optimization.
INFLUENCE OF SAMPLE PREPARATION AND PROTEOFORM SEPARATION FOR IN-DEPTH TOP-DOWN PROTEOMICS

Philipp T. Kaulich¹, Andreas Tholey¹

¹Proteome Research & Bioanalytics, Institute for Experimental Medicine, Christian-Albrechts-Universität zu Kiel, Kiel, Germany

Sample preparation and protein separation are crucial steps for in-depth proteoform identification by top-down proteomics (TDP). As large-scale TDP studies are currently limited to the analysis of proteoforms smaller ~30 kDa,¹ many TDP studies use methods for the enrichment of small proteins as sample preparation, such as gel-based approaches, size-exclusion chromatography, or molecular weight cut-off (MWCO) filter.

We investigated the influence of established TDP sample preparation methodologies in terms of the number, confidence, reproducibility, (artificial) modifications, and physicochemical properties of the identified proteoforms. For this, we lysed human Caco-2 cells, purified the proteins by precipitation, and performed multiple TDP sample preparation and proteoform separation strategies including MWCO filter, PEPPi, solid-phase extraction, size-exclusion chromatography, GELFrEE, organic solvent depletion methods. In addition, we tested several variations of the approaches, e.g., different solid-phase extraction materials or MWCO sizes. We compared the proteoform identification with a 1D-LC-MS/MS strategy, where the proteins were just resuspended after precipitation without further sample treatment. All sample preparations were performed in three replicates and injected twice for LC-MS/MS analysis using two different multi-CV FAIMS methods targeting the low-/medium- and high-molecular-weight range, respectively.²

The proteoforms identified by the different sample preparation methods differ in many properties, such as molecular weight distribution, hydrophobicity, post-translational modifications, and identified protein groups. Notably, different sample preparation methods show a bias for certain proteoform sub-populations, e.g., targeting more hydrophilic, acidic, or larger proteoforms. In addition, most sample preparation methods showed an advantage compared to the 1D-LC-MS/MS analysis in terms of the number of identified proteoforms.

With the knowledge of the biases of identified proteoforms in the different approaches, a rational selection based on the research question will become possible.

Novel Aspect
A large-scale comparison of established TDP approaches shows a significant impact on the identifiable proteoforms.

References

TOP-DOWN PROTEOMIC IDENTIFICATION OF PLASMID-ENCODED PROTEINS FROM PATHOGENIC BACTERIA USING MALDI-TOF-TOF MASS SPECTROMETRY

Clifton K. Fagerquist,¹ Yanlin Shi¹ and Jihyun Park¹;²

¹Western Regional Research Center, Agricultural Research Service, USDA, Albany, CA, USA
²Research Participation Program administered by the Oak Ridge Institute for Science and Education, U.S. Department of Energy. Oak Ridge TN, USA

Abstract

Pathogenic bacteria often carry plasmids (small circular pieces of DNA separate from the host genome) that encode genes that confer survival benefits to a microorganism. Plasmids can encode genes for antibiotic resistance, virulence factors, stress response, etc. Although there is a metabolic cost to retain and maintain these plasmids, the benefits to the host often outweigh the metabolic burden. Plasmids can also be transferred to other bacteria by horizontal gene transfer which can spread antibiotic resistance and virulence factors. Characterizing the protein products of these plasmid genes is critical to understanding their role in bacterial survival and virulence. We have identified and characterized proteins (and their post-translational modifications) of immunity proteins whose genes are encoded by plasmids carried by a Shiga toxin-producing Escherichia coli (STEC) isolated from an agricultural region in Arizona (USA) using plasmid induction techniques, DNA sequencing, Alphafold2 protein structures, MALDI-TOF-TOF mass spectrometry and top-down proteomic analysis. Colicin E8 is a DNAse bacteriocin whose gene and its immunity cognate (ImmE8) are encoded by a ~7 kb plasmid. Colicin E8 attaches to and invades neighboring bacteria destroying their DNA and thus conferring a competitive survival advantage to the host that produced it. ImmE8 inhibits the action of colicin E8 until such time as it is released from the host. Interestingly, ImmE8 was detected with and without an attached acyl group (+308 Da) at its only cysteine residue. Such a PTM has not been previously reported for immunity proteins and suggests that acyl attachment (via a thioester bond) may serve to anchor ImmE8 to the inner membrane of the host until it is needed to inhibit endogenous colicin E8. Alternatively, it may act to anchor ImmE8 to the outer membrane of the host to inhibit the activity of exogenous DNAase bacteriocins.

Novel Aspect

Top-down proteomic characterization of plasmid-encoded proteins of STEC using MALDI-TOF-TOF mass spectrometry and Alphafold2
2D-GELC-FAIMS-MS WORKFLOW FOR IN-DEPTH MIDDLE-DOWN PROTEOMICS

Nobuaki Takemori¹, Ayako Takemori¹, Philipp T. Kaulich², and Andreas Tholey²

¹Advanced Research Support Center, Institute for Promotion of Science and Technology, Ehime University, Ehime, Japan
²Systematic Proteome Research & Bioanalytics, Institute for Experimental Medicine, Christian-Albrechts-Universität zu Kiel, Kiel, Germany

SDS-PAGE is a size-based protein separation method and is widely used for its ease of handling and high reproducibility. In order to utilize SDS-PAGE as a sample preparation method for top-down proteomics, we previously developed two methods for the rapid and efficient recovery of proteins in PAGE gels after electrophoresis: (1) protein elution using BAC cross-linked polyacrylamide gels that can be rapidly dissolved by reduction treatment (BAC-PAGE)¹ and (2) highly efficient passive extraction using CBB and SDS (PEPPI-MS)²,³. In this study, we developed two different middle-down proteomics (MDP) approaches. MDP is an analytical approach in which protein samples are digested with proteases such as Glu-C to generate large peptides (>3 kDa) that are then analyzed by mass spectrometry. This method is useful for characterizing high-molecular-weight proteins that are difficult to detect by top-down proteomics, in which intact proteins are analyzed by mass spectrometry. The first approach is GeLC-FAIMS-MS, a multidimensional separation workflow that combines gel-based prefrionation with LC-FAIMS Orbitrap MS analysis. Whole Glu-C digests of biological samples are first size-fractionated by polyacrylamide gel electrophoresis, followed by RP-LC separation and additional ion mobility fractionation, resulting in a significant increase in peptide length detectable by mass spectrometry. The second approach encompasses an innovative sample pre-frionation workflow, 2D-GeLC-FAIMS-MS, for in-depth MDP by combining BAC-PAGE and PEPPI-MS into a two-dimensional gel fractionation aided by limited Glu-C digestion. In this workflow, samples are first size fractionated by BAC-PAGE and then subjected to in-gel digestion with Glu-C. The resulting middle-down peptides below 50 kDa are further fractionated by second dimension SDS-PAGE with MES running buffer and finally subjected to LC-FAIMS-MS analysis after recovery by PEPPI-MS. The dissolution properties of the BAC gels allow for sample transfer with minimal loss between two PAGEs, resulting in high-resolution sample pre-frionation for in-depth MDP.

PAGE-based fractionation at the protein and peptide levels effectively reduces the complexity of middle-down samples and enables in-depth analysis.

COLLOIDAL ASSEMBLIES FOR ON-CHIP PROTEIN SEPARATION AND TRANSFER FOR TOP-DOWN MS ANALYSIS

Tanushree Dutta¹, Julea Vlassakis¹
¹Department of Bioengineering, Rice University, Houston, Texas

Nanoscale systems can remarkably improve protein recovery by downscaling of processing volumes and efficient chip-to-MS interfacing⁶. Nanofluidic and polymer sieves have been used in microfluidic devices for fast separation of DNA and proteins for fluorescence-based readouts²-⁴ including at the single cell level⁵. For MS-based readout, the focus is more on minimizing (protein) losses⁷,⁸. In particular, several top-down MS sample processing steps lead to substantial losses⁹. Additionally, sample prefractionation is useful for intact protein analysis⁸ but it requires efficient chip-MS interfacing. Here, we report the development of a multifunctional nanosieving system for fractionation and delivery using nanoparticle (NP) assemblies decorated with 1) stimuli-responsive ligands for ‘remote control’ of nanostructure disassembly for efficient chip-to-MS protein transfer and 2) benzophenone (BP) to covalently immobilize protein analytes to the NP upon UV excitation.

Microchannel cast in a polydimethylsiloxane (PDMS) substrate was fabricated using a previous protocol⁹. Gold NPs (15 nm) were functionalized with thiol-modified oligonucleotides and DNA-gold were crystallized by adding linkers to the NP solution and heating (to allow the linker to fully hybridize) and slow cooling through the melting temperature (T_m)¹⁰,¹¹ and modified with benzophenone (BP). SEM images of the gold-DNA showed large assembled networks compared to naked gold (control), which was also confirmed by ToF-SIMS. The stability of the assemblies under applied electric potential and methods to screen surface charges will be tested in a series of experiments for optimizing protein (electrophoretic) separation. For testing protein immobilization, a fluorescently-labeled protein will be immobilized onto gold-DNA assemblies and fluorescent intensities will be compared with and without BP. For achieving NP disassembly, assemblies will be heated at temperature above the T_m for retrieval and transfer for testing protein recovery using QToF.

This work describes a stimuli-responsive immobilization nanomaterial we are designing to sieve and deliver protein for low-cell number MS analysis.

References

ADVANCING EXD-BASED TOP-DOWN ORBITRAP ANALYSIS OF GLYCOPROTEINS BY ACQUISITION AND PROCESSING OF TRANSIENTS AND FULL PROFILE MASS SPECTRA

Yury Tsybin

1 Spectroswiss, 1015 Lausanne, Switzerland

Top-down MS analysis of proteins in general, and glycoproteins in particular, requires exceptional MS/MS performance. In addition, to maximize the information output from the raw (unreduced) data generated with the Orbitrap platforms, it is preferred to process time-domain transients or full profile mass spectra represented in the enhanced FT, or eFT, mode. The full profile eFT mass spectra are readily available on the tribrid Orbitrap platforms. However, the time-domain Orbitrap transients are not readily available and full profile eFT mass spectra reach very large sizes, exceeding 20 GB per LC-MS/MS run.

To enable acquisition and processing of the unreduced data in Orbitrap top-down MS, we employ external high-performance data acquisition systems, FTMS Boosters, and advanced data processing software, Peak-by-Peak. The former provides access to the time-domain transients from any Orbitrap platform and the latter is capable to process datasets of any size, including the full profile eFT mass spectra. Beneficially for top-down MS analysis, we can average data from several technical replicates, thus further enhancing product ion analysis sensitivity and protein sequence coverage.

In this presentation, we will present the recent advances in the top-down MS applications that benefit from the unreduced data processing. Applications in glycoprotein, including monoclonal antibodies, sequencing and modifications analysis demonstrate high analytical specificity, sensitivity, and quantitative precision. The approach intrinsically supports automation and provides transparent data processing.

Furthermore, we will demonstrate the utility of direct (without the untargeted deconvolution) feature extraction by similarity scoring between the experimental and accurately simulated Orbitrap mass spectra. In our approach, we simulate the FTMS and MS/MS spectra in profile mode using the FTMS Simulator (Spectroswiss) for the given experimental parameters and FTMS instrument type and model using suspect databases containing peptide and protein sequences and possible modifications [1, 2].

Novel Aspect

Enabling the acquisition and processing of transients and full profile mass spectra from Orbitraps benefits ExD-based top-down analysis of glycoproteins

ASSESSMENT OF LABELED PURIFIED PROTEINS WITH TOP-DOWN PROTEOMICS

Aron Phong³, John J.H. Shin¹

¹Alkahest – A Grifols Company, San Carlos, CA

Common strategies used in the interrogation of biological pathways in both cells and tissues are through the localization of proteins of interest by microscopy and the identification of their binding partners through immunoprecipitation mass spectrometry (IP-MS). Antibodies or sequence tagged proteins for these tasks are often either commercially unavailable, or time-intensive or cost-prohibitive to obtain. Labeling purified proteins with popular chemical reagents such as N-hydroxysuccinimidobiotin (EZlink) for use in microscopy and IP-MS studies provides an affordable, quick, and practical alternative. However, a robust quality-control workflow is needed to monitor for possible ablation of critical functional domains from over-labeling and to detect any unacceptable degradation in the label over time. Here, we present a simple labeling and quality control workflow that utilizes top-down proteomics (TDP) for rapid assessment of labeling efficiency and label degradation. We use this approach to compare labeling between EZlink and an increasingly popular trans-Cyclooctene-derived label and show the latter suffers from degradation that may impact its functional use.

This work utilizes TDP to rapidly assess the chemical-labeling of purified proteins of interest for use in functional biological assays.
CHARACTERIZATION OF PATIENT SERUM MONOCLONAL PROTEINS BY BIOLAYER INTERFEROMETRY COUPLED WITH HIGH-RESOLUTION MASS SPECTROMETRY

Priscilla S.-W. Yeung¹,², Yajing Liu¹, Run-Zhang Shi¹,², Ruben Y. Luo¹,²

¹Department of Pathology, Stanford University, Stanford, CA, USA
²Clinical Laboratories, Stanford Health Care, Palo Alto, CA, USA

Multiple myeloma is the second most common hematolymphoid malignancy in the United States and has a median survival of 6 years after diagnosis [1, 2]. It is part of a larger group of plasma cell neoplasms that range from monoclonal gammopathy of uncertain significance (MGUS), to smoldering myeloma (SM), to fulminant multiple myeloma (MM), and include extramedullary manifestations such as plasmacytomas or amyloidosis [1, 2]. The most common laboratory test used to screen for plasma cell disorders is serum protein immunofixation electrophoresis (SPIE), which identifies intact monoclonal antibodies (M-proteins) secreted into the serum by clonal populations of plasma cells [1]. Another clinical marker of plasma cell neoplasms is serum free light chains (FLC), which are circulating antibody light chains measured by immunoassay [1]. However, both of these methods can yield results that are hard to interpret in some patient groups [3]. In addition, a recent publication showed that MGUS patients with glycosylated light chains had an increased risk of progression to MM [4], demonstrating that post-translation modifications in these proteins can be used as prognostic biomarkers. In this study, we developed an assay that couples a biolayer interferometry immunocapture step with high-resolution mass spectrometry to determine the masses of serum monoclonal proteins to the 1 Dalton level. This method helps resolve ambiguous SPIE or FLC immunoassay results, and more importantly, allows us to identify post-translational modifications on the M-proteins and FLCs. Because light chain N-glycosylation has been shown to have prognostic value [4], we hypothesize that specific glycosylation patterns and other post-translational modifications on M-proteins and FLCs can be used as clinical biomarkers for monoclonal gammopathies.

This work applies a novel approach to characterize serum monoclonal proteins for discovery of improved biomarkers for plasma cell disorders.

QUADRUPOLE ISOLATION AND CHARACTERIZATION OF SPECIFIC PROTEOFORMS AND PROTEIN COMPLEXES AT A HIGH M/Z RANGE USING ORBITRAP ASCEND

Rafael D Melani¹; Christopher Mullen²; Linda B Lieu²; Jingjing Huang¹; David Bergen¹; Graeme C McAlister¹; Amanda Lee¹, Romain Huguet¹; Vlad Zabrouskov³; Luca Forcini²

¹Thermo Fisher Scientific, San Jose, CA. ²University of Oklahoma, Norman, OK.

Recent advancements in mass spectrometry (MS) and native MS have greatly enhanced the identification capabilities of proteoform and protein complexes. Improved ion transmission, multiple ion activation types, and high-resolution Orbitrap mass analyzers enable more accurate and sensitive detection and characterization of proteoforms. However, precise protein complex and proteoform-specific characterization require high isolation resolution in the high m/z range, especially for native MS. Therefore, we modified an Orbitrap Ascend Tribrid to perform quadrupole isolation with high efficiency and accuracy up to m/z 8,000. We report the characterization of NIST mAb (146 kDa), and tetrameric pyruvate kinase (232 kDa), statically sprayed, under native conditions. For NIST, single glycoforms of the most abundant charge state were quadrupole isolated, and after isolation, the proteoforms were fragmented with different ion activation techniques, including HCD, ETD, EThC, and UVPD. Combining all fragmentation types, we achieve high sequence coverage for a native MS experiment and identified heavy chain fragments containing the different glycan moieties specific for each distinct proteoform. For pyruvate kinase, the most intense charge state of the homotetramer complex, four identical subunits (232 kDa), was isolated in the quadrupole, and the subunits were ejected using HCD. A single subunit proteoform was identified, and further fragmentation confirmed the amino acid sequence and the presence of acetylation at the N-terminus. Subsequently, we isolated the complex with one truncated monomer missing the first 22 N-terminal amino acid residues (229.5 kDa), corresponding to ~10-15% of the intensity of the main complex proteoform. MS² and MS³ experiments annotated and characterized the two distinct proteoforms and confirmed their stoichiometry. The use of quadrupole isolation allows the isolation of closely spaced proteoforms and protein complexes, improves the MS² signal-to-noise ratio, and reduces interference from other ions. It affords deep proteoform/complex characterization on the Tribrid platform in conjunction with different ion activation techniques.

Novel Aspect (12/20): New quadrupole isolation with high efficiency and accuracy up to m/z 8,000
TOP-DOWN CHARACTERIZATION OF NATIVE MONOCLONAL ANTIBODIES OBTAINED WITH ELECTRON CAPTURE DISSOCIATION ON Q-ToF INSTRUMENTS

John Sausen¹,
Yury V. Vasil’ev, Rachel Franklin, Michael C. Hare, Adrian Guthals²,³
Joseph S. Beckman³

¹Agilent Technologies, Americas Field Organization, Wooddale IL, USA
²e-MSion, a part of Agilent., OR, USA
³Linus Pauling Institute, OSU, OR, USA

Monoclonal antibodies (mAbs) are an important class of therapeutics used for treatment of a variety of illnesses from cancer to autoimmune diseases. Development of mAbs requires monitoring for post-translation modifications and structural changes. Mass spectrometry (MS) with bottom-up and middle-down approaches are the most common methods for mAbs analysis. Both presume long more extensive sample preparation procedures and may suffer from a partial loss of information regarding changes in the proteoforms. Top-down MS offers the potential to more effectively characterize proteoform’s, and simplify sample Preparation. This next generation ExD cell demonstrates how electron capture dissociation (ECD) with a top-down approach can be used for improved characterization of mAbs with high efficiency.

Several intact monoclonal antibodies, including NIST mAb, Sigma mAb, Infliximab were fragmented using ECD with and without CID complementary activation. Manual and autotune procedures along with ExDViewer real-time streaming was used to tune the ExD cell for the best total ECD efficiency for both LC and HC monoclonal antibodies, using native carbonic anhydrase.

This work demonstrates exceptional sequence coverage of intact native monoclonal antibodies obtained by ECD fragmentation in a Q-ToF instrument.

*J. Am. Soc. Mass Spectrom.* 2021, 32, 8, 2081–2091, [https://doi.org/10.1021/jasms.0c00482](https://doi.org/10.1021/jasms.0c00482)
INFORMATICS OF INDIVIDUAL ION MASS SPECTROMETRY ENABLES SINGLE CELL TOPODOWN PROTEOMICS

Michael A. R. Hollas,¹ Pei Su,¹ Stanislav S. Rubakhin,² Fatma Ayaloglu Butun,¹ Joseph B. Greer,¹ Bryan P. Early,¹ Ryan T. Fellers,¹ Michael A. Caldwell,³ Jonathan V. Sweedler,² Jared O. Kafader,¹ and Neil L. Kelleher¹

¹Proteomics Center of Excellence, Chemistry of Life Processes Institute, Northwestern University, Evanston, IL
²Beckman Institute and Department of Chemistry, University of Illinois at Urbana-Champaign, Urbana, IL, USA

Here we introduce a novel informatics platform for single-cell Proteoform imaging Mass Spectrometry (scPIMS) which seeks to capture heterogeneity in single cells that is often obscured by population-averaged protein measurements, allowing for the characterization of phenotypic states of cells in complex tissues.¹,² scPIMS integrates a liquid MS imaging probe with individual ion MS (iPIMS), an emerging technology for the detection of single ions, allowing for enhanced sensitivity (>500x) and 10x higher resolving power over traditional Orbitrap spectra.²,³ Using scPIMS, 10,836 rat hippocampal cells were profiled over a 10 day period, detecting 472 and identifying 169 single-cell proteoforms, including established biomarkers like ENOG and GFAP isoforms which stratify brain cell types without antibody staining. Utilizing a subset of selected cell-type markers, scPIMS was able to stratify 2755 cells into three different cell types: astrocytes (1524), microglia (720), and neurons (511). Raw data were collected as continuous chronograms containing discrete peaks corresponding to cell events on glass slides. Peaks were filtered to remove poor quality and colocated cells. Individual ions associated with these single cell features were aggregated into a single cell spectrum. These cell specific individual ions were searched against an isotopic envelope library and matched to specific isotopic peaks. Matching ions were scored based on the theoretical relative intensity of the matching peak, and the mass error of the observed individual ion, calculated via a normal cumulative distribution function with a mean and sigma determined by the theoretical mass and width of the isotopologue peak respectively. An empirical false discovery rate procedure was implemented, where decoy proteoforms were generated, scored alongside proteoform hits, rank ordered, and given a q-value. Cell type scores were generated using FDR controlled proteoforms scores for selected cell type markers, allowing cells to be stratified.

Novel Aspect
A novel informatic approach enables a scalable single-cell proteomics platform utilizing the sensitivity and resolution of individual ion mass spectrometry.

References

Wörner, T.P. et al. Resolving heterogeneous macromolecular assemblies by Orbitrap-based
COMPARISON OF SINGLE MOLECULE PROTEIN SEQENCING AND TOP-DOWN PROTEOMICS

Nickolas P. Fisher¹, Maanasa Narayanaamoorthy¹, Jordan Summers², Danielle Tullman-Ercek²,³, Neil L. Kelleher¹,³,⁴

¹Department of Chemistry, Northwestern University, Evanston, IL; ²Department of Chemical & Biological Engineering, Northwestern University, Evanston, IL; ³Chemistry of Life Processes Institute, Northwestern University, Evanston, IL; ⁴Proteomics Center of Excellence, Northwestern University, Evanston, IL

Abstract

Proteomics has been dominated by mass spectrometry-based methods due to its excellent sensitivity and integration with front-end separation methods for high-throughput analysis of complex biological samples. Though next-generation sequencing (NGS) methods are frequently employed for DNA and RNA sequencing, the application of NGS-like methods to protein samples has been challenging due to the large number of amino acids that encode protein biopolymers. Recently, a single molecule protein sequencing (SMPS) method was realized by Quantum-Si Inc.¹ Using this system, N-terminal amino acid residues are sequenced by monitoring the association and dissociation rates of amino acid recognizers with analyte residues through fluorescence lifetime and intensity measurements. This method provides rich data that can confidently identify 12 of the 20 canonical amino acids and distinguish post-translational modifications (PTMs). Though development of this technology is ongoing, there is great interest in comparing SMPS performance to top-down mass spectrometry proteomics methods to identify overlap in utility and opportunities for cooperativity. To carry out this comparison, we chose to identify the “secretome” from a lab strain of Salmonella, using both the NGS platform and top-down mass spectrometry using collision induced dissociation to generate comparable data. To this end, we employ the Salmonella Type III secretion system (T3SS) to secrete both native and heterogeneous model proteins, providing controls for the project and enabling the evaluation of proteoform investigation and sequencing capability across both methods. Additionally, this comparison may further inform how the population of secretory proteins change upon different modes of T3SS activation. Sequence coverage and ability to identify PTMs along the protein backbone are being compared to understand the advantages and disadvantages of each method for identifying proteoforms. Evaluation of method performance will reveal potential cooperativity between SMPS and top-down mass spectrometry with single molecule counting for improved analysis of complex protein samples.

Novel Aspect

First report on NGS protein sequencing in comparison with top-down mass spectrometry.

References

PROTEOFORM IDENTIFICATIONS IN HUMAN TISSUE USING INTACT MASS, CYSTEINE COUNTING VIA ISOTOPIC CHEMICAL LABELING, AND A PROTEOFORM ATLAS

John G. Pavek¹, Brian L. Frey¹, Dustin C. Frost², Ting-Jia Gu², Lingjun Li², Lloyd M. Smith¹

¹Department of Chemistry, University of Wisconsin-Madison;²School of Pharmacy, University of Wisconsin-Madison

Top-Down proteomics is the dominant method for proteoform characterization in complex mixtures and utilizes both intact mass and fragmentation spectra. While this strategy produces detailed molecular information, it also requires extensive instrument time per proteoform identification, thereby compromising the depth of proteoform coverage accessible on liquid chromatography timescales. Such top-down analysis is necessary for making new proteoform identifications, but once a proteoform has been confidently identified, the extensive characterization it provides may no longer be required for a subsequent identification of the same proteoform. We present a strategy to identify proteoforms in tissue samples based upon the combination of an intact mass determination with a measured count of the number of cysteine residues present in each proteoform. We developed and characterized a NeuCode cysteine tagging chemistry suitable for the efficient and specific labeling of cysteine residues within intact proteoforms, which provides the count of cysteine amino acids present. On simple protein mixtures, the tagging chemistry yields greater than 98% labeling of all cysteine residues, with a labeling specificity of greater than 95%. Similar results are observed on more complex samples. In a proof-of-principle study, proteoforms present in a human prostate tumor biopsy were characterized. Observed proteoforms, each characterized by an intact mass and a cysteine count, were grouped into proteoform families. We observed 2190 unique experimental proteoforms, 703 of which were grouped into 275 proteoform families. We compared the proteoform identification results obtained by this intact mass plus cysteine count strategy using both a conventional database and a more specific proteoform atlas. In addition to their utility for counting cysteines, the NeuCode chemical labels offer promise for precise proteoform quantification.

Novel isotopic chemical labeling reagents enabled cysteine counting, which when combined with a proteoform atlas identified proteoforms without fragmentation spectra.

PTM-FOCUSED TOP-DOWN PROTEOFORM ANALYSIS

Jacob Seay¹, Gustavo Agreda¹, Bryon Drown¹

¹Department of Chemistry, Purdue University, West Lafayette, IN

Precise characterization of proteoforms has greatly improved with advances in high-performance mass spectrometry and separations, but routine discovery top-down proteomics experiments often suffer from abundance biases when studying unenriched samples. Likewise, the analysis of concurrent post-translation modifications (PTMs) is made difficult by their low relative abundance. To combat this problem, enrichment approaches compatible with intact protein mass spectrometry have been developed targeting specific classes of PTM. These PTM-focused TDP analyses will serve as a platform for PTM crosstalk investigation and expand our coverage of precisely characterized proteoforms.

This work applied modern chemical TDP approaches to achieve improved proteoform characterization by PTM-based enrichment.
Top-Down Proteomics Platform Enabled by Photocleavable Surfactant Azo for the Comprehensive Characterization of Endogenous Phospholamban

Holden Rogers¹, David Roberts¹, Eli Larson¹, Jake Melby¹, Kalina Rossler², Austin Carr¹, Kyle Brown¹,³ and Ying Ge¹,⁴,⁵

Department of Chemistry¹; Department of Molecular and Cellular Pathology²; Department of Surgery²; Department of Cell and Regenerative Biology⁴; Human Proteomics Program⁵, University of Wisconsin-Madison, WI

Abstract:

Top-down mass spectrometry (MS)-based proteomics has become a powerful tool for analyzing intact proteins and their associated post-translational modification (PTMs).¹ In particular, membrane proteins play critical roles in cellular functions and represent the largest class of drug targets. However, the top-down MS characterization of endogenous membrane proteins remains challenging, mainly due to their intrinsic hydrophobicity and low abundance. Phospholamban (PLN) is a regulatory membrane protein located in the sarcoplasmic reticulum and is essential for regulating cardiac muscle contraction.² PLN has diverse combinatorial PTMs and their dynamic regulation has significant influence on cardiac contractility and disease. Herein, we have developed a rapid and robust top-down proteomics method enabled by a photocleavable anionic surfactant, Azo, for the extraction and comprehensive characterization of endogenous PLN from cardiac tissue. We employed a two-pronged top-down MS approach using an online reversed-phase liquid chromatography tandem MS (LC-MS/MS) method on a quadrupole time-of-flight (Q-TOF) MS and a direct infusion method via an ultrahigh-resolution Fourier transform ion cyclotron resonance (FTICR) MS. We have comprehensively characterized the sequence and combinatorial PTMs of endogenous human cardiac PLN. We have shown the site-specific localization of phosphorylation to Ser16 and Thr17 by MS/MS for the first time and the localization of S-palmitoylation to Cys36. Moreover, we applied our method to characterize PLN in disease and reported the significant reduction of PLN phosphorylation in human failing hearts with ischemic cardiomyopathy (ICM). Taken together, we have developed a streamlined top-down targeted proteomics method for comprehensive characterization of combinatorial PTMs in PLN toward better understanding the role of PLN in cardiac contractility.

Novel Aspect:

Membrane protein extraction and TDP-MS enabled characterization of phospholamban proteoforms including the first site-specific phosphorylation localization and application to ICM.

References:


ADVANCING TOP-DOWN PROTEIN ANALYSIS BY INTEGRATED TANDEM-TIMS-UVPD-PASEF AND TIME-RESOLVED TANDEM-TIMS APPROACHES.

Christian Bleiholder, Fanny C. Liu, Thais Pedrete, Jusung Lee

Department of Chemistry and Biochemistry, Florida State University, Tallahassee, FL

Top-down proteomics provides insights into sequence variants as well as the localization of post-translational modifications of intact proteins. However, achieving complete sequence coverage via top-down proteomics, a prerequisite for localizing all potential proteoforms, remains technically challenging.

We recently developed a tandem-trapped ion mobility spectrometer/mass spectrometer (tandem-TIMS/MS) equipped with a 213 nm UV laser setup [1, 2]. Ions eluting from the first TIMS cell can be stored in a quadrupole ion trap inserted between the two TIMS cells and dissociated via UV irradiation prior to a second mobility analysis in TIMS-2 cell. We showed that UV photodissociation on protein ions at a pressure range compatible with ion mobility measurements (~2-3 mbar) is feasible.

In this work, we explore the potential of the tandem-TIMS/MS method for native top-down protein analysis in a two-fold approach. In the first approach, parallel accumulation serial fragmentation (PASEF) is incorporated into the tandem-TIMS-UVPD measurements. PASEF is an automated MS/MS workflow which utilizes ions of specific m/z and mobility as precursor ions. Protein fragment ions produced via UVPD are mobility separated in TIMS-2 and subjected to PASEF MS/MS analysis. Additionally, we developed a data analysis workflow for top-down proteomics by using the OMSSA scoring method to increase the accuracy of the MS³ fragment peaks assignment. We discuss the tandem-TIMS-UVPD-PASEF workflow on intact standard reference proteins in the range from 8.6 kDa to ~150 kDa.

In the second approach, we investigate the cross sections of top-down protein fragment ions by time-resolved measurements in tandem-TIMS/MS. Protein fragment ions generated in-situ by collision-induced dissociation or UV photodissociation in tandem-TIMS display conformational heterogeneity that cannot be rationalized by mass or charge alone. For example, distinct ion mobility spectra are observed for in-situ generated \( y_n^{3+} \) ions from proteins avidin, ubiquitin, and cytochrome c despite identical sequence length and charge state, and only minor difference in mass. The \( y_n^{4+} \) ubiquitin ions are trapped in the second TIMS cell for 5 seconds which induces a shift towards smaller cross section in the ion mobility spectra [3]. These structural changes upon trapping in the gas phase underline the presence of metastable fragment ion conformations, which suggests that top-down fragment ions might retain memory of the precursor structure. Overall, the data indicates that the cross sections can be exploited as sequence specific determinants of top-down fragment ions.

Novel Aspect:
Tandem-TIMS/MS coupled with CID/UVPD enables a conformationally-resolved approach to native top-down analysis of protein systems.


CHARACTERIZING HISTONE AND HISTONE-MODIFYING PROTEINS FROM THE GREEN MICROALGA *CHLAMYDOMonas* REINHARDTII BY TOP-DOWN MASS SPECTROMETRY

Emily K. Wright\(^1\), Megan S. Bindra\(^1\), Michael W. Hayes\(^1\), James G. Umen\(^2\) and James J. Pesavento\(^1\)

\(^1\)Saint Mary’s College of California, Moraga CA  
\(^2\)Donald Danforth Plant Science Center, St. Louis MO

Histones are a family of essential proteins that regulate many epigenetic phenomena such as DNA packaging and regulation of gene expression\(^1\). While these proteins have been heavily studied in organisms from the animal kingdom, they are currently understudied in the plant kingdom. The Pesavento Lab aims to fill in these gaps by studying histones and histone-modifying enzymes from the green microalga *Chlamydomonas reinhardtii*. Our lab has uncovered many unique combinations of histone variants and post-translational modifications (PTMs) in *Chlamydomonas*\(^2,3\), most notably methylation of lysine 79 on histone H4 and a bimodal distribution of hypo- or hyperacetylated H3 depending on its lysine 4 methylation state\(^2\). Here, we report on the progress towards characterizing histone H3 and H4 methylation by combining gene synthesis and in vitro transcription/translation of histone methyltransferases with top-down mass spectrometry (TDMS) and Western Blot analysis. We further describe the development of a Python application that greatly facilitates quantitation of intact masses extracted from deconvoluted TDMS data.

The discovery and characterization of histone methyltransferases from the green microalga *Chlamydomonas reinhardtii* deepens our knowledge of algal epigenetics.


DIFFERENTIAL INTERROGATION OF PROTEOFORMS ACROSS THE DYNAMIC RANGE OF PLASMA PROTEOME BY NANOPARTICLES USING A MODIFIED PROTEOGRAPH WORKFLOW

Che-Fan Huang¹, Aniel Sanchez¹, Mrrittika Bhattacharya², Ambika Sundaresan², Michael A. Hollas¹, Michael A. Caldwell¹, Xiaoyan Zhao², Ryan Benz², Asim Siddiqui², Neil L. Kelleher¹

¹Proteomics Center of Excellence, Northwestern University, Evanston, IL;²Seer, Inc., Redwood City, CA.

Proteoforms in human plasma contain valuable biological information and potential biomarkers for clinical applications. However, analysis of the plasma proteome is challenging due to the overwhelming presence of high-abundance proteins that make up more than 90% of the total plasma proteome, including albumin, immunoglobulins, and other abundant proteins [1]. With limited dynamic ranges, depletion of these proteins or enrichment of less abundant proteins are often required prior to proteomics analysis with mass spectrometry [2]. To address this challenge, Seer Inc. developed the Proteograph™ workflow [3], which facilitates deep and broad plasma proteomic measurement at scale. The workflow includes contacting biofluids with engineered nanoparticles to form protein corona. Varying the physicochemical properties of the nanoparticles results in distinct protein corona, which can be analyzed by mass spectrometry. Here we present early results of top-down proteomic analysis from three nanoparticles that differentially interrogate human plasma using a Proteograph workflow modified to generate intact proteins. Pooled human plasma was incubated with three nanoparticles, washed and the proteins were eluted into Lysis buffer (from Seer Proteograph kit). The eluates were extracted for proteins < 50 kDa, identified and quantified for proteoforms using an established top-down LC/MS workflow in discovery mode [4]. With three biological replicates for each nanoparticle, we identified over 1000 proteoforms and demonstrated differential patterns of enrichment and depletion across these nanoparticles and compared to neat plasma. By mapping the identified proteins to the Human Plasma Proteome Project (HPPP) database, preliminary data showed identification of low abundant proteins at less than 10 pg/mL concentration as estimated by HPPP [2]. This work opens a door for deep quantitative analysis of proteoforms in plasma and could reveal biological differences that were previously hindered by abundant proteins present in plasma.

This work applies a novel top-down proteomics approach combined with engineered nanoparticles to achieve deep proteoform characterization in plasma.

COMPARING ALGORITHMS FOR PROTEOFORM IDENTIFICATION IN TOP-DOWN PROTEOMICS

David L. Tabb¹, Kyowon Jeong², Karen Druart³, Megan S. Gant¹, Kyle A. Brown⁴, Carrie Nicora⁵, Mowei Zhou⁶, the Milk Microbiome and Metabolome Team⁶,⁷, Michelle K. McGuire⁷, and Julia Chamot-Rooke⁸*

1. Institut Pasteur, Université Paris Cité, CNRS UAR 2024, Mass Spectrometry for Biology Unit, Paris 75015, France
2. Applied Bioinformatics, Computer Science Department, University of Tübingen, Tübingen, Germany, 72076
3. School of Medicine and Public Health, University of Wisconsin, Madison, Wisconsin, United States of America, 53705
4. Biological Sciences Division, Pacific Northwest National Laboratory, Richland, Washington, United States of America, 99352
5. Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland, Washington, United States of America, 99354
6. Department of Animal, Veterinary, and Food Sciences, University of Idaho, Moscow, Idaho, United States of America, 83844
7. Margaret Ritchie School of Family and Consumer Sciences, University of Idaho, Moscow, Idaho, United States of America, 83844

* julia.chamot-rooke@pasteur.fr

The analysis of complex samples by top-down (TD) mass spectrometry produces complex tandem spectra (MS/MS) which pose a challenge for data analysis. Recently, several TD pipelines have been developed which can be used to generate proteoform identifications from complex TD datasets. These algorithms use spectral matching and match-counting approaches to produce proteoform spectrum matches (PrSMs). For the benefit of the TD proteomics community, we assess state-of-the-art algorithms for top-down identification (ProSight PD, TopPIC, MSPathFinderT, and pTop) in their yield of PrSMs while controlling false discovery rate. We assess the number of PrSMs identified and the processing time required by each of the pipelines included in our comparison engines for E. coli (PX0192947) data acquired on ThermoFisher Orbitrap-class instruments. Finally, we searched for post-translational modifications (PTMs) in proteoforms from bovine milk (PXD031744). We demonstrate that TD identification pipelines produce excellent PrSM yields although only 19% of PrSMs were identified by all four algorithms, and that detection of PTMs is inconsistent between algorithms. By comparing different TD analysis pipelines, we can identify the strengths and weaknesses of different software and recommend the best TD workflows based on number of PrSMs, processing speed, and PTM identification and localisation.
A NEW SPE TIPS METHOD BASED ON AN INNOVATIVE SORBENT FOR FAST AND EFFICIENT PEPTIDE FRACTIONATION IN PROTEOMIC STUDIES

Kaynoush Naraghi1; Mana Shafaei2; Michel Arotçarena1; Florine Hallez1; Cerina Chhuon2; Chiara Guerrera2; Sami Bayoudh1

1 Affinisep, Le Houlme,
2 France Affinisep USA LLC, Miami, FL
3 INSERM – US24 SFR Necker Proteome, Paris, France

Peptide fractionation for full proteome characterization is very challenging, especially in the case of complex samples. The objective of this study was to develop a simplified procedure for the efficient and fast fractionation of peptides at basic pH, to contribute to further simplification of peptide separation and analysis.

A new reversed-phase sorbent, based on small sorbent particles tightly embedded in a monolithic membrane packed in SPE StageTips was used for the fractionation of peptides resulting from the enzymatic proteolysis of HEK293 cell lysate, and the results were compared to a reference commercial fractionation kit.

Eight fractions were obtained from each the commercial column and the SPE StageTips, with an acetonitrile gradient. Each fraction then was evaporated to dryness before being re-suspended in an appropriate solvent for nanoLC-MS/MS analysis.

The total number of proteins identified and the percentage of peptides eluting in only one fraction, were similar for both sorbents, with a good distribution of peptides over the eight fractions. Yet, it appeared that the fractionation on the new sorbent presented several advantages compared to the reference kit. For one it can be stored dry at room temperature while the commercial columns have to be stored at 4°C in a storage buffer. Moreover, due to the SPE StageTips format, the time required for the evaporation of each fraction is almost halved compared to the commercial columns.

Thus, the new sorbent appears as a promising solution for the fractionation of complex samples and the generation of spectral libraries, since it leads to an increase of more than 25% in the number of proteins identified, compared to unfractionated samples. Finally, this new sorbent offers flexibility of format and capacity, since it is also available as spin columns, for high amounts of peptides, or as 96 well plates, for high throughput experiments.

Keywords
Proteomics; peptide; fractionation; LC-MS/MS ; deep proteome