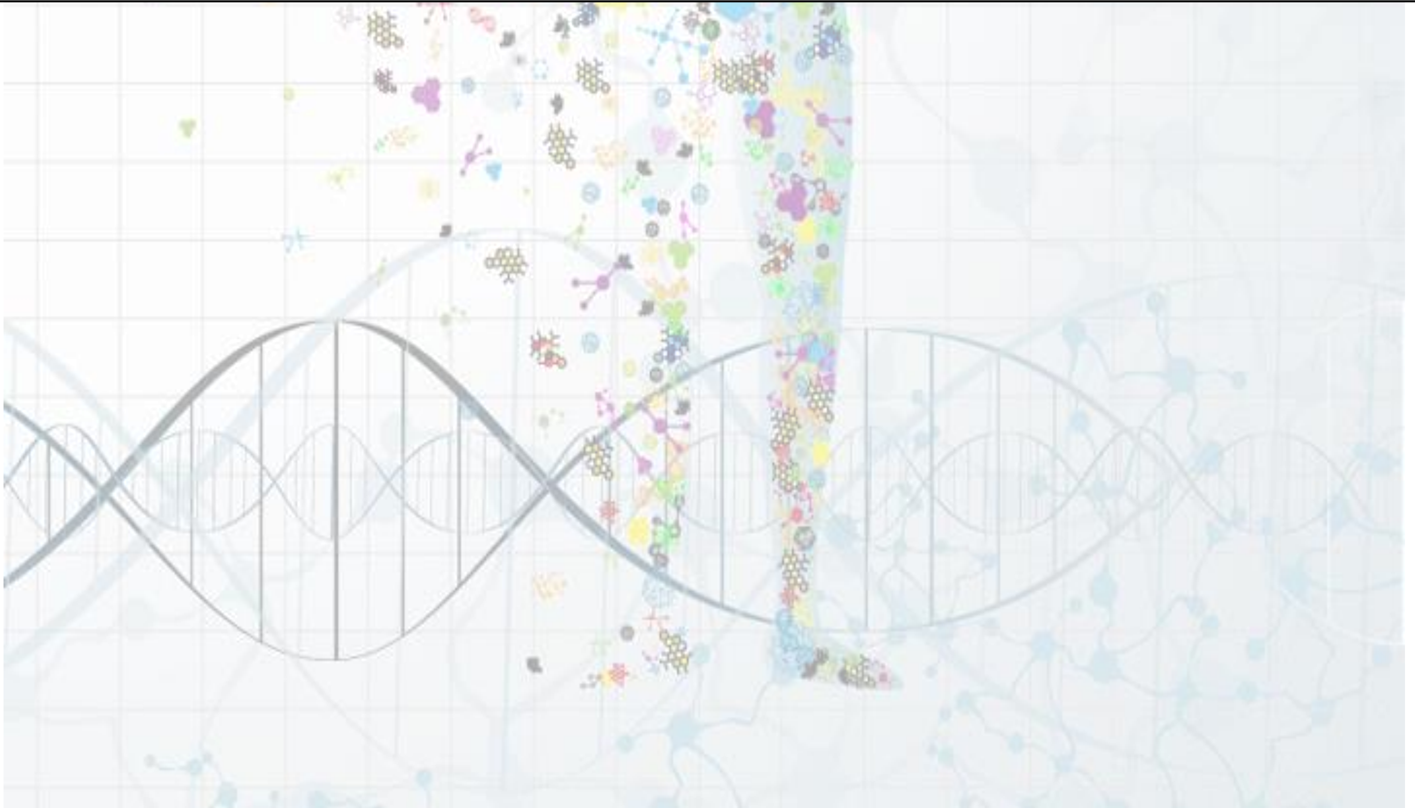
2023 INTERNATIONAL TOP-DOWN PROTEOMICS SYMPOSIUM

POSTER ABSTRACTS



CONNECTING PROTEOFORMS TO STRUCTURE WITH CLINIC

*[1]Melanie Cheung See Kit, [1]Ankita B. Gurav, [1]Ashlyn N. Dollar, [1]Erika Renbarger, [1,2]**Ian K Webb***
[1]Department of Chemistry and Chemical Biology, Indiana University – Purdue University Indianapolis;
[2]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 **C**hemical cross**L**inking **I**dentified under **N**ative or **I**ntact **C**onditions (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

[1]*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.

1 Milos Cejkov et al. (2021) J. Am. Soc. Mass Spectrom. 32(3): 762-771.

2 Luca Fornelli et al. (2018) Anal. Chem. 90(14): 8421-8429.

3 Richard D Smith et al. (1990) J. Am. Soc. Mass Spectrom. 1(1): 53-65.

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[1]

[1]*Quantum-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.

[1] B. Reed, et al, *Science*, 378, 186-192 (2022).

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

Kyowon Jeong[1,2], Ayesha Feroz[1], Konstantin Nagornov[3], Jihyung Kim[1,2], Timo Sachsenberg[1,2],Yury O.Tsybin[3], and Oliver Kohlbacher[1,2,4]

[1]*Applied Bioinformatics, Department for Computer Science, University of Tübingen, Sand 14, 72076 Tübingen, Germany*

[2]*Institute for Bioinformatics and Medical Informatics, University of Tübingen, Sand 14, 72076 Tübingen, Germany*

[3]*Spectroswiss, 1015 Lausanne, Switzerland*

[4]*Translational 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 decoy-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)

[2] Jeong et al., *Cell Systems*, 10, 213-218 (2020)

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

David Wissel[1], Madison Mehlferber[2], Marina Panteli[1], Mayank Murali[2], Saikat Bandyopadhyay[2], Roger Volden[3], Elizabeth Tseng[3], Gloria Sheynkman[2], Mark D. Robinson[1]

[1]*Department of Molecular Life Sciences, University of Zurich, Zurich CH, 8057*

[2]*Department of Molecular Physiology and Biological Physics, University of Virginia, Charlottesville VA, 22903*

[3]*Pacific 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.

[1] Miller, Rachel M., et al. "Enhanced protein isoform characterization through long-read proteogenomics." *Genome Biology* 23.1 (2022): 1-28.

[2] de Souza, Vladimir BC, et al. "Transformation of alignment files improves performance of variant callers for long-read RNA sequencing data." *Genome Biology* 24.1 (2023): 91.

[3] Poplin, Ryan, et al. "A universal SNP and small-indel variant caller using deep neural networks." *Nature Biotechnology* 36.10 (2018): 983-987.

[4] Martin, Marcel, et al. "WhatsHap: fast and accurate read-based phasing." *BioRxiv* (2016): 085050.

[5] Pardo-Palacios, Francisco, et al. "SQANTI3: curation of long-read transcriptomes for accurate identification of known and novel isoforms." *bioRxiv* (2023): 2023-05.

[6] Wenger, Craig D., and Joshua J. Coon. "A proteomics search algorithm specifically designed for high-resolution tandem mass spectra." *Journal of proteome research* 12.3 (2013): 1377-1386.

[7] Solntsev, Stefan K., et al. "Enhanced global post-translational modification discovery with MetaMorpheus." *Journal of proteome research* 17.5 (2018): 1844-1851.

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

Michael Poltash[1], Zoe Zhang[2], Haichuan Liu[2], Andy Mahan[1], Hirsh Nanda[1], Partha Chowdhury[1]

[1]Janssen Pharmaceuticals R&D, Spring House, PA, [2]SCIEX 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. Kaulich[1], Andreas Tholey[1]

[1]*Proteome 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

[1]Melani, R. D., Gerbasi, V. R., Anderson, L. C., Sikora, J. W., Toby, T. K., Hutton, J. E., Butcher, D. S., Negrão, F., Seckler, H. S., Srzentic, K., Fornelli, L., Camarillo, J. M., LeDuc, R. D., Cesnik, A. J., Lundberg, E., Greer, J. B., Fellers, R. T., Robey, M. T., DeHart, C. J., Forte, E., Hendrickson, C. L., Abbatiello, S. E., Thomas, P. M., Kokaji, A. I., Levitsky, J. and Kelleher, N. L. (2022). The Blood Proteoform Atlas: A reference map of proteoforms in human hematopoietic cells. *Science*. 375, 411–418.

[2]Kaulich, P. T., Cassidy, L., Winkels, K. and Tholey, A. (2022). Improved Identification of Proteoforms in Top-Down Proteomics Using FAIMS with Internal CV Stepping. *Anal. Chem.* 94, 3600–3607.

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

Clifton K. Fagerquist,[1] Yanlin Shi[1] and Jihyun Park[1,2]

[1]*Western Regional Research Center, Agricultural Research Service, USDA, Albany, CA, USA*

[2]*Research 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 Takemori[1], Ayako Takemori[1], Philipp T. Kaulich[2], and Andreas Tholey[2]

[1]*Advanced Research Support Center, Institute for Promotion of Science and Technology, Ehime University, Ehime, Japan*
[2]*Systematic 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 prefractionation 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-fractionation 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-fractionation 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.

[1] N.Takemori, et al, *Anal. Chem.*, 89, 8244-8250 (2017).
[2] A.Takemori, et al, *J. Proteome Res.*, 19, 3779-3791 (2020).
[3] A.Takemori, et al, *Anal. Chem.,* 94, 12815-12821 (2022).

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

Tanushree Dutta[1], Julea Vlassakis[1]
[1]*Department of Bioengineering, Rice University, Houston, Texas*

Nanoscale systems can remarkably improve protein recovery by downscaling of processing volumes and efficient chip-to-MS interfacing[1]. Nanofluidic and polymer sieves have been used in microfluidic devices for fast separation of DNA and proteins for fluorescence-based readouts[2-4] including at the single cell level[5]. For MS-based readout, the focus is more on minimizing (protein) losses[1, 6]. In particular, several top-down MS sample processing steps lead to substantial losses[7]. Additionally, sample prefractionation is useful for intact protein analysis[8] 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 protocol[9]. 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

[1] Y. Zhu, et al, *Nat. Commun.,* 9, 1-10 (2018).
[2] M. Azim, et al, *Electrophoresis* 38, 342-349 (2017).
[3] N. Shaabani, et al, *Electrophoresis* 37, 2602-2609 (2016).
[4] Y. Zeng, et al, *Analytical chemistry* 79, 2289-2295 (2007).
[5] J. Vlassakis, et al, *Nat. Commun.,* 12, 4969 (2021).
[6] Y. Zhu, et al, *Angew. Chem.,* 130, 12550-12554 (2018).
[7] M.V. Holt, et al, *J. Am. Soc. for Mass Spectrom.,* 30, 2548-2560 (2019).
[8] A. Takemori, et al, *J. Proteome Res.* 19, 3779-3791 (2020).
[9] P. Abdel-Sayed, et al, *Anal. Chem.* 89, 9643-9648 (2017).
[10] S. J. Hurst, et al, *Anal. Chem.* 78, 8313-8318 (2006).
[11] R.J. Macfarlane, et al, *Science* 334, 204-208 (2011).

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

[1] K. N. Nagornov, A. N. Kozhinov, N. Gasilova, L. Menin, Y. O. Tsybin, J. Am. Soc. Mass Spectrom., 31, 9, 1927–1942 (2020)
[2] K. N. Nagornov, A. N. Kozhinov, N. Gasilova, L. Menin, Y. O. Tsybin, J. Am. Soc. Mass Spectrom., 33, 7, 1113–1125 (2022)

ASSESSMENT OF LABELED PURIFIED PROTEINS WITH TOP-DOWN PROTEOMICS

Aron Phong[1], **John J.H. Shin[1]**

[1]*Alkahest – 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. Yeung[1,2], Yajing Liu[1], Run-Zhang Shi[1,2], Ruben Y. Luo[1,2]

[1]*Department of Pathology, Stanford University, Stanford, CA, USA*
[2]*Clinical 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.

[1] S.V. Rajkumar, *Am J Hematol* 97: 1086-107 (2022).
[2] N. van de Donk et al, *Lancet* 397: 410-27 (2021).
[3] C. Sarto et al, *Clin Biochem* 95: 28-33 (2021).
[4] A. Dispenzieri et al, *Leukemia* 34: 2749-53 (2020).

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

Rafael D Melani[1]; Christopher Mullen[1]; Linda B Lieu[2]; Jingjing Huang[1]; David Bergen[1]; Graeme C McAlister[1]; Amanda Lee[1], Romain Huguet[1]; Vlad Zabrouskov[1]; Luca Fornelli[2]

[1]Thermo Fisher Scientific, San Jose, CA. [2]University 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, EThcD, 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. MS^2 and MS^3 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 MS^2 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 Sausen[1],
Yury V. Vasil'ev, Rachel Franklin, Michael C. Hare, Adrian Guthals[2,3]
Joseph S. Beckman[3]

[1]*Agilent Technologies, Americas Field Organization, Wooddale IL, USA*
[2] *e-MSion, a part of Agilent., OR, USA*

[3]*Linus 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

INFORMATICS OF INDIVIDUAL ION MASS SPECTROMETRY ENABLES SINGLE CELL TOPDOWN 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. Kelleher[1]

[1]*Proteomics Center of Excellence, Chemistry of Life Processes Institute, Northwestern University, Evanston, IL*
[2]*Beckman 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 (I^2MS), an emerging technology for the detection of single ions, allowing for enhanced sensitivity (>500×) and 10× higher resolving power over traditional Orbitrap spectra.[2-4] 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 colocalized 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

1. Su, P. et al. Highly multiplexed, label-free proteoform imaging of tissues by individual ion mass spectrometry..Sci. Adv. 8, eabp9929 (2022)
2. Melani, R.D. et al. The Blood Proteoform Atlas: A reference map of proteoforms in human hematopoietic cells. Science 375, 411-418 (2022).
3. Kafader, J.O. et al. Multiplexed mass spectrometry of individual ions improves measurement of proteoforms and their complexes. Nature methods 17, 391-394 (2020).
4. McGee, J.P. et al. Isotopic Resolution of Protein Complexes up to 466 kDa Using Individual Ion Mass Spectrometry. Anal. Chem. 93, 2723-2727 (2021).

Wörner, T.P. et al. Resolving heterogeneous macromolecular assemblies by Orbitrap-based

COMPARISON OF SINGLE MOLECULE PROTEIN SEQENCING AND TOP-DOWN PROTEOMICS

Nickolas P. Fisher[1], Maanasa Narayanamoorthy[1], Jordan Summers[2], **Danielle Tullman-Ercek**[2,3], Neil L. Kelleher[1,3,4]

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

1) Reed, Brian D., Michael J. Meyer, Valentin Abramzon, Omer Ad, Omer Ad, Pat Adcock, Faisal R. Ahmad, et al. "Real-Time Dynamic Single-Molecule Protein Sequencing on an Integrated Semiconductor Device." *Science* 378, no. 6616 (October 14, 2022): 186–92. https://doi.org/10.1126/science.abo7651.

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

John G. Pavek[1], **Brian L. Frey[1]**, Dustin C. Frost[2], Ting-Jia Gu[2], Lingjun Li[2], Lloyd M. Smith[1]

[1]Department of Chemistry, University of Wisconsin-Madison; [2]School 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.

[1] J.G. Pavek, B.L. Frey, D.C. Frost, T.J. Gu, L. Li, L.M. Smith; submitted to *Analytical Chemistry*.

PTM-FOCUSED TOP-DOWN PROTEOFORM ANALYSIS

Jacob Seay[1], Gustavo Agreda[1], **Bryon Drown**[1]

[1]*Department 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 Rogers[1], David Roberts[1], Eli Larson[1], Jake Melby[1], Kalina Rossler[2], Austin Carr[1], Kyle Brown[1,3] and Ying Ge[1,4,5]

Department of Chemistry[1]; Department of Molecular and Cellular Pathology[2]; Department of Surgery[3]; Department of Cell and Regenerative Biology[4]; Human Proteomics Program[5], 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:

(1) Whitelegge, J. P. Integral Membrane Proteins and Bilayer Proteomics. *Anal. Chem.* **2013**, *85* (5), 2558–2568. https://doi.org/10.1021/ac303064a.

(2) MacLennan, D. H.; Kranias, E. G. Phospholamban: A Crucial Regulator of Cardiac Contractility. *Nat. Rev. Mol. Cell Biol.* **2003**, *4* (7), 566–577. https://doi.org/10.1038/nrm1151.

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 MS^3 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.

[1] F. C. Liu, M.E. Ridgeway, M.A. Park, C. Bleiholder, Tandem trapped ion mobility spectrometry, Analyst 2018, 143, 2249-2258.

[2] 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.

[3] F. C. Liu, S. R. Kirk, K. A. Caldwell, T. Pedrete, F. Meier, C. Bleiholder, Tandem Trapped Ion Mobility Spectrometry/Mass Spectrometry (tTIMS/MS) Reveals Sequence-Specific Determinants of Top-Down Protein Fragment Ion Cross Sections, Analytical Chemistry 2022, 94, 8146–8155.

CHARACTERIZING HISTONE AND HISTONE-MODIFYING PROTEINS FROM THE GREEN MICROALGA *CHLAMYDOMONAS REINARDTHII* 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 mases extracted from deconvoluted TDMS data.

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

(1) Strahl, B. D.; Allis, C. D. The Language of Covalent Histone Modifications. *Nature* **2000**, *403* (6765), 41–45. https://doi.org/10.1038/47412.

(2) Rommelfanger, S. R.; Zhou, M.; Shaghasi, H.; Tzeng, S.; Evans, B. S.; Umen, J. G.; Pesavento, J. J. An Improved Top-Down Mass Spectrometry Characterization of Chlamydomonas Reinhardtii Histones and Their Post-Translational Modifications. **2021**. https://doi.org/10.1021/jasms.1c00029.

(3) Khan, A.; Eikani, C. K.; Khan, H.; Iavarone, A. T.; Pesavento, J. J. Characterization of Chlamydomonas Reinhardtii Core Histones by Top-Down Mass Spectrometry Reveals Unique Algae-Specific Variants and Post-Translational Modifications. *Journal of Proteome Research* **2018**, *17* (1), 23–32. https://doi.org/10.1021/acs.jproteome.7b00780.

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

Che-Fan Huang[1], Aniel Sanchez[1], Mrittika Bhattacharya[2], Ambika Sundaresan[2], Michael A. Hollas[1], Michael A. Caldwell[1], Xiaoyan Zhao[2], Ryan Benz[2], Asim Siddiqui[2], Neil L. Kelleher[1]

[1]*Proteomics Center of Excellence, Northwestern University, Evanston, IL;* [2]*Seer, 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.

[1] R. S. Tirumalai, et al, Mol. Cell. Proteom., 2, 1096-1103 (2003).

[2] E. W. Deutsch, et al, *J. Proteome Res.,* 20 (12), 5241–5263 (2021).

[3] J. E. Blume, et al, Nat. Commun., 11: 3662 (2020).

[4] R. D. Melani, et al, *Science,* 375 (6579), 411-418 (2022).

COMPARING ALGORITHMS FOR PROTEOFORM IDENTIFICATION IN TOP-DOWN PROTEOMICS

David L. Tabb[1], Kyowon Jeong[2], Karen Druart[1], Megan S. Gant[1], Kyle A. Brown[3], Carrie Nicora[4], Mowei Zhou[5], the Milk Microbiome and Metabolome Team[4,6,7], Michelle K. McGuire[7], and Julia Chamot-Rooke[1*]

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 Shafaei**2; Michel Arotçarena1; Florine Hallez1; Cerina Chhuon2; Chiara Guerrera2; Sami Bayoudh1

1 Affinisep, Le Houlme,

2France Affinisep USA LLC, Miami, FL

3INSERM – 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.

Key words

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

PROTEOFORM ATLASING: PAST, PRESENT, AND FUTURE

Ryan T. Fellers[1], Hartmut Schlüter[2], Members of Working Group 2 of HPfAtlas Initiative[3]

[1]Proteomics Center of Excellence, Northwestern University, Evanston, IL; [2]Section/Core Facility Mass Spectrometric Proteomics, University Medical Center Hamburg-Eppendorf (UKE), Martinistraße 52, 20246 Hamburg, Germany; [3]Consortium for Top-Down Proteomics

Since the advent of mass spectrometry of intact proteins and top-down proteomics, practitioners have looked for ways to publish lists of identified proteoforms for community use. The current state of the art consists of an "Alpha" proteoform atlas hosted by the CTDP and two "Beta" human proteoforms atlases [1,2] hosted by Northwestern University. While the existing solutions have had success (e.g., integration into UniProt), we feel now is the time to consider the next generation of proteoform atlases. Future incarnations should continue to involve the community to refine requirements and improve utility. Specifically, it is crucial to place a high importance of the functions and biological processes of specific proteoforms. This will necessitate the creation and improvement of atlases that are more targeted and functional in nature [3], with specific user interface elements to make such interactions clear. It is also vital to increase the links to existing resources: tighter coupling to UniProt to provide a proteoform centric experience, creation of datasets in the ProteomeXchange consortium [4] (e.g., PRIDE, MassIVE) to support spectral evidence if available, and deep links to specific gene within top-tier protein atlases (e.g., Human Protein Atlas [5]). With the increasing reach of top-down proteomics, now is the time to create atlases to serve not only a discovery based, gene-centric need, but also the disease- and organ-centric, biological view of proteomics. Ultimately, this should result in a web-based system that can handle large datasets and targeted, functional and biological process-focused descriptions coherently and serve them to an ever-expanding community.

This work discusses the current state of proteoform atlases and suggests additional annotations of functions and biological processes of proteoforms.

[1] Hollas, Michael A R et al. "The Human Proteoform Atlas: a FAIR community resource for experimentally derived proteoforms." Nucleic acids research vol. 50,D1 (2022): D526-D533. doi:10.1093/nar/gkab1086

[2] Melani, Rafael D et al. "The Blood Proteoform Atlas: A reference map of proteoforms in human hematopoietic cells." Science (New York, N.Y.) vol. 375,6579 (2022): 411-418. doi:10.1126/science.aaz5284

[3] Proteinopathy, in preparation.

[4] Deutsch, Eric W et al. "The ProteomeXchange consortium at 10 years: 2023 update." Nucleic acids research vol. 51,D1 (2023): D1539-D1548. doi:10.1093/nar/gkac1040

[5] Uhlén, Mathias et al. "Proteomics. Tissue-based map of the human proteome." Science (New York, N.Y.) vol. 347,6220 (2015): 1260419. doi:10.1126/science.1260419

AUTOMATED WORKFLOW FOR HIGHLY-SENSITIVE FRAGMENTATION IN ONCOPROTEOMICS USING THE SAMPLESTREAM PLATFORM AND INDIVIDUAL ION MASS SPECTROMETRY

Raveena Gupta[1], Bryon Drown[1], John P. McGee[1], Jared O. Kafader[1], Michael A. Caldwell[1], Phil Compton[1], Neil Kelleher1

[1]Department of Chemistry, Northwestern University, Evanston, IL

Dysregulation of signaling pathways is a common mechanism that contributes to oncogenesis, metastasis, and drug resistance. Disentangling the combinatorial post-translational modification (PTM) patterns that encode these biological signals is challenging, as peptide-based mass spectrometry strategies ("bottom-up" MS) degrade information on PTM stoichiometry. However, robust top-down mass spectrometry workflows incorporating automation promise to increase access to proteoform-level measurements including precise characterization of PTMs and their stoichiometries, opening a new window into cancer proteomics. Automated sample cleanup using SampleStream removes interfering substances and enhances sensitivity and reproducibility while reducing manual sample handling. Coupling SampleSteam with direct injection of samples for individual ion mass spectrometry (I2MS) analysis further advances the goals of automation and reproducibility and aids in localizing modifications by providing much deeper sequence coverage than traditional tandem-MS. Coupling antibody-based enrichment and SampleStream cleanup with I2MS top-down analysis provides a powerful new workflow for the determination of oncoproteoforms.

MEK1 obtained by immunoprecipitation from A375 metastatic melanoma cancer cells was introduced to the SampleStream system where a size-selective membrane concentrates retained protein while removing interfering matrix components. Concentrated and cleaned samples were then directly introduced to a Q Exactive-HF mass spectrometer. The baseline proteoform landscape was observed and quantified using MS1 ensemble measurements. In the case of MEK1, phosphorylations between treatments and mutant cell lines were observed and and phospho-site localization and occupancy were determined by I2MS2.

The overall degree of MEK1 phosphorylation decreased upon MAPK pathway inhibition with vemurafenib (upstream BRAF mutation inhibitor) and trametinib (MEK inhibitor) treatments. Furthermore, we compared the phosphorylations of MEK1 isolated from isogenic A375 cell lines that bear activating mutations in MEK1, KRAS, and NRAS conferring resistance to BRAF inhibition. While KRAS and NRAS mutations did not affect the basal level of MEK1 phosphorylation, they prevented the durable reduction of MEK1 phosphorylation by BRAF inhibition. This integrated approach should find wide application in evaluating drug responses in signaling pathways from models and patients.