ENHANCED ION ACTIVATION AND DISSOCIATION OF INTACT PROTEINS IN NATIVE MASS SPECTROMETRY BASED TOP-DOWN PROTEOMICS

Michael F. Keating
April 17th, 2018

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Bachelor of Science
(Honors Chemistry)
in the University of Michigan
2018

This thesis has been read and approved by Brandon T. Ruotolo

Signed: _____ Date: 4/16/18

Faculty advisor e-mail: bruotolo@umich.edu Phone: 734-615-0198

ENHANCED ION ACTIVATION AND DISSOCIATION OF INTACT PROTEINS IN NATIVE MASS SPECTROMETRY BASED TOP-DOWN PROTEOMICS

Michael F. Keating
April 17th, 2018

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Bachelor of Science
(Honors Chemistry)
in the University of Michigan
2018

This thesis has been read and approved by Brandon T. Ruotolo

Signed: _____ Date: ____/____/____

Faculty advisor e-mail: _____ Phone: _____

Table of Contents

Acknowledgements

I've been fortunate enough to have been influenced by a great deal of wonderful people during my time here at the University of Michigan including friends, graduate student mentors, and research advisors. Thank you to Professor Nathaniel Szymczak, for giving me my first opportunity to do research. Thank you to my first graduate student mentor, Lily Hale, for leading by example and pushing me to develop the skills required to do good science. While I'm not pursuing the research that we conducted together, it's safe to say I wouldn't be where I am today without all of the exceptional mentoring you gave me. I am grateful to Professor Brandon Ruotolo for allowing me to participate in cutting edge research doing top-down proteomics during the twilight of my undergraduate education. I'd like to acknowledge Dan Polasky for not only putting up with my constant stream of questions, but also for somehow having an answer to each and every one of them. He also found the time to give me all of the training required to make this thesis possible. To all members of the Ruotolo lab, thank you for making me feel welcome and creating a fun and productive work environment. A special thank you to all of the friends I've made at the University of Michigan, especially my friends in the chemistry department. For all of the times I found the coursework particularly challenging, I could count on you to keep me on track and on the path to success. To Maddie McCabe, Anushka Gupta, and Guy Burke, thank you for being such awesome friends. Some of my best memories have been made studying with you, being in class with you, and eating bagels with you. I'm sad that we will all be in different parts of the country soon, but I'm excited to watch you all succeed. It is absolutely essential that we continue to meet up and enjoy bagels together for the rest of our lives.

I am also grateful to Purdue University and Professor R. Graham Cooks for allowing me to join for the summer as an REU student. I can't think of a better place to learn about mass spectrometry and its applications. The ability to work with ambient ionization and mass spectrometry based imaging technologies has had a tremendous impact on me, so much that I plan to pursue related research in my graduate career.

None of the acknowledgements above would exist without the love and support provided by my family, obviously also including Alexander Dils. My parents have always emphasized the important of taking education seriously, something I have adopted myself. I cannot thank you enough for supporting me in all of my pursuits. You've worked incredibly hard to give me every possible opportunity. Thank you to my brothers James, Alex, and Joseph, for laughing at all of my jokes. More seriously, James' journey through science has had a considerable influence on my own personal journey. I'm looking forward to being your future peer in analytical chemistry. Alex's contagious work ethic has always motivated me to work harder and be a better version of myself. Joseph has always had my back. Thank you for keeping me grounded and always being so supportive.

To all those not explicitly mentioned, thank you for all of your support.

Chapter 1

Introduction

Part 1. Protein Structure – The Proteoform

Proteins are critical components of the cellular machinery; acting as molecular machines that perform vital functions necessary to sustain life e.g. cell maintenance, replication, destruction, etc. Protein structure in structural biology can be divided into four levels of increasing complexity: primary, secondary, tertiary, and quaternary structures (Figure 1).[18] Primary structure is the sequence of amino acids that make up its polypeptide chain. Secondary protein structure are local folds of amino acids e.g. alpha helices, beta sheets, beta barrels, etc. Tertiary structure refers to how the protein folds and organizes in three dimensional space. The highest degree of protein structure, quaternary structure, references the number and arrangement of multiple folded proteins that make up a multi-protein complex.[1]

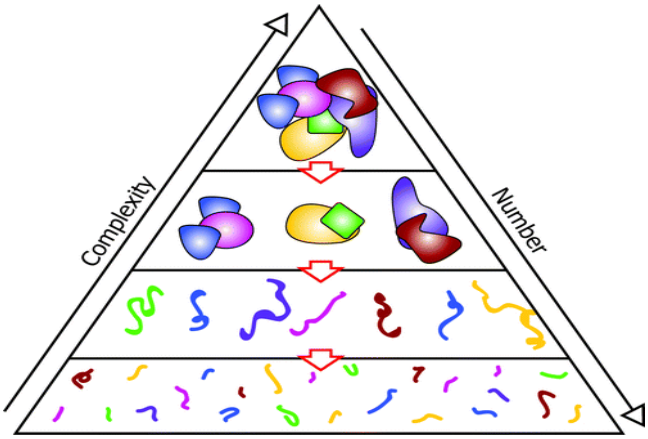

Figure 1. The four degrees of protein structure: primary, secondary, tertiary, and quaternary.[18]

To further increase the complexity, proteins can undergo post-translational modification (PTM), chemical changes that occur on individual amino acids after the sequence has been translated from its genetic instructions. PTMs range from the cleavage of peptide bonds to the covalent attachment of entirely new chemical moieties e.g. phosphorylations and glycosylations.[2] The term "proteoform" refers to all possible forms in which the protein product of a single gene can be found (Figure 2).[3] This includes all forms of genetic variation and all possible post-translational modifications. The complexity of protein structure and the vast number of possible

proteoforms highlight the need for analytical tools that can provide insight into all four degrees of protein structure.

Some currently available and emerging techniques used in protein analysis are X-ray crystallography, Nuclear magnetic resonance (NMR) spectroscopy, and mass spectrometry (MS). X-ray crystallography and NMR spectroscopy can provide high resolution structural data on

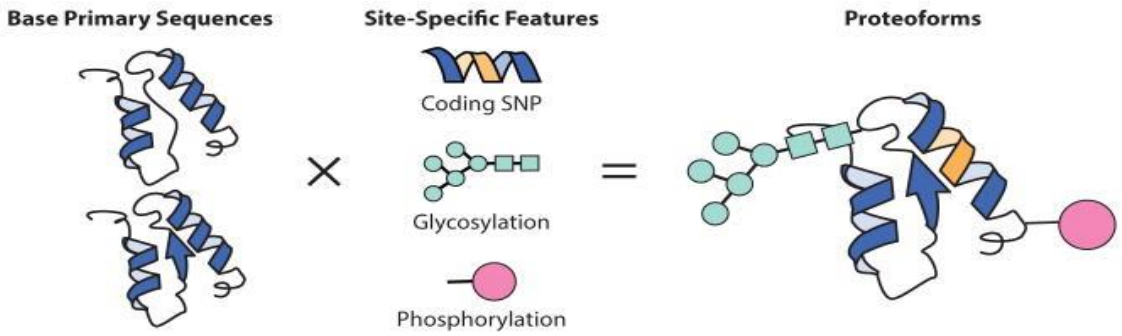

Figure 2. Post-translational modifications and other site-specific features contribute to individual proteoforms.[3]

proteins and protein complexes, but it is highly challenging to employ these technologies in the study of large heterogeneous protein complexes. Mass spectrometry based instrumentation is not restricted by complex size and has been used to study protein complexes of up to several megadaltons.[9,20] The use of X-ray crystallography is also entirely dependent on the ability to crystallize the protein/protein complex of interest. NMR spectroscopy requires pretreating the sample with expensive isotopic labels. Mass spectrometry does not share these limitations.[16] Mass spectrometry of intact proteins and protein complexes has the potential to provide information about every level of protein structure ranging from sequences and post-translational modifications to the structure of whole protein assemblies.[6] This, coupled with a reduced sample

requirement compared to both XRD and NMR (picomoles versus millimoles), make the optimization of mass spectrometry based technologies for protein analysis highly desirable.

Part 2. Mass Spectrometry Based Top-Down and Bottom-Up Proteomics

Proteomics is defined as the comprehensive analysis of cellular proteins. There are two general approaches used in proteomics: top-down and bottom-up.[4] Bottom-up proteomics employs an enzyme, typically trypsin, to digest proteins into peptides prior to analysis. Top-down proteomics analyzes intact proteins without first using an enzyme to digest the cellular proteoforms into peptides.[8] While the peptides generated in bottom-up approaches are easily solubilized and chromatographically separated before MS analysis, a large degree of information is forfeited upon enzymatic digestion of the proteoforms present.[5] More specifically, digestion eliminates connectivity between any PTMs identified and their incorporation into specific proteoforms. The digestion process can also obfuscate the location and identity of post-translational modifications, in that all of the sequence is not always recovered during the analysis. The top-down approach conserves the connectivity between PTMs and proteoforms, as well as often providing improved information on the identify and location of PTMs.[6,7]

Part 3. Top-Down Proteomics using Native Mass Spectrometry

Most proteins form non-covalent multiprotein assemblies to carry out their biological function.[16] As a result, complete characterization of these complexes requires maintaining non-covalent interactions that influence higher degrees of protein structure. Preserving the native state is possible by keeping the protein in aqueous solution and manipulating ionic strength and pH to mimic physiological conditions.[10] Recently, a number of research groups have attempted

merging top-down sequencing efforts with native MS. As above, such an experiment has the key advantage of connecting protein complex proteoforms with the identification of individual PTMs. A typical top-down proteomics experiment coupled with native mass spectrometry can be effectively described in three key steps: ion generation, ion activation/fragmentation, and mass analysis.

Part 4. A Typical Native Mass Spectrometry Experiment

For simplicity, a typical experiment will be described in the specific context of the mass spectrometry instrumentation used to collect all data described in Chapter 3 of this dissertation. All data were collected on a quadrupole ion mobility time-of-flight mass spectrometer (Synapt G2 HDMS, Waters, Milford, Ma). The instrument is equipped with a nano-electrospray ionization (nESI) source, a quadrupole mass analyzer, a T-wave ion mobility separator (IM), and a time-of-flight (ToF) mass analyzer arranged in tandem (Figure 3). The first step in the experiment is to generate ions from a sample of interest. nESI is well suited for native mass

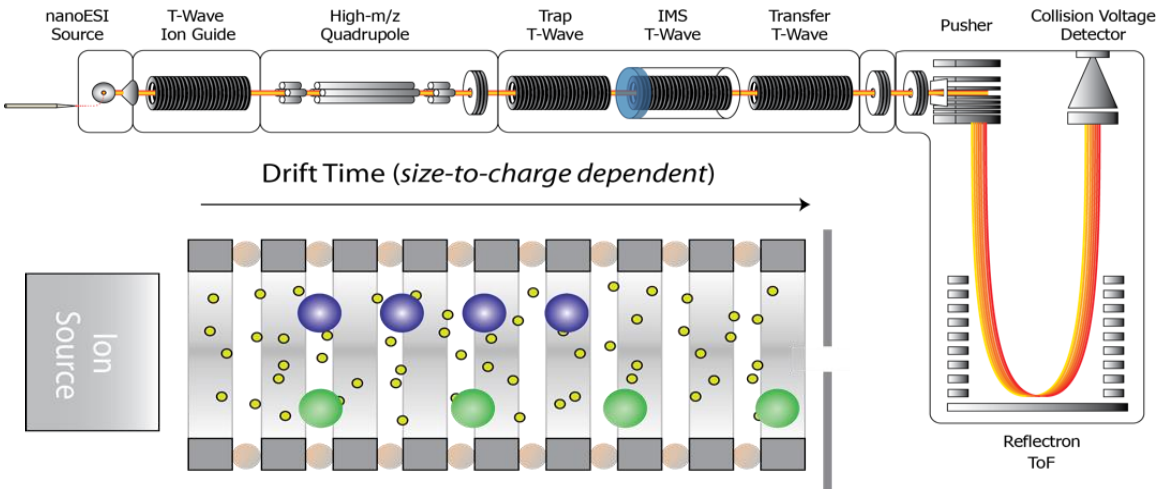

Figure 3. A Schematic diagram of the Synapt G2, quadrupole-ion mobility-time-of-flight mass spectrometry instrument used in this work.

spectrometry because it is a soft ionization technique capable of preserving non-covalent interactions in protein assemblies.

Gas-phase protein ions are generated in the nESI source region by applying high electrostatic voltage to the tip of a capillary charged with a protein containing solution. The high electric field causes exposed liquid at the tip of the capillary to form a Taylor cone, which emits droplets containing charged biomolecular species. Emitted droplets encounter neutral gas molecules as they enter the instrument which removes excess solvent. As solvent is removed the charges in each droplet experience increasing Coulombic repulsion. When the Coulombic repulsion overcomes the surface tension holding the droplet together, droplet fission occurs. This process repeats itself until virtually all solvent has been removed, leaving behind multiply charged protein ions.[21]

Multiply charged protein ions are then transmitted through the "T-Wave Ion Guide" and into the "High m/z quadrupole". The quadrupole can be operated as an additional ion guide, or be used to isolate a particular m/z ion packet for subsequent mass analysis. In top-down proteomics analysis, the quadrupole is operated so that only a single charge state will be transmitted further into the instrument.[5] Ions permitted to travel through the quadrupole enter the "Tri-wave region." This region is where two key processes take place: collision induced dissociation (CID) and ion mobility (IM) separation. In CID, ions that have been selected in the quadrupole mass analyzer are accelerated into a population of neutral gas molecules. Collisions with neutral gases convert the ion's kinetic energy into internal energy until the ion undergoes fragmentation.[22] CID is discussed in greater detail in Chapter 2. Regardless of whether CID was employed to induce

fragmentation, quadrupole selected ions enter the mobility cell of the "Tri-wave region" where they are IM separated based on their ability to travel through a chamber filled with a population of neutral gas molecules. Larger proteins experience more collisions with the neutral gas molecules, taking longer to traverse the mobility cell.[20] Ions that have been separated in the mobility cell are transferred to the pusher. The pusher pulses periodically to inject a portion of the ion beam into the flight tube of the time-of-flight (ToF) mass analyzer. The flight tube is a field-free drift tube operated under vacuum. Ions are separated in the flight tube based on their m/z-dependent flight times to the detector.[36]

Part 5. Challenges in Top-Down Proteomics using Native Mass Spectrometry

Mass spectrometry of intact proteins and protein complexes has the potential to provide a transformative level of information on biological systems, ranging from sequence and post-translational modification analysis to the structures of whole protein assemblies.[12-20] This ambitious goal requires the efficient fragmentation of both intact proteins and the macromolecular, multicomponent machines they collaborate to create through non-covalent interactions. Current activation technologies struggle to provide sufficient activation to efficiently fragment these assemblies. This lack of fragmentation limits the amount of the protein sequence that can be identified. State of the art mass spectrometry compatible activation technologies have only reported meager sequence coverage of large protein assemblies. In this thesis, we aim to improve the sequence coverage of intact protein complexes by using chemical derivatization technology coupled with CID to enhance top-down fragmentation. Improving technologies to achieve the fragmentation required to comprehensively analyze cellular protein composition remains perhaps the greatest challenge facing current efforts in top-down type

analyses and is essential to realize the full potential of proteomics.[9-11] Currently available activation technologies are discussed in Chapter 2.

Chapter 2

Ion Activation and Fragmentation Technologies Compatible with Native Mass Spectrometry

Part 1. Post-Ionization Activation Technologies

The most common method of ion activation for use with MS is collision induced dissociation (introduced in Chapter 1). Like the name suggests, CID induces fragmentation of charged species by accelerating them into a population of neutral gas molecules. Each collision converts some portion of the ion's kinetic energy into internal energy. This rise in internal energy results in ion fragmentation. Applying CID in native MS to activate intact proteins and protein complexes is challenging because it struggles to generate enough fragmentation to characterize them effectively.[22] The lack of CID mediated fragmentation in protein complexes can be attributed to their ability to participate in multiple dissociation pathways (Figure 4).[38] Possible dissociation pathways of macromolecular assemblies include unfolding followed by dissociation, dissociation without unfolding, or peptide fragment loss without dissociation. Protein assemblies that participate in the first two pathways are especially difficult to sequence because they utilize the energy gained in the CID process to unfold/dissociate a protein from the complex before fragmenting into peptides that can be used for sequence analysis.

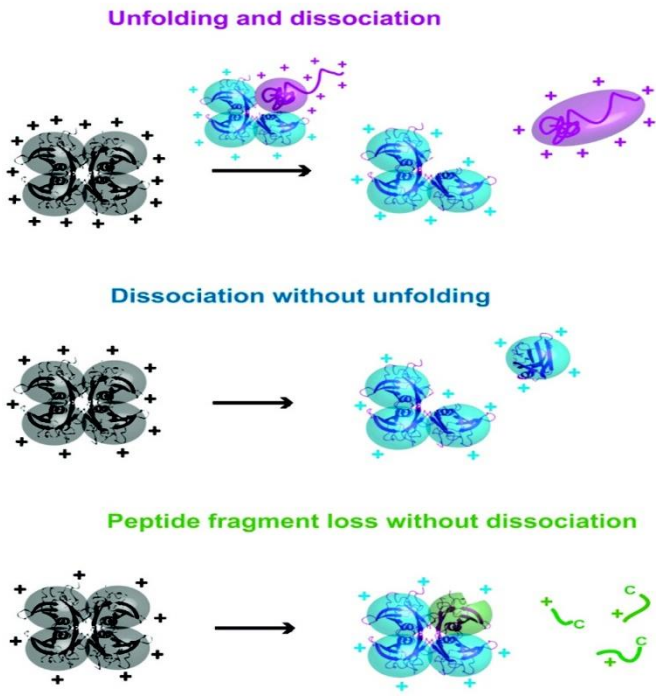

Figure 4. Dissociation pathways available to CID activated protein assemblies.[38]

This drawback has led to the development of other post-ionization activation technologies including ultraviolet photodissociation (UVPD)[23-25], electron transfer and capture dissociation (ETD[26,27] and ECD[28-30]), and infrared multiphoton dissociation (IRMPD).[31]

UVPD operates by irradiating trapped protein ions with energetic photons, typically 193 nm. These high energy photons are absorbed by the amide backbone which results in dissociation. UVPD has demonstrated conservation of labile PTMs and the ability to localize them in the proteoform. This method of activation has produced net sequence coverages of approximately 50% in ~20 kDa proteins and reduced net sequence coverages in larger proteins.[23-25, 39]

ECD induces fragmentation by introducing low energy electrons to trapped gas-phase protein ions. ECD has demonstrated the ability to conserve labile PTMs. This activation technique was originally developed exclusively for use with Fourier-Transform Ion Cyclotron Resonance (FT-ICR) instrumentation. The restriction of this activation technique to expensive FT-ICR mass spectrometers catalyzed the development of ETD. ETD induces fragmentation by using charged electron donating reagents to energize ions in an exothermic electron attachment process. Unlike ECD, ETD is not limited to a specific subset of mass spectrometry instrumentation. Like ECD, this technology has demonstrated the ability to conserve labile PTMs. These methods have only demonstrated modest sequence coverage for large proteins and protein complexes.

IRMPD induces fragmentation in a similar manner to UVPD. Trapped protein ions are irradiated with low energy photons, typically 10.6 um, which corresponds to ~ 0.1 eV of energy per

photon. A large number of these low energy photons must be absorbed to gain sufficient energy to dissociate an intact protein.

Employing UVPD or IRMPD requires implementing expensive laser modules to generate the energetic photons responsible for fragmentation. ETD requires using radical anion reagents to transfer electrons to biomolecules. While these methods of activation have demonstrated an improved ability to fragment proteins and proteins complexes relative to traditional CID, they are far less available and limited to a small subset of MS instrumentation. More importantly, these activation technologies still lack the ability to generate the sequence coverage required to actualize the analysis of protein complex proteoforms.[33]

Part 2. Chemical Modifications for Improvement and Study of Ion Activation

Applying chemical derivatization technologies to covalently tether charge bearing functional groups to a substrate and influence the way that it fragments when energized is an attractive technique to apply to top-down protein sequencing. The ability to covalently tether fixed positive charges to peptides prior to MS analysis and CID activation has been demonstrated in the form of sulfonium based reagents but adapting this technology to native mass spectrometry is challenging.[32]

Adapting any form of covalent modification to top-down proteomics requires a high labeling specificity. Ideally, modifications can be localized to one or two amino acid residues. Multiple potential sites of modification can result in complex isobaric mixtures of modified species. This

complex mixture of species can dissociate into even more complicated peptide ions and impede sequencing attempts.[35]

Fixed charges that have been tethered to intact proteins and protein complexes dissociate at lower activation energies than is required to induce fragmentation of the biomolecule itself. The loss of the charged tags during activation does not improve fragmentation of intact proteins. In fact, it only serves to complicate the data analysis process. Fixed charges that remain tethered through the CID process can alter the dissociation pathways of protein assemblies and have the potential to improve fragmentation efficiency in native mass spectrometry experiments.[33]

Chapter 3

Trimethyl Pyrilium Modification for Enhanced Top-Down Analysis

Part 1. Precedence

Polasky et. al. recently reported the use of an inexpensive pyrilium reagent, trimethyl pyrilium (TMP), to tether fixed positive charges to lysine residues of intact protein complexes (Figure 5). Notably, the pyrilium reagent does not dissociate during the CID activation process and provides

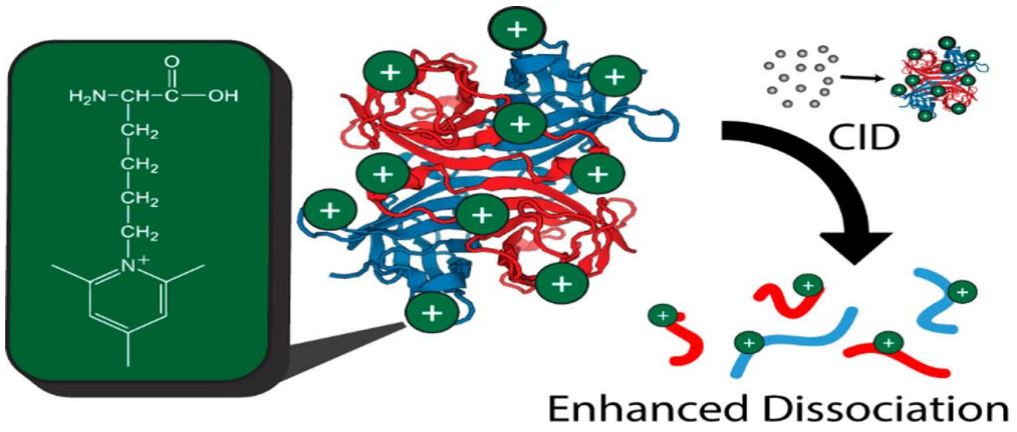

Figure 5. Reaction of the pyrilium with a primary amine results in a pyridinium derivative with a fixed positive charge on the nitrogen atom of the former amine. Plus signs indicate positive charges localized to lysine residues or N-termini throughout the protein complex.[33]

increased fragmentation of the native protein assembly.[33] The ability to manipulate the charged environment of the substrate using CID stable fixed charges while preserving the non-covalent interactions characteristic of native protein complexes was explored in Avidin, alcohol dehydrogenase, and ovalbumin systems. Avidin, alcohol dehydrogenase, and ovalbumin are tetramers with masses of 64, 148, and 170 kDa respectively. The generation of TMP fragments from these native assemblies enabled sequencing of previously inaccessible regions of the complexes. In the case of ADH, employing TMP modification resulted in the identification of 57 additional residues not accessible when using ECD or ETD methods of activation. This translated to a total top-down sequence coverage of ~50%. This level of coverage is unprecedented for assemblies of this size and a dramatic improvement when compared to sequence information available by using UVPD, ECD, or ETD activation technologies.

In this thesis, we explore the details surrounding this cheap chemical derivatization reagent (TMP) coupled with the most widely available fragmentation technique (CID) by targeting smaller protein systems to ascertain how individual TMP modification states influence overall protein sequence coverage. Such information can then be fed back into our ongoing efforts to optimize the TMP modification for top-down proteomics applications more generally.

Part 2. Applying Trimethyl Pyrilium Modification to a New Protein

Current instrumentation lacks the mass resolving power to discern between TMP and modified states for large protein complexes. As such, in order to assess the impact of individual TMP modifications on the CID fragmentation chemistry of a protein, we have chosen to study the Small EDRK rich factor (SERF) protein. SERF has recently been identified as a positive regulator of protein aggregation, aiding in the construction of amyloid fibrils that are characteristic of many human neurodegenerative disorders e.g. Parkinon's Disease and Alzheimer's Disease.[37] A monomeric protein containing a large number of lysine residues relative to all other amino acids making up its sequence (14 lysines/68 total residues) and weighing 8 kDa, SERF has been identified as a good candidate for exploring trimethyl pyrilium modification technology.[37]

Part 3. TMP Modification Efficiency

Applying TMP labeling to a new system requires first validating the ability to modify the lysine residues in that system. While SERF possesses 14 lysine residues that have the potential to be modified by TMP, factors such as solubility of the protein in the reaction buffer and the degree to which each lysine residue is exposed to the solvent can influence reactivity. In applying this

chemical derivatization technology to SERF, we have observed that the number of TMP modifications covalently tethered has a significant dependence on the reaction stoichiometry of the labeling reagent relative to the amount of protein in solution. A general reaction scheme was followed while varying the equivalents of the trimethyl pyrilium labeling reagent (Figure 6).

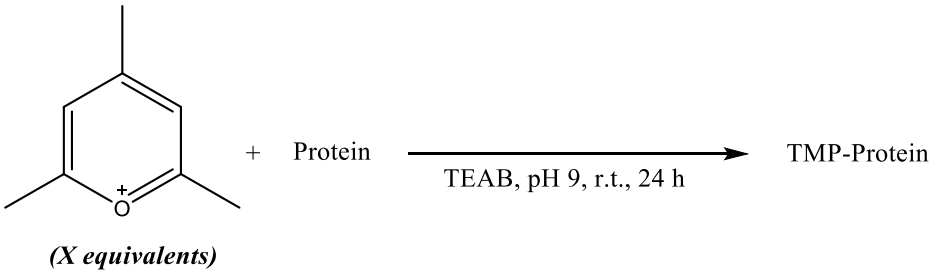

Figure 6. Generalized scheme of trimethyl pyrilium reacting with protein (SERF). Except for equivalents of TMP, all conditions described were kept constant for all modification reactions.

Applying the same set of reaction conditions, save for the equivalents of trimethyl pyrilium relative to protein in solution, we can manipulate the number of modifications applied to the system. The reaction was performed using 25, 100, 250, and 1000 equivalents of TMP with respect to protein. This set of four reaction conditions has allowed us to successfully resolve each individual modification state of SERF; ranging from zero to fourteen lysines modified (no modification versus complete modification). The entire range of TMP tagging is most easily observed in the +7 charge state. Each TMP added increases mass by 104.06 Daltons. In the 7+ charge state, m/z shifts of 14.9 are indicative of the modification series (Figure 7). The number of modifications increases as the number of equivalents of TMP in solution increases.

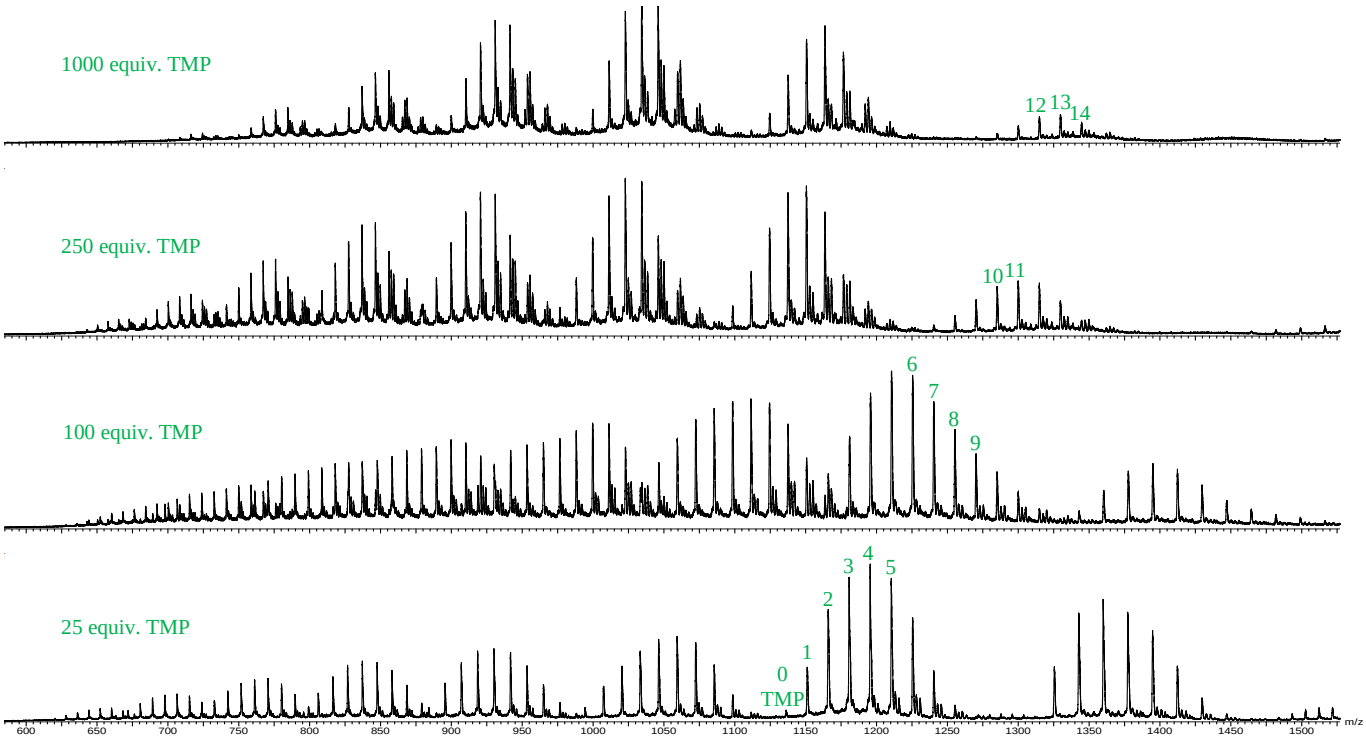

Figure 7. Stacked mass spectra highlighting the full range of TMP modification states achievable in SERF system. Increasing equivalents of TMP give rise to an increased number of modifications.

Part 4. Effect of TMP Modifications on Required Fragmentation Energy

After successfully producing the full range of modification states in SERF, we investigated the effect that the number of TMP modifications had on the conditions necessary to induce fragmentation of the intact protein using CID. The degree of fragmentation depends on the amount of kinetic energy the ion possesses when it enters the "Tri-wave region." The kinetic energy of ions entering this region can be manipulated by raising/lowering the trap collision voltage (Trap CE). Trap CE acts as a potential difference between the "High m/z quadrupole" and the "Tri-wave region" and is used to accelerate/decelerate ions for CID experiments.[20] The minimum Trap CE required to fragment a quadrupole selected ion until its intensity decreased by

50% was evaluated for TMP modification states of sufficient intensity (0-10 modifications) (Figure 8).

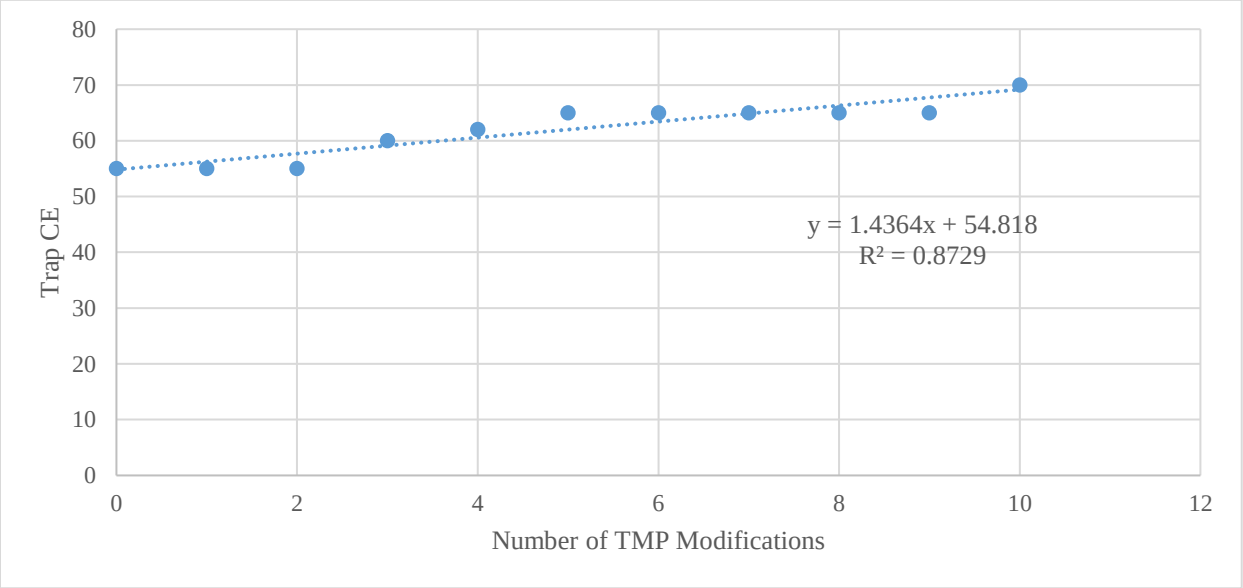

Figure 8. Minimum Trap CE required to reduce the intensity of quadrupole selected parent ion by 50% as a function of number of TMP modifications.

The increase in Trap CE required to fragment increasingly modified SERF could be explained by the increase in total mass of the ion. Each trimethyl pyrilium tag adds 104.06 Daltons, which is 1.3% of the mass of the entire protein. SERF that has been completely modified has a mass ~20% larger relative to the unmodified species. This is not unprecedented, higher energies are often required to activate and dissociate larger protein complexes. The small increase in energy required to induce fragmentation of the modified species is outweighed by the generation of informative new peptide fragments.

Part 5. Effect of TMP Modifications on Fragments Generated

The effect of TMP modifications on peptide fragments generated is easily visualized by comparing sequence coverage maps. Sequence coverage maps quickly display whether or not

any peptides that were generated in the CID process can be used to identify residues present in the protein system. These maps also distinguish whether or not detected peptide fragments include the N- or C- terminus. Sequence coverage maps comparing unmodified SERF and singly modified SERF are shown (Figure 9). Details regarding data analysis and generation of sequence coverage maps is discussed in the Experimental section.

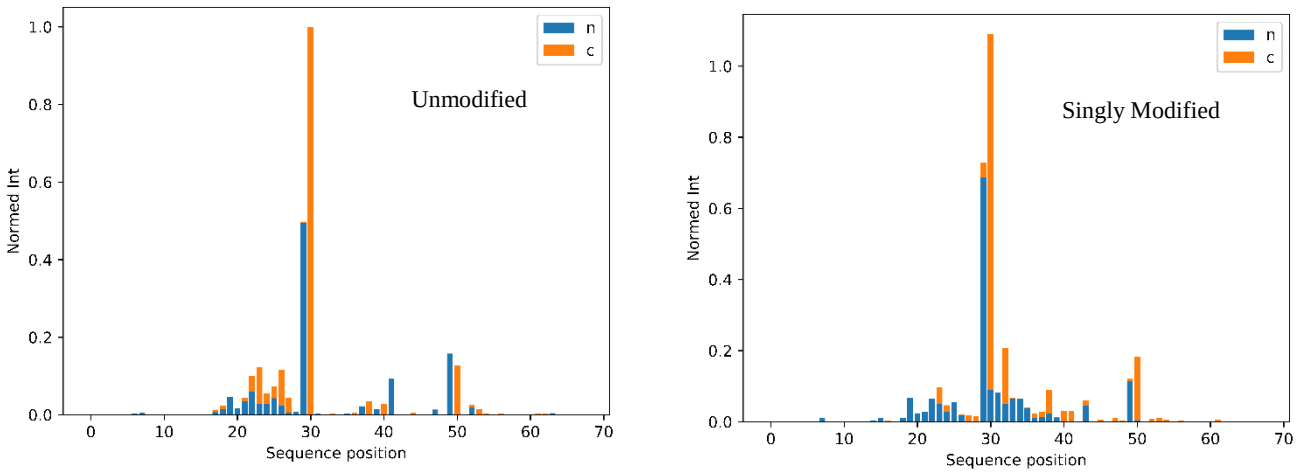

Figure 9. Sequence coverage maps highlighting differences observed between control (unmodified) and singly modified SERF.

The addition of a single fixed charge dramatically improved the ability to generate CID induced peptide fragments that can be used to identify residues 30-40 in the protein sequence. Not only did the covalent tethering of one TMP modification to SERF improve coverage of residues 30-40, but there was no loss of information relative to the unmodified species. This improvement was quantified by comparing the difference in total normalized intensity for ions corresponding to residues 30-40 (Figure 10).

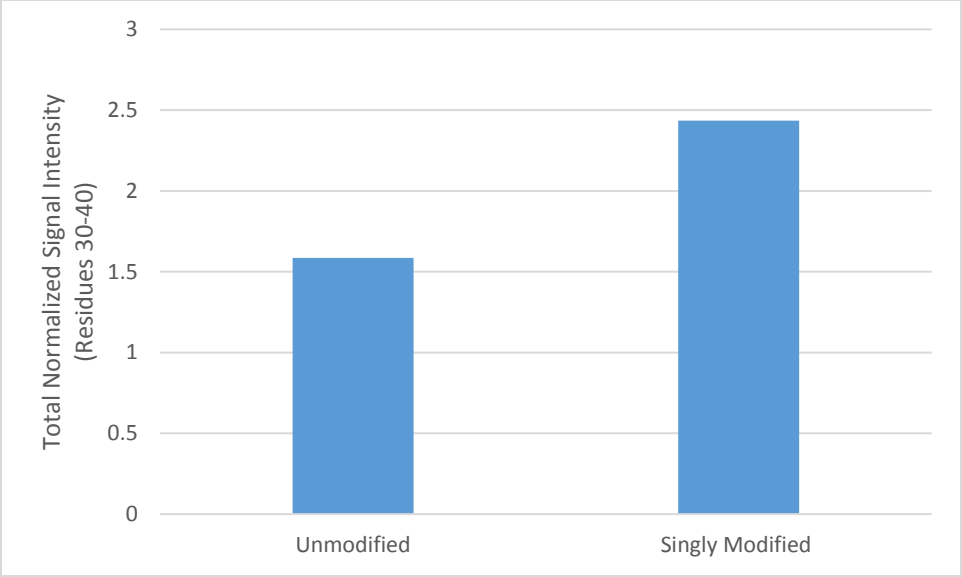

Figure 10. Total normalized signal intensity for ions corresponding to residues 30-40 in unmodified and singly modified SERF.

The addition of one TMP modification produced a 54% increase in total normalized intensity corresponding to amino acid residues 30-40 relative to the unmodified variant.

CID induced fragmentation of unmodified SERF struggled to provide sequence information relevant to positions 30-69. An increased number of modifications, 10 TMPs, provided an even more dramatic improvement in overall sequence information from fragmentation of the modified protein (Figure 11).

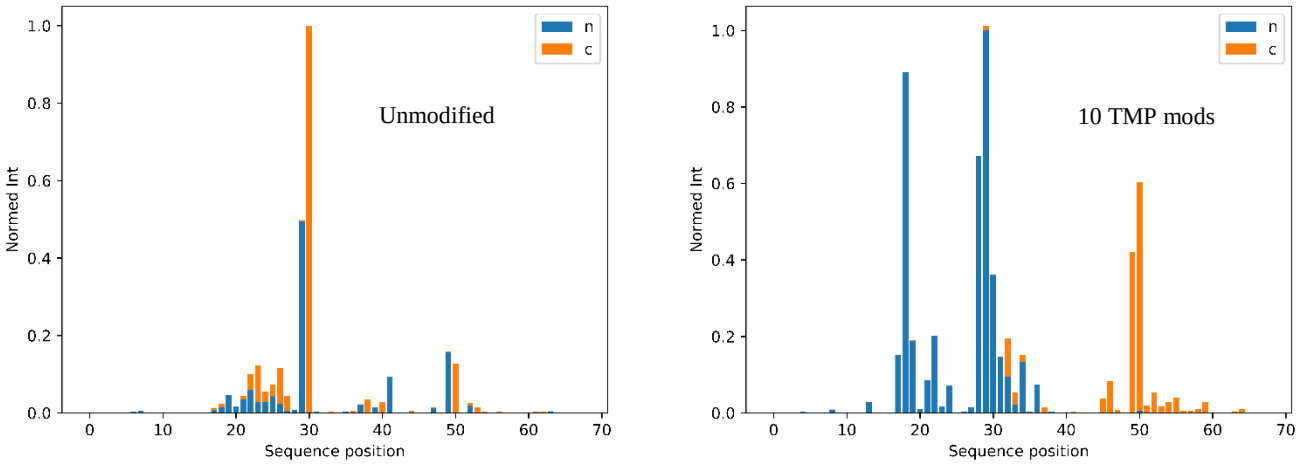

Figure 11. Sequence coverage maps highlighting differences observed between control (unmodified) and SERF with 10 TMP modifications.

Fixing 10 CID stable positive charges to lysine residues in SERF in the form of trimethyl pyrilium has facilitated enhanced fragmentation of the intact protein. This enhanced fragmentation has resulted in an increase in sequence coverage relative to the unmodified protein. The improvement in sequence coverage was quantified by comparing the total number of amino acid residues covered between unmodified and highly modified (10 TMPs) protein variants (Figure 12). An amino acid residue was considered "covered" if the intensity of an ion corresponding to that residue exceeded 1% of the total normalized intensity. Fragmentation of the highly modified species provided 30% more sequence coverage compared to the unmodified protein.

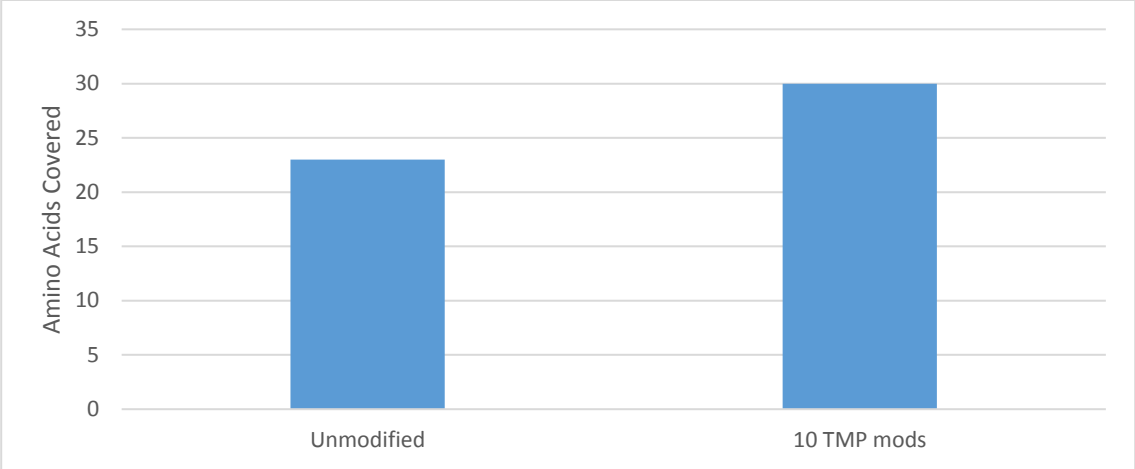

Figure 12. Number of amino acid residues covered in unmodified and highly modified (10 TMPs) SERF.

The coupling of cheap chemical derivatization technology in the form of TMP modification with the most widely available fragmentation technique (CID) has proved highly beneficial; especially in the context of protein sequence coverage.

Part 6. Future Directions

We have observed discrepancies between the numbers of fixed charges added in the form of TMP modifications and the charge state of the protein ion e.g. fourteen fixed charges corresponding to an ion in the +7 charge state. This is evidence that has led us to believe the addition of fixed charges to the protein system is promoting zwitterion formation through charge pairing, presumably with some acidic residues in the protein. While the addition of CID-stable fixed charges changes the energy landscape to produce new peptide fragments, zwitterion formation provides a stabilizing effect, partially mitigating these benefits. Acidic residues can be protected to eliminate their ability to pair with nearby positive charges.[35] By capping the acidic residues before tethering fixed positive charges, we will investigate the role of zwitterions in

charge dependent fragmentation pathways to inform future improvements in dissociation of intact proteins.

Part 7. Experimental Methods

Chemical modification.

SERF from *Saccharomyces cerevsiae* was overexpressed in E. Coli and purified by ion exchange chromatography and size exclusion chromatography. SERF was dissolved in 100 mM triethylammonium bicarbonate (TEAB), pH 9, to make solutions containing 100 uM protein for chemical modification. 2,4,6-Trimethyl pyrylium (TMP) tetrafluoroborate (Alfa Aesar, Haverhill, MA) was dissolved in 100 mM TEAB, pH 9, vortexed for ten seconds to dissolve, and quickly added to protein solutions at varying molar excess relative to the protein complex concentration. Reaction solutions were briefly vortexed and allowed to react for 24 hours at room temperature. Following modification, proteins were buffer exchanged into 100 mM ammonium acetate pH 7.4 (Sigma Aldrich) with P6 microspin columns (BioRad Laboratories, Hercules, CA) according to manufacturer instructions. Buffer exchanged samples were either analyzed immediately or flash frozen with liquid nitrogen and stored at -80 ◦C prior to analysis.

Ion Mobility-Mass Spectrometry.

A quadrupole ion mobility-time of flight mass spectrometer (Synapt G2 HDMS, Waters, Milford, MA) was used for all ion mobility experiments. 5 uL of buffer-exchanged protein solution (100 uM) was transferred to a gold-coated borosilicate capillary (0.78mm i.d., Harvard Apparatus, Holliston, MA) for direct infusion. Instrumental settings were optimized to preserve intact protein complexes prior to activation: capillary voltage 1.5 kV, sample cone 40 V,

extraction cone 0 V. Gas flows (mL/min): source: 30, trap: 2, helium cell: 200, IMS: 90. IMS Settings: wave velocity: 300 m/s, wave height: 5.4 V, bias: 10 V. Backing pressure was set to 2.6 mbar. A single charge state of each protein complex was selected and collisionally activated in the trap cell (trap collision voltage: ranging from 50 V to 70 V) prior to ion mobility separation. Scans were combined for 10 minutes to obtain sufficient signal to noise ratios.

Data Analysis

Ion mobility mass spectrometry fragmentation data was analyzed using a recently developed software suite consisting of two programs: IMTBX (IM Toolbox) and Grppr (Grouper).[34] This software was specifically designed to process data generated in complex top-down proteomics experiments and allowed for a highly automated workflow. The first step in the analysis workflow requires picking peaks out of raw data and is performed using IMTBX. The second step in the analysis workflow requires isotopic peak grouping and is performed by Grppr. IMTBX added together or averaged all ion mobility scans. It then applies a 2D Gaussian filter for smoothing. Local maxima in the averaged or summed scans are found and used to fit 2D Gaussians. Averaged data with a signal-to-noise ratio of less than 2 was filtered. Final peak lists generated by IMTBX were then imported into Grppr for isotopic grouping. Grppr has two available algorithms for isotopic cluster detection: "Convex" and "Averagine". The Averagine algorithm was used in all data analysis. This algorithm calculates theoretical isotopic distributions for all peaks in the list created by IMTBX and compared to the data. Sequence coverage plots were generated using an additional in-house script written in Python 3.5.

.

References

1. Dayhoff, M. O. & Foundation., N. B. R. *Atlas of protein sequence and structure. [Vol. 1]*, (National Biomedical Research Foundation, 1969).

2. Smith, L. M., Kelleher, N. L. & Proteomics, T. C. for T. D. Proteoform: a single term describing protein complexity. *Nat. Methods* **10,** 186 (2013).

3. Toby, T. K., Fornelli, L. & Kelleher, N. L. Progress in Top-Down Proteomics and the Analysis of Proteoforms. *Annu. Rev. Anal. Chem.* **9,** 499–519 (2016).

4. Yates, J. R. The Revolution and Evolution of Shotgun Proteomics for Large-Scale Proteome Analysis. *J. Am. Chem. Soc.* **135,** 1629–1640 (2013).

5. Catherman, A. D., Skinner, O. S. & Kelleher, N. L. Top Down proteomics: Facts and perspectives. *Biochem. Biophys. Res. Commun.* **445,** 683–693 (2014).

6. Siuti, N. & Kelleher, N. L. Decoding protein modifications using top-down mass spectrometry. *Nat. Methods* **4,** 817 (2007).

7. Savaryn, J. P., Catherman, A. D., Thomas, P. M., Abecassis, M. M. & Kelleher, N. L. The emergence of top-down proteomics in clinical research. *Genome Med.* **5,** 53 (2013).

8. Chait, B. T. Mass Spectrometry: Bottom-Up or Top-Down? *Science* **314,** 65–66 (2006).

9. Heck, A. J. R. Native mass spectrometry: a bridge between interactomics and structural biology. *Nat. Methods* **5,** 927 (2008).

10. Leney, A. C. & Heck, A. J. R. Native Mass Spectrometry: What is in the Name? *J. Am. Soc. Mass Spectrom.* **28,** 5–13 (2017).

11. Chandler, S. A. & Benesch, J. L. P. Mass spectrometry beyond the native state. *Curr. Opin. Chem. Biol.* **42,** 130–137 (2018).

12. Li, H., Wongkongkathep, P., Van Orden, S. L., Ogorzalek Loo, R. R. & Loo, J. A. Revealing Ligand Binding Sites and Quantifying Subunit Variants of Noncovalent Protein Complexes in a Single Native Top-Down FTICR MS Experiment. *J. Am. Soc. Mass Spectrom.* **25,** 2060–2068 (2014).

13. Chait, B. T., Cadene, M., Olinares, P. D., Rout, M. P. & Shi, Y. Revealing Higher Order Protein Structure Using Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **27,** 952—965 (2016).

14. Sharon, M., Taverner, T., Ambroggio, X. I., Deshaies, R. J. & Robinson, C. V. Structural Organization of the 19S Proteasome Lid: Insights from MS of Intact Complexes. *PLOS Biol.* **4,** (2006).

15. Breuker, K., Jin, M., Han, X., Jiang, H. & McLafferty, F. W. Top-Down Identification and Characterization of Biomolecules by Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **19,** 1045–1053 (2008).

16. Robinson, C. V, Sali, A. & Baumeister, W. The molecular sociology of the cell. *Nature* **450,** 973 (2007).

17. Xie, Y., Zhang, J., Yin, S. & Loo, J. A. Top-Down ESI-ECD-FT-ICR Mass Spectrometry Localizes Noncovalent Protein-Ligand Binding Sites. *J. Am. Chem. Soc.* **128,** 14432–14433 (2006).

18. Benesch, J. L. P., Ruotolo, B. T., Simmons, D. A. & Robinson, C. V. Protein Complexes in the Gas Phase: Technology for Structural Genomics and Proteomics. *Chem. Rev.* **107,** 3544–3567 (2007).

19. Pukala, T. L. *et al.* Subunit Architecture of Multiprotein Assemblies Determined Using Restraints from Gas-Phase Measurements. *Structure* **17,** 1235–1243 (2009).

20. Ruotolo, B. T., Benesch, J. L. P., Sandercock, A. M., Hyung, S.-J. & Robinson, C. V. Ion mobility–mass spectrometry analysis of large protein complexes. *Nat. Protoc.* **3,** 1139 (2008).

21. Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F. & Whitehouse, C. M. Electrospray ionization for mass spectrometry of large biomolecules. *Science* **246,** 64–71 (1989).

22. Jennings, K. R. The changing impact of the collision-induced decomposition of ions on mass spectrometry. *Int. J. Mass Spectrom.* **200,** 479–493 (2000).

23. Bowers, W. D., Delbert, S. S., Hunter, R. L. & McIver Jr, R. T. Fragmentation of oligopeptide ions using ultraviolet laser radiation and Fourier transform mass spectrometry. *J. Am. Chem. Soc.* **106,** 7288–7289 (1984).

24. Shaw, J. B. *et al.* Complete Protein Characterization Using Top-Down Mass Spectrometry and Ultraviolet Photodissociation. *J. Am. Chem. Soc.* **135,** 12646–12651 (2013).

25. Moon, J. H., Shin, Y. S., Cha, H. J. & Kim, M. S. Photodissociation at 193 nm of some singly protonated peptides and proteins with m/z 2000–9000 using a tandem time-of-flight mass spectrometer equipped with a second source for delayed extraction/post-acceleration of product ions. *Rapid Commun. Mass Spectrom.* **21,** 359–368.

26. Syka, J. E. P., Coon, J. J., Schroeder, M. J., Shabanowitz, J. & Hunt, D. F. Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proc. Natl. Acad. Sci.* **101,** 9528–9533 (2004).

27. Lermyte, F. & Sobott, F. Electron transfer dissociation provides higher-order structural information of native and partially unfolded protein complexes. *Proteomics* **15**, 2813–2822 (2015).

28. McLafferty, F. W. *et al.* Electron capture dissociation of gaseous multiply charged ions by fourier-transform ion cyclotron resonance. *J. Am. Soc. Mass Spectrom.* **12,** 245–249 (2001).

29. Zhang, H., Cui, W., Wen, J., Blankenship, R. E. & Gross, M. L. Native Electrospray and Electron-Capture Dissociation FTICR Mass Spectrometry for Top-down Studies of Protein Assemblies. *Anal. Chem.* **83,** 5598–5606 (2011).

30. Li, H. *et al.* Structural Characterization of Native Proteins and Protein Complexes by Electron Ionization Dissociation-Mass Spectrometry. *Anal. Chem.* **89,** 2731–2738 (2017).

31. Little, D. P., Speir, J. P., Senko, M. W., O'Connor, P. B. & McLafferty, F. W. Infrared Multiphoton Dissociation of Large Multiply Charged Ions for Biomolecule Sequencing. *Anal. Chem.* **66,** 2809–2815 (1994).

32. Mendoza, V. L. & Vachet, R. W. Probing protein structure by amino acid-specific covalent labeling and mass spectrometry. *Mass Spectrom. Rev.* **28,** 785–815 (2009).

33. Polasky, D. A. *et al.* Fixed-Charge Trimethyl Pyrilium Modification for Enabling Enhanced Top-Down Mass Spectrometry Sequencing of Intact Protein Complexes. *Anal. Chem.* **90,** 2756–2764 (2018).

34. Avtonomov, D. M., Polasky, D. A., Ruotolo, B. T. & Nesvizhskii, A. I. IMTBX and Grppr: Software for Top-Down Proteomics Utilizing Ion Mobility-Mass Spectrometry. *Anal. Chem.* **90,** 2369–2375 (2018).

35. Zhang, H., Wen, J., Huang, R. Y.-C., Blankenship, R. E. & Gross, M. L. Mass spectrometry-based carboxyl footprinting of proteins: Method evaluation. *Int. J. Mass Spectrom.* **312,** 78–86 (2012).

36. Coles, J. & Guilhaus, M. Orthogonal acceleration — a new direction for time-of-flight mass spectrometry: Fast, sensitive mass analysis for continuous ion sources. *TrAC Trends Anal. Chem.* **12,** 203–213 (1993).

37. Yoshimura, Y. *et al.* MOAG-4 promotes the aggregation of α-synuclein by competing with self-protective electrostatic interactions. *J. Biol. Chem.* **292,** 8269–8278 (2017).

38. Pagel, K., Hyung, S.-J., Ruotolo, B. T. & Robinson, C. V. Alternate Dissociation Pathways Identified in Charge-Reduced Protein Complex Ions. *Anal. Chem.* **82,** 5363–5372 (2010).

39. Greer, S. M. & Brodbelt, J. S. Top-Down Characterization of Heavily Modified Histones Using 193 nm Ultraviolet Photodissociation Mass Spectrometry. *J. Proteome Res.* **17,** 1138–1145 (2018).