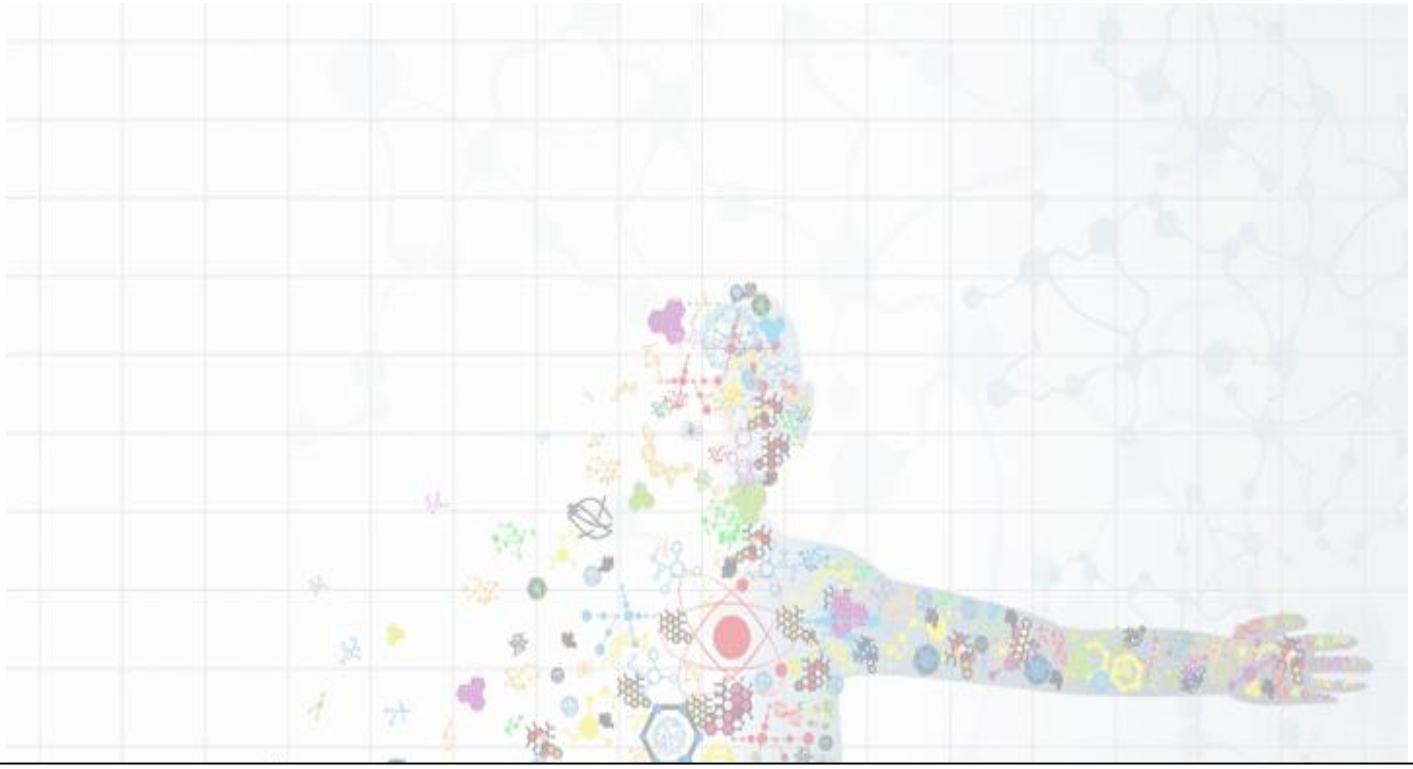
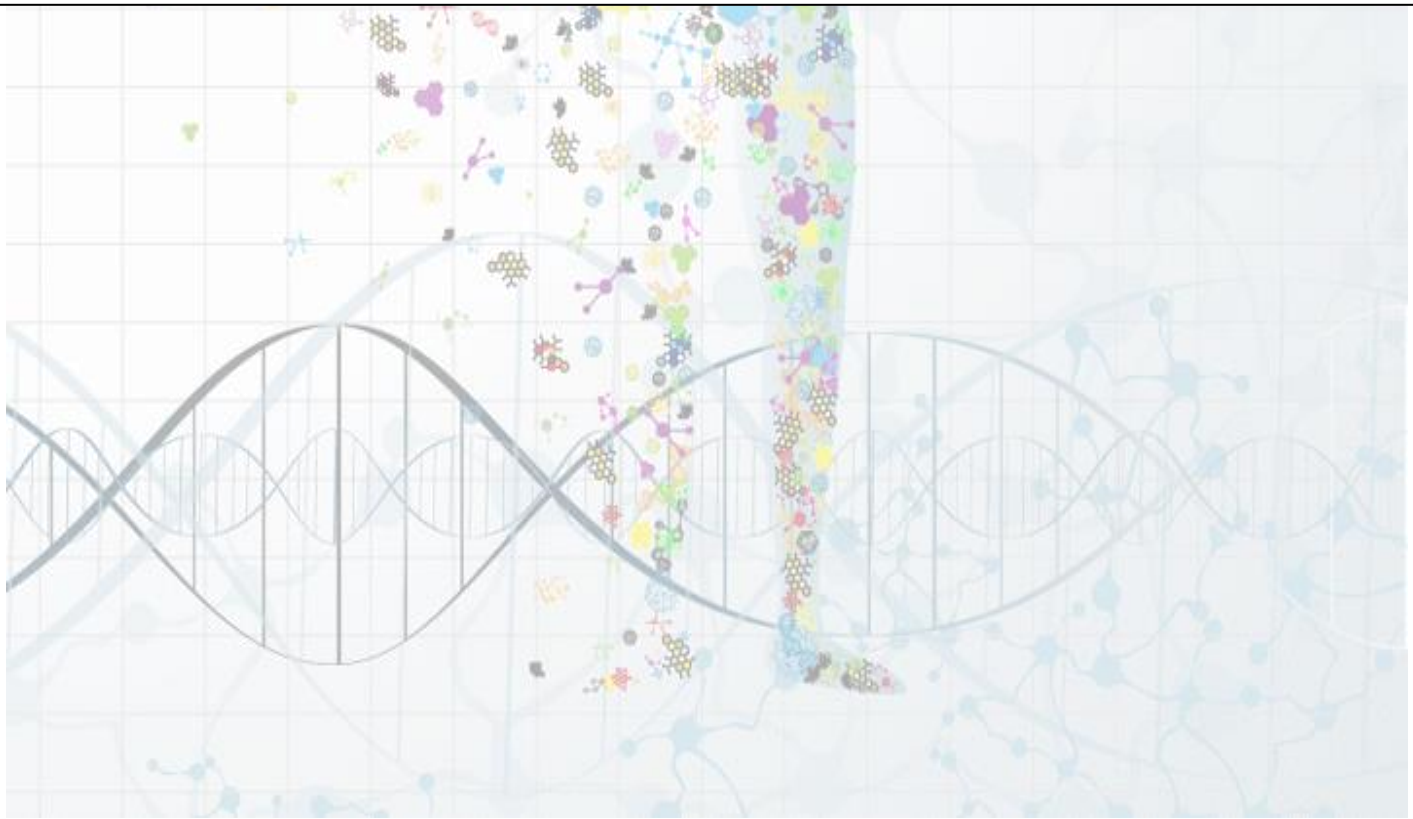


2023 INTERNATIONAL TOP-DOWN PROTEOMICS SYMPOSIUM SPEAKER ABSTRACTS



NEW FRONTIERS IN PROTEOMICS - PROTEOFORMS, PROTEOFORM FAMILIES, AND THE HUMAN PROTEOFORM PROJECT

Lloyd M. Smith[1]

[1]*Department 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[1].

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 todays 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 level[2,3].

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.

[1] L. M. Smith, et al, Nat. Meth., 10(3), 186-187 (2013).

[2] L. M. Smith, et al, Science 359(6380),1106-1107 (2018).

[3] L. M. Smith, et al, Sci. Adv. 7, eabk0734 (2021).

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.

[1] C. Wu et al, *ACS Nano* 16, 1025-35 (2022).

[2] X. Kang et al, N. Kelleher, *ACS Nano* 17, 5412-20 (2023).

A SYSTEMS BIOLOGY APPROACH TO DISCOVERY OF CLINICALLY ACTIONABLE PROTEIN ISOFORMS

Gloria M. Sheynkman[1,2,3,4]

[1]*Department of Molecular Physiology and Biological Physics, University of Virginia, Charlottesville, VA*
[2]*Department of Biochemistry and Molecular Genetics, University of Virginia, Charlottesville, VA*
[3]*Center for Public Health Genomics, University of Virginia, Charlottesville, VA*
[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.

[1] R. Miller, et al., *Genome Biology,* 23:69 (2022).

[2] M. Mehlferber, et al., *RNA Biology,* 19(1):1228-1243 (2022).

[3] A. Abood, et al., bioRxiv. Preprint (2023).

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
Richard D. LeDuc, Veit Schwämmle, Michael R. Shortreed, Anthony J. Cesnik, Stefan K. Solntsev, Jared B. Shaw‖, Maria J. Martin, Juan A. Vizcaino, Emanuele Alpi, Paul Danis, Neil L. Kelleher, Lloyd M. Smith, Ying Ge, Jeffrey N. Agar, Julia Chamot-Rooke, Joseph A. Loo, Ljiljana Pasa-Tolic, and Yury O. Tsybin, J. Proteome Res., 17, 1321–1325 (2018)

[2] Best practices and benchmarks for intact protein analysis for top-down mass spectrometry
Daniel P. Donnelly, Catherine M. Rawlins, Caroline J. DeHart, Luca Fornelli, Luis F. Schachner, Ziqing Lin, Jennifer L. Lippens, Krishna C. Aluri, Richa Sarin, Bifan Chen, Carter Lantz, Wonhyeuk Jung, Kendall R. Johnson, Antonius Koller, Jeremy J. Wolff, Iain D. G. Campuzano, Jared R. Auclair, Alexander R. Ivanov, Julian P. Whitelegge, Ljiljana Paša-Tolić, Julia Chamot-Rooke, Paul O. Danis, Lloyd M. Smith, Yury O. Tsybin, Joseph A. Loo, Ying Ge, Neil L. Kelleher & Jeffrey N. Agar, Nature Methods, 16, 587–594 (2019)

[3] A five-level classification system for proteoform identifications
Lloyd M. Smith, Paul M. Thomas, Michael R. Shortreed, Leah V. Schaffer, Ryan T. Fellers, Richard D. LeDuc, Trisha Tucholski, Ying Ge, Jeffrey N. Agar, Lissa C. Anderson, Julia Chamot-Rooke, Joseph Gault, Joseph A. Loo, Ljiljana Paša-Tolić, Carol V. Robinson, Hartmut Schlüter, Yury O. Tsybin, Marta Vilaseca, Juan Antonio Vizcaíno, Paul O. Danis & Neil L. Kelleher, Nature Methods, 16, 939–940 (2019)

[4] Interlaboratory Study for Characterizing Monoclonal Antibodies by Top-Down and Middle-Down Mass Spectrometry

Kristina Srzentić, Luca Fornelli, Yury O. Tsybin, Joseph A. Loo, Henrique Seckler, Jeffrey N. Agar, Lissa C. Anderson, Dina L. Bai, Alain Beck, Jennifer S. Brodbelt, Yuri E. M. van der Burgt, Julia Chamot-Rooke, Sneha Chatterjee, Yunqiu Chen, David J. Clarke, Paul O. Danis, Jolene K. Diedrich, Robert A. D'Ippolito, Mathieu Dupré, Natalia Gasilova, Ying Ge, Young Ah Goo, David R. Goodlett, Sylvester Greer, Kim F. Haselmann, Lidong He, Christopher L. Hendrickson, Joshua D. Hinkle, Matthew V. Holt, Sam Hughes, Donald F. Hunt, Neil L. Kelleher, Anton N. Kozhinov, Ziqing Lin, Christian Malosse, Alan G. Marshall, Laure Menin, Robert J. Millikin, Konstantin O. Nagornov, Simone Nicolardi, Ljiljana Paša-Tolić, Stuart Pengelley, Neil R. Quebbemann, Anja Resemann, Wendy Sandoval, Richa Sarin, Nicholas D. Schmitt, Jeffrey Shabanowitz, Jared B. Shaw, Michael R. Shortreed, Lloyd M. Smith, Frank Sobott, Detlev Suckau, Timothy Toby, Chad R. Weisbrod, Norelle C. Wildburger, John R. Yates III, Sung Hwan Yoon, Nicolas L. Young, and Mowei Zhou, J. Am. Soc. Mass Spectrom., 31, 1783–1802 (2020)

[5] Top-Down Mass Spectrometry of Native Proteoforms and their Complexes: A Community Study
Tanja Habeck, Kyle Brown, Benjamin Des Soye, Carter Lantz, Mowei Zhou, Novera Alam, Md Amin Hossain, Wonhyeuk Jung, James Keener, Michael Volny, Jesse W. Wilson, Yujia Ying, Jeffrey Agar, Paul Danis, Ying Ge, Neil Kelleher, Huilin Li, Joseph A. Loo, Michael Marty, Ljiljana Paša-Tolić, Wendy Sandoval, Frederik Lermyte, Nature Methods, submitted.

[6] The Human Proteoform Project: Defining the human proteome
Lloyd M. Smith, Jeffrey N. Agar, Julia Chamot-Rooke, Paul O. Danis, Ying Ge, Joseph A. Loo, Ljiljana Paša-Tolić, Yury O. Tsybin, Neil L. Kelleher, and The Consortium for Top-Down Proteomics, Science Advances, 7, eabk0734 (2021)

[7] *Proteoform: a single term describing protein complexity*
Smith, L., Kelleher, N. & The Consortium for Top-Down Proteomics., Nat Methods 10, 186–187 (2013)

5

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

Joseph A. Loo[1]

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

[1] A. Doerr, *Nature Methods*, 5, 24 (2008).

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

Dimitris Papanastasiou

[1]*Institution/Affiliation (Arial, Font size 10, italicized)*

Fasmatech 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 EID 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].

[1] Papanastasiou D. et al, *J. Am. Soc. Mass Spectrom.,* 33(10), 1990-2007 (2022). DOI: 10.1021/jasms.2c00214

[2] Wei J., et al, *Chem. Sci*. 14, 6695-6704, (2023). DOI:10.1039/d3sc00870c
[3] Smyrnakis A. et al, Accepted in *Anal. Chem.*

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 (ncIEF)-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 ncIEF-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.

[1] McCool EN, Xu T, Chen W, Beller NC, Nolan SM, Hummon AB, Liu X, Sun L. *Sci Adv.*, 8, eabq6348 (2022)
[2] Xu T, Wang Q, Wang Q, Sun L. *Anal Chem.*, 95, 9497-9504 (2023)
[3] Xu T, Han L, Sun L. *Anal Chem.*, 94, 9674-9682 (2022)

8

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

Francisco A. Fernandez-Lima1

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%.

9

TOWARDS LONG-READ SINGLE-MOLECULE PROTEIN SEQUENCING ON AN ARRAY OF UNFOLDASE-COUPLED NANOPORES

Keisuke Motone[1], Daphne Kontogiorgos-Heintz[1], **Jeff Nivala**[1,2]

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

[1] Cardozo et al. Multiplexed direct detection of barcoded protein reporters on a nanopore array. *Nat. Biotechnol.* 2022.

[2] Nivala et al. Unfoldase-mediated protein translocation through an α-hemolysin nanopore. *Nat. Biotechnol.* 2013.

MINIATURIZED QUANTITATIVE LOW-INPUT (NANO)-TOP-DOWN PROTEOMICS REVEALS INFORMATION ABOUT PROTEOFORM NOT ACCESIBLE 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

[1] Leipert J, Tholey A. Lab Chip, 19: 3490-3498, (2019).
[2] Kaulich PT, Cassidy L, Winkels K, Tholey A. Anal Chem, 94: 3600-3607, (2022).
[3] Leipert J, Kaulich PT, Steinbach MK, Steer B, Winkels K, Blurton C, Leippe M, Tholey A. *Angew Chem Int Ed*, e202301969, (2023).
[4] Steinbach MK, Leipert J, Blurton C, Leippe M, Tholey A. J Proteome Res, 21: 1986-1996, (2022).

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

Michael J. MacCoss[1], Javier Antonio Alfaro[2,3,4], Danielle A. Faivre[1], Christine C. Wu[1],

Meni Wanunu[5] & Nikolai Slavov[6,7]

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.

MacCoss et al, Nature Methods volume 20, pages 339–346 (2023)

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

Amy E. Herr[1,2]

[1]*Department of Bioengineering, University of California, Berkeley, Berkeley, CA;* [2]*Chan Zuckerberg Biohub San Franciso, 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.

13

TOWARD SINGLE-CELL PROTEOFORMS AND PROTEOFORM IMAGING

Ljiljana Paša-Tolić[1], Kevin J. Zemaitis[1], Dušan Veličković[1], Mowei Zhou[2], James M. Fulcher[1], Yen-Chen Liao[1], Ying Zhu[3], Sarah M. Williams[1], David J. Degnan[1], Lisa M Bramer[1], William R. Kew[1]

[1]*Pacific Northwest National Laboratory, Richland, WA, United States*
[2]*Department of Chemistry, Zhejiang University, Hangzhou, Zhejiang, China*
[3]*Genentech, 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.

[1] K.J. Zemaitis, D. Veličković, W. Kew, K.L. Fort, M. Reinhardt-Szyba, A. Pamreddy, Y. Ding, D. Kaushik, K. Sharma, A.A. Makarov, M. Zhou M, and L. Paša-Tolić L, *Anal Chem*. 94, 12604-12613 (2022).

[2] D.J. Degnan, K.J. Zemaitis, L.A. Lewis, L.A. McCue, L.M. Bramer, J.M. Fulcher, D. Veličković, L. Paša-Tolić, M. Zhou. *J Am Soc Mass Spectrom.* doi: 10.1021/jasms.3c00180. Online ahead of print. (2023).
[3] K.J. Zemaitis KJ, M. Zhou M, W. Kew W, and L. Paša-Tolić, *J Am Soc Mass Spectrom.*, 34, 328-332 (2023).

14

DIGITIZING PROTEOFORM BIOLOGY WITH SINGLE-MOLECULE MASS SPECTROMETRY

Neil L. Kelleher[1],[2]

[1]Proteomics Center of Excellence, Northwestern University, Evanston, IL; [2]Chemistry of Life Processes Institute, Northwestern University, Evanston, IL

Since the completion of the Human Genome Project, much has been made of the need to bridge the gap from genes and traits. As a key nexus for the many interacting '-omes' (genome, transcriptome, proteome, metabolome, etc.), the proteome should offer a tight link between genotype and phenotype. Proteoforms, or all of the precise molecular forms of a protein, capture all sources of variability in protein composition (i.e., SNPs, isoforms, post-translational modifications), and thus provide crucial insights into regulation and function. Now, "single ion" mass spectrometry is poised to convert genes to proteoform signatures at a far faster rate. Recently we developed proteoform imaging mass spectrometry (PiMS), with individual ion mass spectrometry. This platform has been extended now to single-cell Proteoform imaging Mass Spectrometry (scPiMS), boosting cell processing rates by >20-fold in the field while detecting proteoforms from single cells.

15

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.

[1] X. Yang, et al, *Cell*, 164, 805-817 (2016).

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), neutravidin (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.

[1] Vidal, M.; Cusick, M. E.; Barabási, A.-L. Interactome Networks and Human Disease. Cell 2011, 144 (6), 986–998.
[2] F. C. Liu, M.E. Ridgeway, M.A. Park, C. Bleiholder, Tandem trapped ion mobility spectrometry, Analyst 2018, 143, 2249-2258.
[3] F. C. Liu, M.E. Ridgeway, M.A. Park, C. Bleiholder, Tandem-trapped ion mobility spectrometry/mass spectrometry (tTIMS/MS): a promising analytical method for investigating heterogenous samples, Analyst 2022, 147, 2317-2337.
[4] F. C. Liu, T. C. Cropley, M. E. Ridgeway, M. A. Park, C. Bleiholder, Structural analysis of the glycoprotein complex avidin by tandem-trapped ion mobility spectrometry – mass spectrometry (tandem-TIMS/MS), Analytical Chemistry 2020, 92 (6), 4459-4467.

[5] F. C. Liu, M. E. Ridgeway, J. S. R. V. Winfred, N. C. Polfer, J. Lee, A. Theisen, C. A. Wootton, M. A. Park, C. Bleiholder, Tandem-trapped ion mobility spectrometry/mass spectrometry coupled with ultraviolet photodissociation, Rapid Communications in Mass Spectrometry 2021, 35 (22), e9192.

17

CHEMICAL PROTEOMIC STRATEGIES TO DISCOVER PROTEOFORM-SPECIFIC SMALL-MOLECULE PROBES

Benjamin F. Cravatt1

1Department of Chemistry, The Scripps Research Institute, La Jolla, CA 92037

Advances in DNA sequencing have radically accelerated our understanding of the genetic basis of human disease. However, many of human genes encode proteins that remain uncharacterized and lack selective small-molecule probes. The functional annotation of these proteins should enrich our knowledge of the biochemical pathways that support human physiology and disease, as well as lead to the discovery of new therapeutic targets. To address these problems, we have introduced chemical proteomic technologies that globally profile the functional state of proteins in native biological systems. Prominent among these methods is activity-based protein profiling (ABPP), which utilizes chemical probes to map the activity state and small-molecule interactions of large numbers of proteins in parallel. In this lecture, I will briefly explain ABPP and focus on the implementation of advanced ABPP platforms for proteome-wide ligand discovery, with a particular emphasis on the discovery of proteoform-specific small molecule-protein interactions.

DIRECT DETERMINATION OF MEMBRANE PROTEIN COMPLEXES FROM CELLULAR MEMBRANES THROUGH NATIVE TOP-DOWN MS

Wonhyeuk Jung[1,2]#; Aniruddha Panda[1,2]#; Caroline Brown[1,2]; Jeff Coleman[1,2];

James E. Rothman[1,2]; Kallol Gupta[1,2]

[1]*Nanobiology Institute, Yale University, West Campus ISTC, 850 West Campus Drive, West Haven, CT - 06516, U.S.A*

[2]*Department 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-bilayers[(1,2)]. 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

1. Panda A, Giska F, Duncan AL, Welch AJ, Brown C, McAllister R, Hariharan P, Goder JND, Coleman J, Ramakrishnan S, Pincet F, Guan L, Krishnakumar S, Rothman JE, Gupta K Direct determination of oligomeric organization of integral membrane proteins and lipids from intact customizable bilayer. Nature Methods. 2023 Jun;20(6):891–7.

2. Panda A, Brown C, Gupta K Studying Membrane Protein–Lipid Specificity through Direct Native Mass Spectrometric Analysis from Tunable Proteoliposomes J Am Soc Mass Spectrom. 2023 Jul 11. doi: 10.1021/jasms.3c00110. Online ahead of print.

TOP-DOWN PROTEOMICS FOR CARDIAC PRECISION MEDICINE AND CLINICAL DIAGNOSIS

Ying Ge, PhD[1]

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

(1)Melby, J.A.; Roberts, D.S.; Larson, E.J.; Brown, K.A.; Bayne, E.B.; Jin, S.; Ge, Y. Novel Strategies to Address the Challenges in Top-Down Proteomics J. Am. Soc. Mass Spectrom., 2021, 32, 1278-1294.

(2)Brown, K.A.; Melby, J.A.; Roberts, D.S.; Ge, Y Expert Rev Proteomics, 2020, 17, 719-733.

(3)Brown, K. A.; Chen, B.; Guardado-Alvarez, T.; Lin, Z.; Hwang, L.; Ayaz-Guner, S.; Jin, S.; Ge, Y. Nature Methods, 2019, 16, 417-420.

(4)Tiambeng T, Roberts DS, Brown KA, Zhu Y, Chen B, Wu Z, Mitchell SD, Guardado-Alvarez TM, Jin S, Ge Y. Nature Commun. 2020, 11, 3903.

(5)*Tucholski, T.; Cai, W.; Gregorich, Z.; Bayne, E.; Mitchell, S.; de Lange, W.; McIlwain, S.; Wrobbel, M.; Karp, H.; Hite, Z.; Vikhorev, P. G., Marston, S. B.; Lal, S.; Li, A.; dos Remedios, C.; Kohmoto, T.; Hermsen, J.; Kamp, T.; Ralphe J. C.; Moss, R.L.; Ge, Y.; Proc. Natl. Acad. Sci. U. S. A. 2020, 117, 24691-24700.*

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[1]. 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[2].
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[4] and equipped with a FTMS Booster) will also be presented[4].

Novel Aspect
Proteoforms as potential biomarkers for clinical applications

[1] L. M. Smith, et al, *Sci Adv 7* (46), eabk0734 (2021).
[2] M. Dupre, et al, *J Proteome Res 20* (1), 202-211 (2021).
[3] M. Dupre, et al. *Anal Chem 93* (30), 10627-10634 (2021).
[4] D. Papanastasiou, et al, *J. Am. Soc. Mass Spectrom. 33*(10), 1990–2007 (2022).

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. he 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-ternimal 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.

ACCELERATING DISEASE MARKER IDENTIFICATION AND DRUG DISCOVERY BY CLINICAL PROTEOFORMICS

Hartmut Schlüter[1]

[1] *Section of Mass Spectrometric Proteomics, University Medical Center Hamburg-Eppendorf, University of Hamburg, Hamburg, Germany*

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

[1] K. Jeong K, et al., *Cell Syst.* 10, 213-218 (2020)
[2] Z. Lin, et al., *J Am Soc Mass Spectrom.* 29, 1284-1294 (2018)
[3] M. Kostrzewa, *Expert Rev Proteomics*. 15,193-202 (2018)

23

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

Mariel Coradin[1] and **Benjamin A. Garcia**[2]

[1]*University of Pennsylvania School of Medicine, Philadelphia PA 19104*
[2]*Washington 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.