# UCLA
## UCLA Electronic Theses and Dissertations

**Title**

Development of Electrospray Ionization-Mass Spectrometry for Analysis of Water Soluble and Membrane Proteins and Educational Protocols for an Analytical Chemistry Class

**Permalink**

https://escholarship.org/uc/item/0726179b

**Author**

Jung, Wonhyeuk

**Publication Date**

2021

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Development of Electrospray Ionization-Mass Spectrometry for Analysis of Water Soluble and

Membrane Proteins and Educational Protocols for an Analytical Chemistry Class

A dissertation submitted in partial satisfaction of the requirements for the degree Doctor of

Philosophy in Biochemistry, Molecular and Structural Biology

by

Wonhyeuk Jung

2021

ABSTRACT OF THE DISSERTATION

Development of Electrospray Ionization-Mass Spectrometry for Analysis of Water Soluble and Membrane Proteins and Educational Protocols for an Analytical Chemistry Class

by

Wonhyeuk Jung

Doctor of Philosophy in Biochemistry, Molecular and Structural Biology

University of California, Los Angeles, 2021

Professor Joseph Ambrose Loo, Chair

Native mass spectrometry (MS) is a branch of MS analysis in which the structure of the target analytes of interest are kept intact and remaining in their "native" functional structure (as much as possible). This approach was made possible by the development of electrospray ionization (ESI), a soft ionization technique that does not fragment the target analyte during the ionization process while inducing multiple charging. The multiply charged biomolecules, in turn, can be subjected to fragmentation via collisional activation with a non-reactive gas such as nitrogen. This approach of combining native MS with fragmentation-based analysis, termed native top-down MS analysis, can be applied to large biomolecules such as membrane proteins to gain structural insights.

Membrane proteins present unique challenges to conventional high-resolution structural techniques due to their hydrophobic nature. However, they are responsible for various physiological phenomena and account for 60% of known druggable targets in the cell. Thus, there is a need for an approach that can overcome issues with membrane protein analysis while complementing other biophysical techniques used to probe protein structure.

Here, how native top-down MS can play this role is presented. The effects of non-ionic saccharide-based detergents, a commonly used class of detergents for membrane protein solubilization, on the resulting charge states of soluble proteins is investigated to gain insights into the mechanism of ESI. The MS-fragmentation patterns from collisionally activated dissociation of membrane proteins and membrane protein-lipid complexes are compared. How new insights into the lipid binding sites can be gained by detecting lipid-bound MS-fragments is presented. The result of the study indicates that native top-down MS analysis can provide unique structural insights for membrane proteins and their non-covalent interactions.

When the analytical goal is to investigate the atomic composition of the target analyte, an ionization approach in which the sample is fully atomized before MS analysis is preferred instead. Inductively coupled plasma ionization, which atomizes and ionizes the sample via a plasma, can be coupled with MS analysis (ICP-MS) to quantify heavy metal contamination in complex samples. A protocol for ICP-MS analysis of commercial fish products for mercury contamination detection developed to aid an analytical chemistry class for instruction of undergraduate chemistry students is presented.

The dissertation of Wonhyeuk Jung is approved.

Jose Rodriguez

James Bowie

Julian Whitelegge

John Colicelli

Joseph Loo, Committee Chair

University of California, Los Angeles

2021

TABLE OF CONTENTS

LIST OF FIGURES

**Chapter 3:**

**Chapter 4:**

# LIST OF TABLES

# ACKNOWLEDGEMENTS

VITA

**Education & Degrees:**

University of Southern California, Keck School of Medicine, Los Angeles, CA

Sept. 2014 – May 2016

Master of Science, Major in Biochemistry and Molecular Biology

Korea University, College of Life Sciences and Biotechnology, Seoul, South Korea

Mar. 2006 – Mar. 2014

Bachelor of Science in Life Sciences

**Publications:**

Choi, D., Ramu, S., Park, E., Jung, E., Yang, S., Jung, W., Choi, I., Lee, S., Kim, K. E., Seong, Y. J., Hong, M., Daghlian, G., Kim, D., Shin, E., Seo, J. I., Khatchadourian, V., Zou, M., Li, W., De Filippo, R., … Hong, Y. K. (2016). Aberrant activation of notch signaling inhibits PROX1 activity to enhance the malignant behavior of thyroid cancer cells. Cancer Research, 76(3), 582–593.

Donnelly, D. P., Rawlins, C. M., DeHart, C. J., Fornelli, L., Schachner, L. F., Lin, Z., Lippens, J. L., Aluri, K. C., Sarin, R., Chen, B., Lantz, C., Jung, W., Johnson, K. R., Koller, A., Wolff, J. J., Campuzano, I. D. G., Auclair, J. R., Ivanov, A. R., Whitelegge, J. P., … Agar, J. N. (2019). Best practices and benchmarks for intact protein analysis for top-down mass spectrometry. Nature Methods, 16(7), 587–594.

Zenaidee, M. A., Lantz, C., Perkins, T., Jung, W., Loo, R. R. O., & Loo, J. A. (2020). Internal Fragments Generated by Electron Ionization Dissociation Enhance Protein Top-Down Mass Spectrometry. Journal of the American Society for Mass Spectrometry, 31(9), 1896–1902.

**Awards:**

2020 Michael E. Jung Excellence in Teaching Award, UCLA Chemistry and Biochemistry Department

**Oral Presentations:**

Conference on Ion Chemistry and Mass Spectrometry, January 17-19, 2020, Lake Arrowhead, CA [Detergents and Chemical Additives' Effect on Charge State Distributions During Mass Spectrometry Analysis]

68th American Society for Mass Spectrometry Conference, June 1-12, 2020, On-line conference [Effect of Nonionic Saccharide-based Detergents and Supercharging Agents on Native Mass Spectrometry of Membrane Proteins]

69th American Society for Mass Spectrometry Conference, November 1-4, 2021, Philadelphia, PA [Native Top-Down Mass Spectrometry to Probe the Structural Effects of Lipid Binding to Membrane Proteins]

**Introduction**

**1.1 Mass spectrometry as a tool for proteoform characterization and the development of 'soft' ionization techniques**

The number of human genes is estimated to be ~22,000, which is a relatively low number considering our biological diversity and complexity.[1] However, during transcription, most of the human genes can go through alternative splicing, and thus the number of potential isoforms climbs to ~70,000.[2] Further diversity can stem from post-translational modifications (PTMs) and for certain proteins such as immunoglobulins and T-cell receptors, somatic recombination can result in billions of potential protein variants.[2] In 2013, 'proteoform' was introduced as a term[3] to comprehensively depict this protein complexity (**Figure 1**).
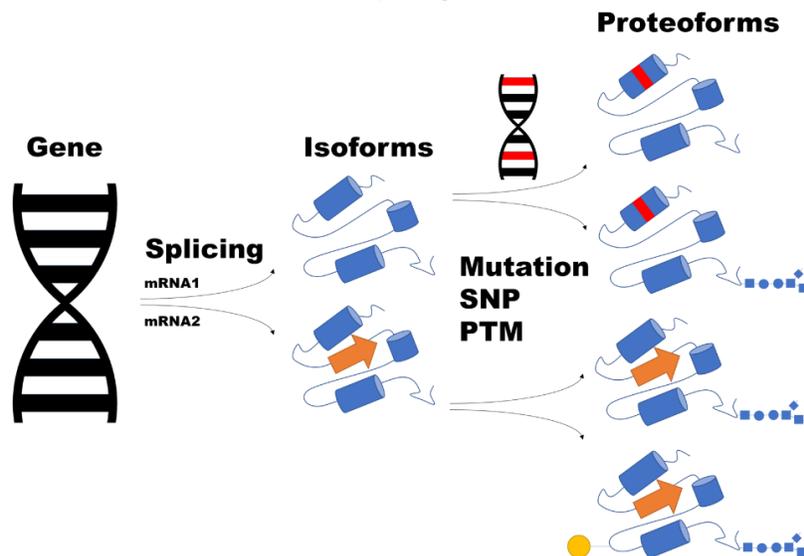


**Figure 1. Cartoon depiction of how multiple proteoforms can arise from a single gene.** From a single gene, alternative splicing, or the use of different promoters and/or translation start sites can result in different isoforms. Then, mutations, and single nucleotide polymorphisms (SNPs) can introduce primary sequence changes (represented in red). PTMs, such as N-glycosylation and phosphorylation, can introduce additional variations, resulting in an exponential increase in the number of proteoforms.

Mass spectrometry (MS) has emerged as a powerful method for proteoform investigation. To perform MS analysis, target analytes of interest are ionized and then introduced to the mass spectrometer to determine the mass-to-charge ratio (commonly denoted as m/z), which in turn yields accurate mass information. MS is extremely sensitive and can reliably detect targets in the fmol (10-15 mol) range with high accuracy.[4] This sensitivity and accuracy, in turn, can be used to detect proteoforms that are difficult to analyze. For instance, new proteoforms of the inward rectifier potassium channel (Kir) 2.1, an integral membrane protein, which is critical for the maintenance of the resting membrane potential and phase-3 repolarization of the cardiac action potential in the heart, were identified through MS-based analysis.[5] Membrane proteins are a challenging class of proteins for proteoform characterization (see section 1.5 for more details) and this study highlights the capacity of MS analysis to handle difficult samples.

Proteins may form various non-covalent interactions to perform biological functions, which introduces additional layers of complication to the proteoform landscape. Thus, there has been a push for further MS development to enable structural studies with focus on higher-order structure/interaction. Again focusing on membrane protein studies, Yen et al. captured the interaction between the human purinergic receptor P2Y1R, a class A G-protein coupled receptor (GPCR), and its ligands through native MS analysis.[6] Laganowsky et al., in another landmark study, investigated through combination of crystallography, biochemical assay, and MS how lipid interactions modulate membrane proteins' structure and function.[7] These studies show the potential of MS to provide a new way to capture and probe non-covalent interactions important for physiological phenomena.

One of the breakthroughs that made these scientific advances possible was the development of 'soft' ionization techniques (little to no fragmentation of the target analyte during the ionization process) such as electrospray ionization (ESI)[8] and matrix-assisted laser desorption/ionization (MALDI).[9,10] ESI has especially been successful in biomolecule analysis with the development of

nanoESI (nESI),[11] in which the flow rate is drastically reduced by reducing the diameter of the emitter tip. With the advent of nESI, the sample consumption rate was minimized while ionization efficiency increased.[12,13]

**1.2 Mechanism of ESI and charge manipulation during MS analysis**

**1.2.1 Three major ESI models**

The ESI source is an electrochemical cell that disperses analyte solution into charged droplets. The droplets shrink in size as they evaporate and goes through Coulomb fission process, which is caused by the increase in charge density. The analytes inside the resulting nanodroplets are eventually released into the gas phase. The source region of mass spectrometers is typically heated to facilitate droplet evaporation. The vacuum of the mass spectrometer's interior helps the desolvation process and the transmission of the ions. Although ESI can be done in positive and negative polarity, the discussion here will be largely restricted to the positive ion mode, which is commonly used for protein measurements[14] (**Figure 2**). The precise mechanism of ESI is still under debate and three major models for ions' release from the nanodroplets have been proposed. Dole's charge residue model (CRM)[15] for relatively large, globular proteins, Iribarne

and Thomson's ion evaporation model (IEM)[16] for relatively small ions, and Konermann's chain ejection model (CEM)[17,18] for denatured, elongated proteins (**Figure 3**).
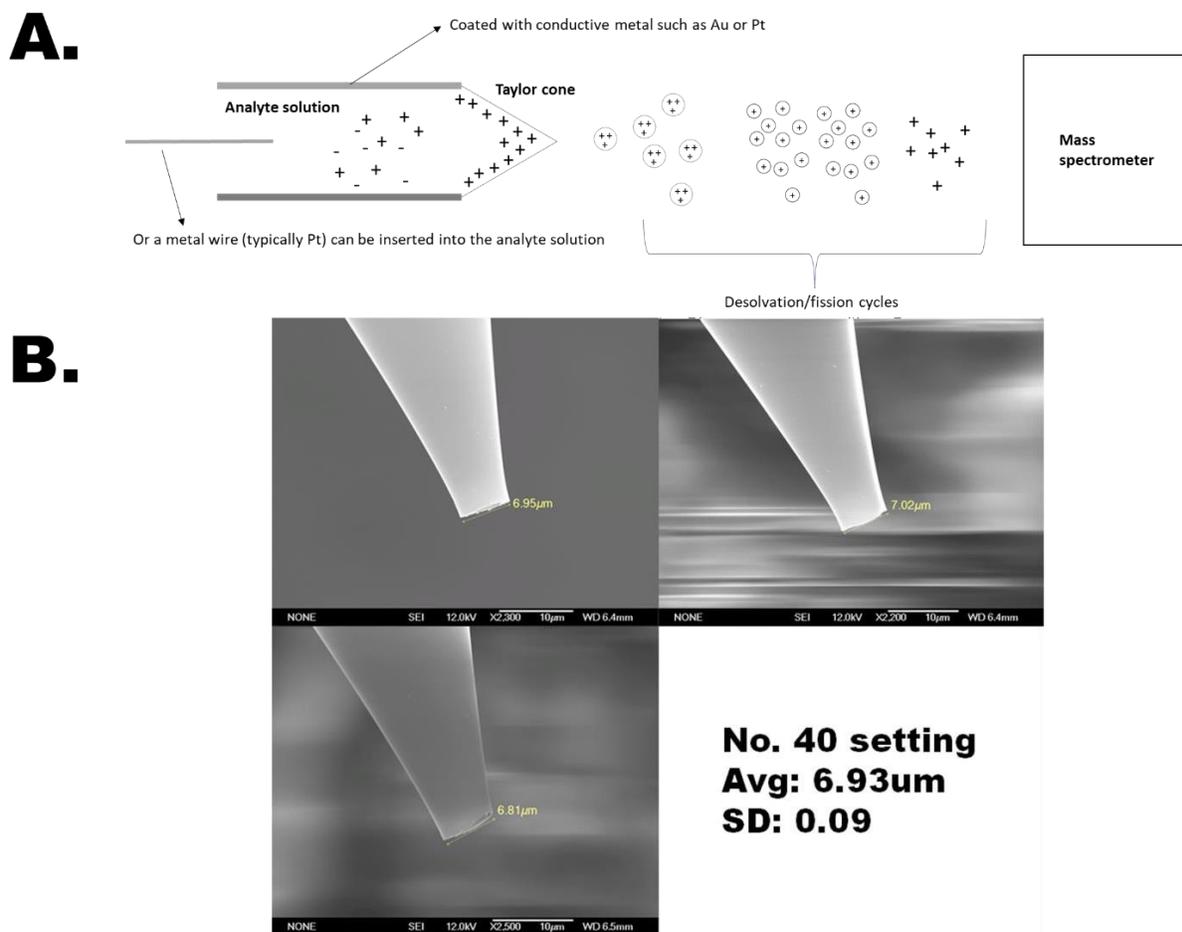


**Figure 2. Cartoon depiction of positive mode ESI and scanning electron microscope images (SEMs) of nESI emitter tip.**

A. By introducing a charge potential along the emitter (typically, the borosilicate tips are coated with conductive metal such as gold or platinum or a platinum wire is inserted into the sample solution), Taylor cone formation is induced. For positive mode ESI, droplets with excess positive charge are ejected. As droplets undergo evaporation, the droplets shrink in size and elevation of charge density leads to Coulomb fission. Eventually, nanodroplets form and ionized target analytes are released into the gas phase. B. SEM pics of nESI emitter tips coated with platinum. These tips were utilized for membrane proteins analysis (see section 1.5 and Chapter 3 for more details). Typically, borosilicate tips are heated and then pulled with a specific force and speed to reach an optimal diameter for the ionization of the target analyte (see Chapter 3 material and method section for more details).

Consider a nanodroplet containing volatile solvent with one large and approximately spherical ion with radius $R_i$. Complete evaporation of the solvent will yield a target analyte ion with the number of charges that is similar to the charge the droplet had an instant before. The radius of the ion and the maximum charge that the ion can harbor will be dictated by the Rayleigh's limiting charge, $q_R$, as depicted in the following equation:

$$q_R e = 8\pi \left( \gamma \varepsilon_0 R_i^3 \right)^{1/2} \quad \text{(Equation 1)}$$

where $e$ is the elementary charge, $\gamma$ is the surface tension of the droplet, and $\varepsilon_0$ is the permittivity of the surrounding medium, which is typically the vacuum for MS analysis.[19] A nanodroplet undergoing evaporation and Coulomb explosions will harbor charge approximately between 70 to 100% of qR.[20] Fernandez de la Mora examined soluble proteins whose size ranged from 17 kDa to 116 kDa and found that the ESI of these proteins results in 68%-88% of $q_R$ when sprayed with an aqueous solvent.[20] Also, Reading et al., found that the charge state of membrane proteins solubilized with saccharide-based detergents closely follows Rayleigh's limit charge.[21] These studies indicate that CRM can predict the charge state when the solvent is close to physiological pH and the protein keeps its native, globular structure throughout the ionization process.

When the target analyte is relatively small and compact while carrying sufficient amount of

charge, IEM can take place from nanodroplets with the diameter approximately at 5.5 nm.[22]

Aliyari et al., conducted molecular dynamic (MD) simulation studies on ESI of ubiquitin across a

wide range of pH and found that a mixture of CRM and IEM-based ionization takes place at pH

4.4 and below.[22] Experimental data agreed well with the expected charge state and the

adduction formation from the MD simulation data, indicating that proteins can go through IEM

during ESI. Thus, for folded proteins, ESI can take place through mixture of CRM and IEM

based on the size, compactness, and charge density of the protein and the size of the

nanodroplet as ion ejection is taking place.
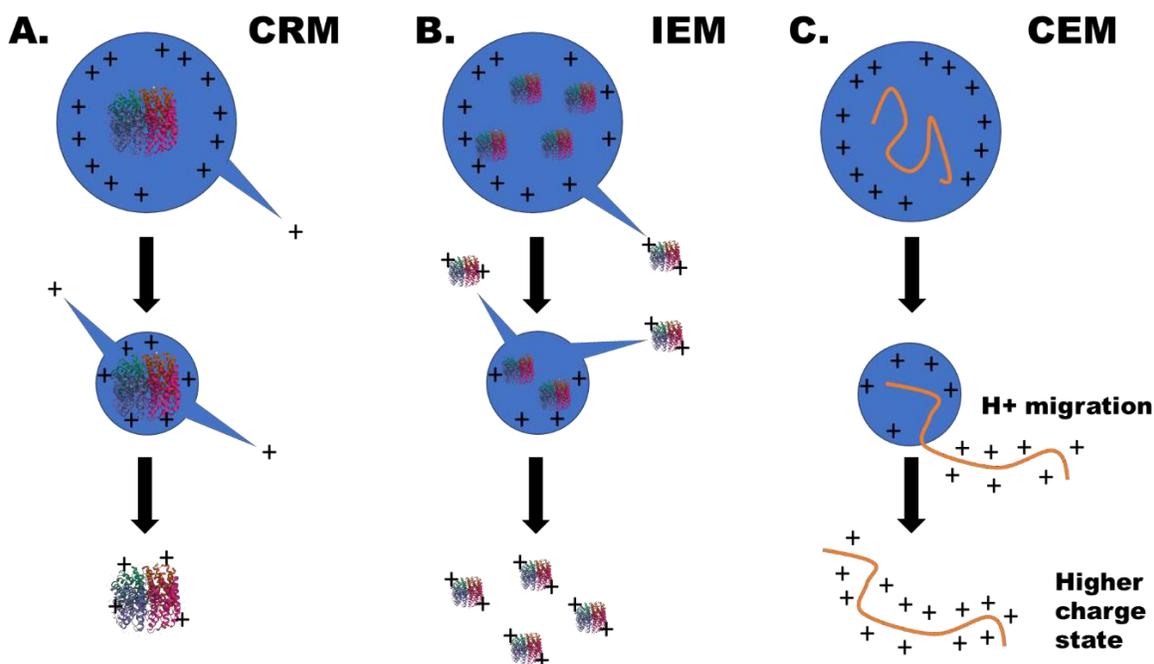


**Figure 3. Cartoon depiction of three major ESI models.**

A. The charge residue model stipulates that throughout the desolvation process, excess
charges are ejected out and the proteins remain inside the nanodroplets. Eventually, complete
evaporation takes place, and the remaining charge is deposited onto the protein. Thus, the final
number of charge that the targe analyte will acquire will be mostly decided by the surface
tension of the nanodroplets and the size of the final nanodroplet. B. Ion evaporation model, in
contrast, specifies that ejection of the charge from the nanodroplets involves the target analytes.
IEM will take place when the protein of interest carries a relatively high charge while maintaining
a compact structure and when the ion is ejected from a relatively large nanodroplet. C. Chain
ejection model explains why denatured, elongated proteins harbor relatively higher charge state
compared to intact, globular proteins. As nanodroplets shrink in size, the denatured protein's

exposed nonpolar residues drive the protein chains to the droplet's surface. Then, a stepwise ejection takes place as protons migrate along the elongated protein. The protein will be coated with charge and the final charge state will be higher compared to CRM or IEM-based ionization.

### 1.2.2 Challenges to CRM and introduction to supercharging

CRM asserts that a relatively large, globular protein will display Rayleigh charge, the charge at which Coulomb repulsion balances with the solvent surface tension[23] of the nanodroplet that is similar to the sphere approximation of the protein. Since the final charge state is decided by the surface tension and size of the protein, lower surface tension should lead to lower charge states and the initial droplet size or the flow rate of the ESI should have no effect. However, non-ionic saccharide-based detergents treatment to the sample prior to ESI, which decreases the overall surface tension of the solvent, results in an increase in charge.[24] Non-ionic saccharide-based detergents is commonly used for membrane protein solubilization.[25] Thus, understanding how this class of detergents affects the resulting charge state distributions (CSDs) during MS-based analysis holds great utility for not only fundamental investigation of the ESI mechanism but also MS-based membrane protein analysis. To gain further insights, correlation between the non-ionic saccharide-based detergents' physicochemical properties and their supercharging behavior (see below) was investigated. The results of this study will be discussed in more detail in Chapter 2.

As mentioned above, a limited addition of small molecules prior to ESI can induce a shift to higher CSDs. This phenomenon was named 'supercharging' by Williams et al.[26] The mechanism behind supercharging has been under heavy debate[19,27–30] and the arguments presented are invariably tied to the ESI mechanism, a topic that is also under an active debate.[19,22,23,27,31–35] Throughout this discussion, the term 'supercharging' has expanded to include a variety of approaches for charge increase during ESI. For instance, Cassou et al. introduced the term 'electrothermal supercharging', in which the temperature elevation of the

emitter tip, the spray potential increase, and the solution ionic strength was utilized to cause denaturation of the protein, which leads to overall increase of CSDs.[36] Flick et al. described charge increase induced by metal cation adduction to the target analytes as supercharging also.[37] To avoid confusion, from here, the definition of 'supercharging' will be restricted to shift to higher CSDs induced by limited chemical agent treatment prior to ESI that does not involve intentional denaturation of the protein or adduct formation.

One of the first researched supercharging agents by Williams et al. was m-nitrobenzyl alcohol (m-NBA), which has a surface tension below water. When m-NBA was added to a methanol/water solution prior to ESI, CSDs of short oligonucleotide anions, which lack defined structure, increased.[38] Lomeli et al. showed that m-NBA can be added to ESI solutions to induce supercharging without harming non-covalent interactions. The increases in charge ranged from 8% for the 690 kDa α7β7β7α7 20S proteasome complex to 48% additional charge for the zinc-bound 29 kDa carbonic anhydrase-II protein.[39] These two studies indicate that m-NBA induced supercharging occurs without inducing denaturation of the protein while lowering the surface tension of the solvent, presenting a clear challenge to CRM in addition to the detergent caused supercharging mentioned above.

Also, Li et al. found that increasing the flowrate of the nESI lowers the average charge state and that a small but reproducible shift to higher CSDs is observed for both proteins and peptides when the diameter of the emitter tip is reduced.[40] This pattern was especially noticeable when the analyte concentration was increased to 20 μM. These results indicate that there are other factors that affect the final charge state in addition to the size and the surface tension of the final nanodroplets that contain the target analyte.

**1.2.3 Evidence for CEM and supercharging agent induced denaturation during ESI**

Both CRM and IEM focus on target analytes that have relatively compact structures. In contrast, CEM aims to explain how denatured, extended proteins ionize during ESI. Ahadi et al. conducted MD simulation studies on hydrophobic and hydrophilic extended polymers and found that during ESI, hydrophobic polymers migrate towards the surface of the nanodroplet and then gets rapidly expelled in a stepwise manner. The ejection was observed when the nanodroplet harbored charge beyond a certain threshold. Hydrophilic chains, in contrast, remained inside the nanodroplet during the entire simulation period. Thus, the data implies that polymer expulsion from the droplet is mediated by an interplay of hydrophobic and electrostatic factors.[17]

Donor et al. further investigated how supercharging affects the ESI of extended polymers. Head-to-tail-linked polyubiquitins (Ubq1−11) were used to determine the size and the charge state scaling laws for unfolded protein ions formed by supercharging while eliminating amino acid composition as a potential confounding factor.[22] Through circular dichroism analysis, the study found that the supercharging condition does not cause protein denaturation in the solution phase. For the native Ubq1−11, the average charge state was found to scale as (mass of the protein)$^{0.55\pm0.01}$, which was in good agreement with the previously published CRM-based prediction.[20] In contrast, the maximum charge state of supercharged Ubq1−11 scaled according to a $(z - 1) \times \ln(z - 1)$ relationship (where z represents the observed max charge state), which was derived analytically by treating the protein as a line segment with uniformly spaced point charges, indicating that supercharging causes denaturation as the ionization is taking place rather than before the ionization. Finally, reduced bovine serum albumin (BSA), β-lactoglobulin, and lysozyme were compared to their disulfide-unreduced counterparts to investigate the internal disulfide bonds' effect on supercharging. Supercharging of reduced BSA, β-lactoglobulin, and lysozyme resulted in a marked increase in observed charge state compared to their unreduced counterparts.[22] These results strongly support a supercharging mechanism

9

whereby unfolding and ejection of protein chains occurs during the electrospray process, that is, a CEM-like model.

## 1.2.4 Three regime view of ESI and supercharging mechanism explanation based on Brønsted acid-base properties

In 2014, Ogorzalek Loo et al. introduced a three-regime view of ESI, with solution, intermediate, and gas phase regimes. The following is a brief summary of the ESI and supercharging mechanism from the publication.[27] During ESI, like-charge repulsion dictates that ions cannot increase charge, regardless of whether they unfold in the gas phase. They can only maintain, reduce, or redistribute it. Thus, the intermediate regime is the only region where ions can gain charge beyond that defined by solution equilibria. Also, in order for wrong-way ionization (solutions sprayed with electrical biases opposite the analyte polarity in solution) to take place,[41] ions need to flip charge in the intermediate regime before being released into the gas phase. Thus, the accumulated observed data supports an intermediate regime in which the target analytes undergoing ESI harbor properties between the solution and the gas phase.

How is charge disbursed from a decomposing droplet in the intermediate regime? Let us consider a singly charged analyte that is being sprayed from an ESI source. If we designate the total charge that a droplet N holds as $Q_{TN}$, the charge that is emitted from the droplet (i.e., the charge acquired by the ion that is being ejected out of the droplet) can be described by the following equation:

$$\Delta Q_{TN} = Q_{TN} - Q_{DN} = Q_{ZN} + Q_{SN} + Q_{BN} + Q_{AHN} \qquad \text{(Equation 2)}$$

where $Q_{DN}$ is the charge that droplet holds, $Q_{ZN}$ is the charge of any ions of fixed charge Z+ (such as $Na^+$ or $K^+$), $Q_{SN}$ is the charge that solvent of the droplet holds, $Q_{BN}$ is the is the charge

that buffer has, and Q$_{AHN}$ is the charge the analyte will acquire. Let us designate the probability of charges being ejected out into the gas phase as k$_g$. Then, equation 2 can be rewritten as:

$$\Delta Q_{TN} = k_{gZ}(Z^+)_N + k_{gS}k_S(S)_N(H^+)_N + k_{gB}k_B(B)_N(H^+)_N + k_{gAH}k_{AH}(A)_N(H^+)_N \quad \text{(Equation 3)}$$

where terms in parentheses represent concentrations at the point where charge is disbursed from a decomposing droplet. From equation 3, the following conclusions can be reached.

- Analyte response will be linear with concentration as long as significantly fewer analyte ions are emitted from droplets than solvent or buffer ions. When the analyte concentration increases to a point at which protonated analyte emission accounts for most of the protons lost by the droplet, the response will be nonlinear.

- The concentration at which analyte response deviates from linearity is independent of the amount of Na+, K+, N(CH ) +, or other fixed charge ions in the solution.

- In cone-jet mode, the ESI current is roughly independent of flow rate. Consequently, the analyte, solvent, and buffer ion signals should also be roughly independent of flow rate.

Focusing on the charge acquired by the analyte, let us now examine how a multiply charged analyte can be represented with the same approach.

$$Q_{AHN} = k_{gAH}k_{AH}(A)_N(H^+)_N \quad \text{(Equation 4)}$$

$$Q_{AH2N} = 2 * k_{gAH2}k_{AH2}k_{AH}(A)_N(H^+)_N^2 \quad \text{(Equation 5)}$$

$$Q_{AH3N} = 3 * k_{gAH3}k_{AH3}k_{AH2}k_{AH}(A)_N(H^+)_N^3 \quad \text{(Equation 6)}$$

Then, equation 3 can be expanded as follows for a triply charged ion.

$$\Delta Q_{TN} = k_{gZ}(Z^+)_N + k_{gS}k_S(S)_N(H^+)_N + k_{gB}k_B(B)_N(H^+)_N + k_{gAH}k_{AH}(A)_N(H^+)_N + 2 *$$

$$k_{gAH2}k_{AH2}k_{AH}(A)_N(H^+)_N^2 + 3 * k_{gAH3}k_{AH3}k_{AH2}k_{AH}(A)_N(H^+)_N^3 \quad \text{(Equation 7)}$$

From equation 7, the following conclusions can be reached.

11

- For the same ESI current and flow rate, spraying from a smaller orifice diameter can yield higher charge states. A smaller diameter emitter opening yields a larger number of droplets smaller in volume. Smaller primary droplets evaporate a larger percentage of their solvent before reaching the Rayleigh limit and they attain higher excess charge concentrations prior to ion ejection.

- If the flow rate of the ESI is increased, droplets' size will generally increase and lead to decrease in average charge state.

- If a multiply charged analyte's concentration increases with little change in ESI current, and a significant portion of the charge emitted by the ESI source is in the form of gas phase protonated analyte ions, analyte CSDs will shift to lower charge state. This is due to higher analyte concentration reducing $(H^+)_N$ significantly, as more protons associate with the added analyte.

Given this information, there are several characteristics that supercharging agents should have to induce charge increase during positive mode ESI. First, the agent should have lower volatility than the solvent. This is because once the analyte is released into the gas phase from the intermediate phase, the extent to which an analyte will lose charge by transferring protons to species with higher gas phase basicity depends on the latter's gas phase population and, hence, volatility. Also, chemical agents that have relatively lower solution and intermediate phase basicity compared to the solvent should be better supercharging agents. These hypotheses were examined by exploring the supercharging capacity of various nitriles and amides. The results of the study will be discussed in more detail in Chapter 2.

## 1.2.5 Fragmentation techniques, and the utility of charge state manipulation

What is the value of increasing charge state during ESI analysis? Higher charge state can improve mass accuracy in high-resolution MS.[23] Also, increasing the charge state of the target analyte leads to greater fragmentation efficiency in electron capture dissociation (ECD),[42] collisionally activated dissociation (CAD),[43] and infrared multiphoton dissociation (IRMPD).[43] Cleavage of the C–N (peptide) bond of proteins/peptides can be achieved by vibrational activation of ionized targe analytes by collisions with neutral gas molecules, known as CAD, or by multiple photon-induced activation and fragmentation, known as IRMPD.[43] Cleavage of the C-C bond along the backbone of the proteins/peptides can be induced by the absorption of one or more high-energy UV photons at specific wavelength (most commonly used are 157, 193, or 266 nm wavelengths) by the target analyte.[45] Finally, cleavage of the N-C bond can be instigated by low energy, ~1 eV, electron capture by multiply charged peptide cations via ECD.[42] Alternatively, N-C bonds can also be cleaved with electrons with higher energy, ~10 eV, which can be useful when the target analyte has low charge. This approach has been named electron induced dissociation (EID).[46] The nomenclature for these fragmentations is shown in **Figure 4**. How these fragmentation approaches can be utilized to gather structural information with MS analysis will be discussed further in section 1.4.3.

Fragmentation-based analysis can be used on target analytes that has been ionized to gain a wide variety of information such as in-depth sequencing, the discovery of novel proteoforms, and the quantification of disease-associated PTMs.[49] However, since native proteins and protein complexes tend to have compact structure, the CSD tend to be lower, which in turn decreases ECD, IRMPD, and CAD efficiency. Supercharging can be used to counter this effect. For instance, Yin et al. studied how supercharged native proteins-ligand complexes can be more efficiently fragmented.[49] Adenylate kinase (AK; or myokinase) belongs to a family of nucleoside monophosphate kinases and maintains the cellular equilibrium concentration of adenylate

nucleotides. AK-ATP complexes were subjected to MS analysis via ESI, and the maximum

charge state observed for AK-ATP increased from 10+ to 18+ upon addition of 200mM

sulfolane, another commonly used supercharging agent.[50] The fragmentation efficiency of AK-

ATP complex increased with charge state for both CAD and ECD approaches and resulted in

many more ligand-bound fragments.
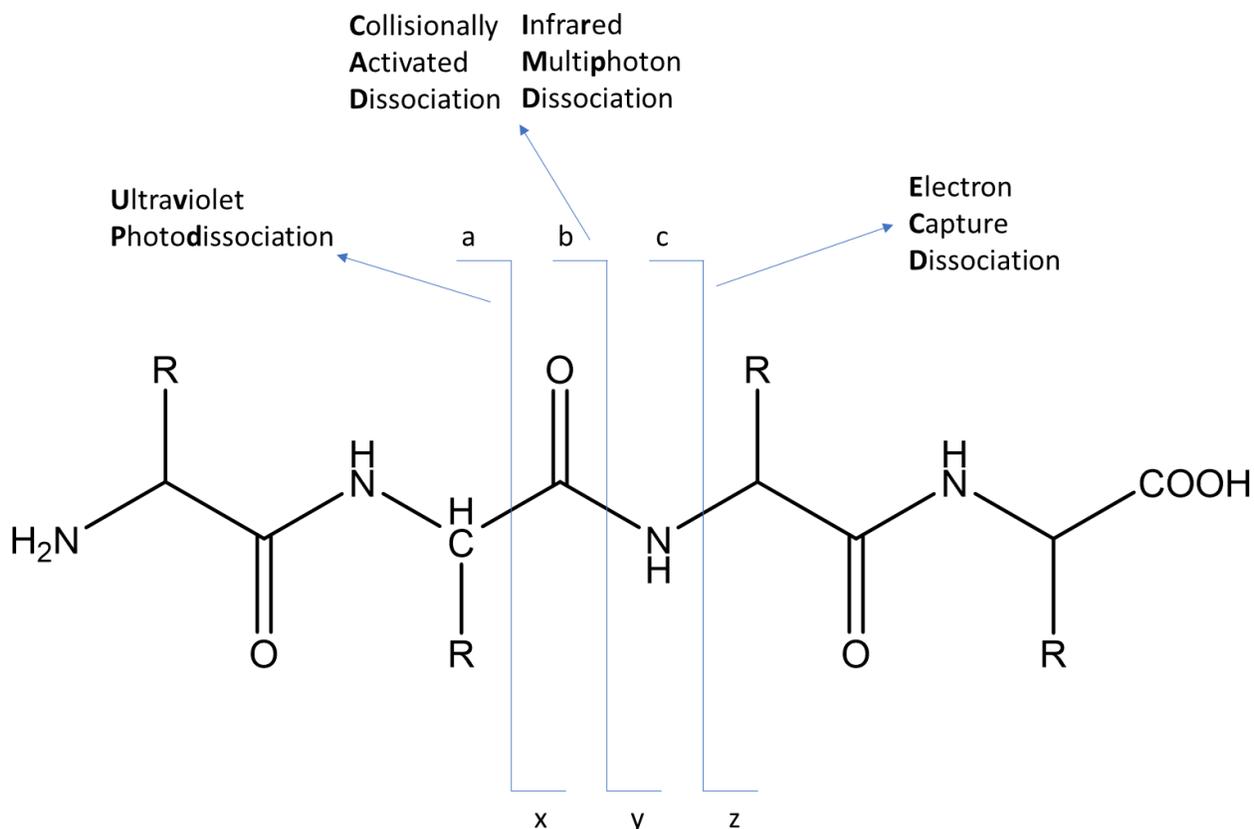


**Figure 4. Fragmentation nomenclature for peptides and proteins: a/x, b/y, and c/z ions according to cleavage of backbone bonds.**

Various fragmentation approaches result in different cleavages of the protein/peptide backbone. The ions that include the N-terminus are denoted as a, b, and c ions and the corresponding ions that include the C-terminus as x, y, and z ions.


In addition, carbonic anhydrase II (CA-II), which is a 29kDa zinc metalloenzyme that catalyzes

the hydration of carbon dioxide to carbonic acid,[51] was fragmented by CAD. A divalent zinc ion

is an essential cofactor for CA-II and the x-ray crystallography structure shows that $Zn^{2+}$ is

coordinated to three histidyl residues (His-94, His-96 and His-119) and a water molecule.[52]

Interestingly, supercharging CA-II did not improve the dissociation efficiency overall but rather resulted in formation of product ions that retain the zinc ligand (e.g., b192 fragment + Zn).[49] This study indicates that supercharging agents can be used to not only enhance fragmentation efficiency but also to facilitate ligand-bound fragment generation.

In another study, Keener et al. showed that stability of membrane protein-nanodisc complexes can be manipulated by supercharging agent treatment.[53] Nanodiscs consist of a nanoscale lipid bilayer surrounded by two membrane scaffold protein (MSP) belts.[54] Nanodiscs are particularly well-suited for native MS due to their monodispersity, homogeneity, and optimal size. However, nanodiscs do not efficiently release naked membrane proteins under conventional native MS conditions and are also not stable enough to detect the intact nanodisc with the membrane protein inside. The study found that supercharging agent, propylene carbonate,[55] stabilizes the membrane protein-nanodisc complex during ESI while glycerol carbonate[56] caused destabilization. Although the cause of this contrasting effect from two supercharging reagents with similar gas phase basicity remain unclear, the utility of supercharging for nanodisc MS analysis was evident.


**1.3 ICP-MS analysis for educational purposes**

Until now the discussion has been focused on ESI, which is an ionization technique suited for biomolecules, especially when multiple charging is desired. However, when the analytical goal is to investigate the atomic composition of the target analyte, an ionization method in which the sample is fully atomized before MS analysis would be preferred. In the following sections, inductively coupled plasma ionization, which atomizes and ionizes the sample via plasma before MS analysis, will be introduced. This particular ionization approach coupled with MS analysis (ICP-MS) is especially useful for quantifying heavy metal contamination in complex samples.

Also, how a protocol for ICP-MS analysis of commercial fish products for mercury contamination detection was developed to teach an analytical chemistry class will be discussed.

## 1.3.1 Ionization technique for atomic composition investigation – inductively coupled plasma ionization

The plasma source of ICP-MS consists of three basic parts, which are the plasma torch, radio frequency (RF) induction coil and the power supply (**Figure 5**). The gas (usually argon) that is used to form the plasma (or plasma gas) is passed between the outer and middle tubes at a flow rate of 12– 17 L/min. A second gas flow (auxiliary gas) passes between the middle tube and the sample injector at 1 L/min and is used to change the position of the base of the plasma relative to the tube and the injector. A third gas flow (nebulizer gas) also at 1 L/min brings the sample, in the form of a fine droplet aerosol, from the sample introduction system and physically punches a channel through the center of the plasma.[57] **Figure 6** describes plasma generation process step by step.[58]
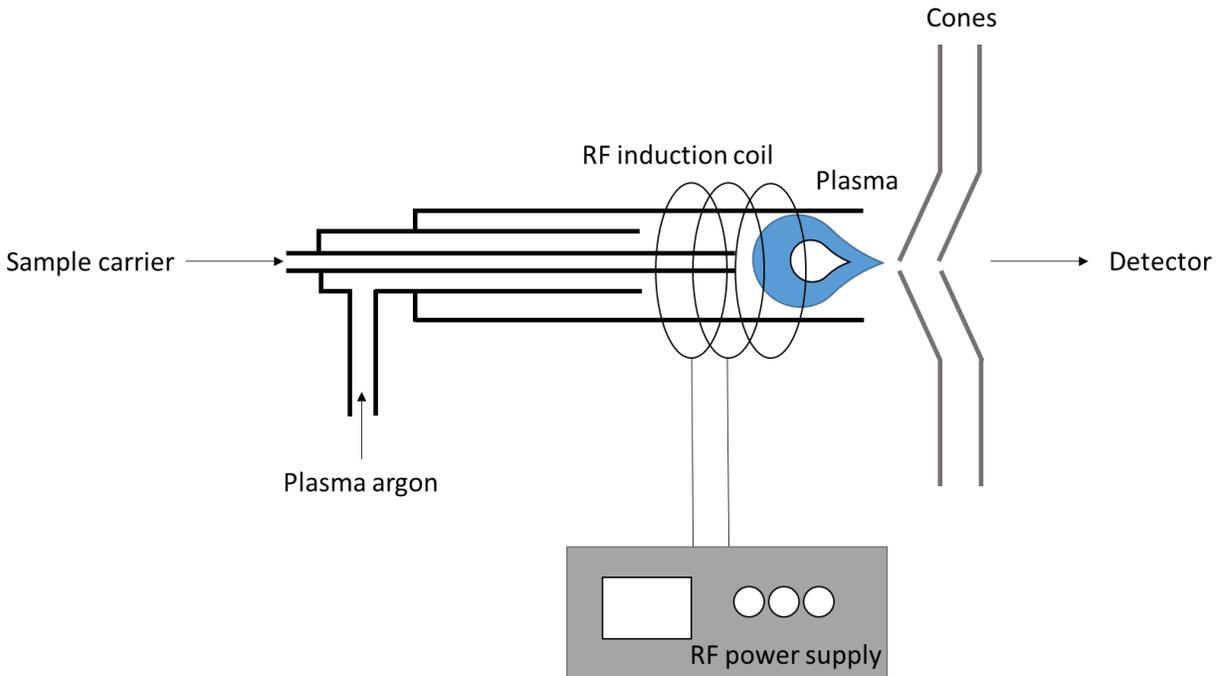
**Figure 5. Schematic view of plasma torch and RF coil.**

The plasma source of ICP-MS consists of the plasma torch, the radio frequency (RF) coil and the power supply.

The ICP discharge is then sustained within the torch and load coil as RF energy is continually transferred to it through the inductive coupling process. The amount of energy required to generate argon ions in this process is approximately 15.8 eV (first ionization potential), which is enough energy to ionize most of the elements in the periodic table.

When the sample droplets encounter the plasma, desolvation takes place first. With the water molecules stripped away, the sample moves further into the plasma and changes first into a gaseous form and then into a ground state atom. The final process of conversion of an atom to an ion is achieved mainly by collisions of energetic argon electrons (and to a lesser extent, by

argon ions) with the ground state atom. The ion then emerges from the plasma and is directed into the interface of the mass spectrometer.[58]
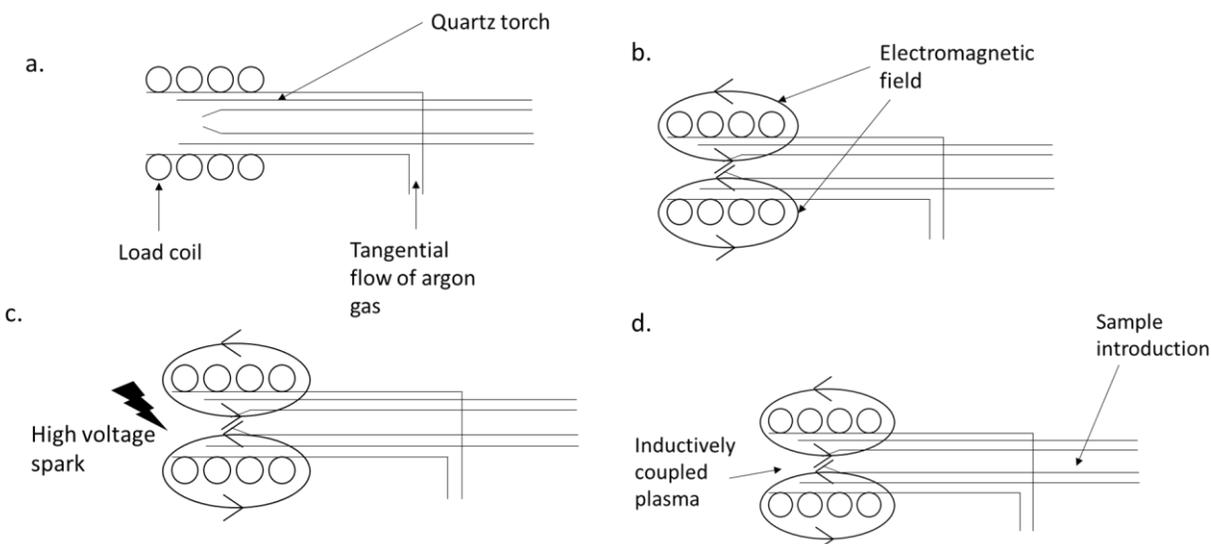


**Figure 6. Schematic of an ICP torch and load coil showing ICP formation process.**

(a) A tangential flow of argon gas is passed between the outer and middle tube of the quartz torch. (b) RF power is applied to the load coil, producing an intense electromagnetic field. (c) A high-voltage spark produces free electrons. (d) Free electrons are accelerated by the RF field, causing collisions and ionization of the argon gas and the ICP is formed at the open end of the quartz torch.

### 1.3.2 General overview of ICP-MS analysis

The sample, which is usually in a liquid form, is pumped at 1 mL/min, with a peristaltic pump into a nebulizer, where it is converted into a fine aerosol with argon gas at about 1 L/min. The fine droplets of the aerosol, which represent only 1–2% of the sample, are separated from larger droplets by means of a spray chamber. The fine aerosol then emerges from the exit tube of the spray chamber and is transported into the plasma torch via a sample injector.

Once the ions are produced in the plasma, they are directed into the mass spectrometer via the interface region, which is maintained at a vacuum of 1–2 Torr with a mechanical roughing pump. This interface region consists of two metallic cones (usually nickel), called the sampler and a

skimmer cone, each with a small orifice (0.6–1.2 mm) to allow the ions to pass through to the ion optics, where they are guided into the mass separation device.

Ions that have been successfully extracted from the interface region are directed into the main vacuum chamber by a series of electrostatic lens, called ion optics. The operating vacuum in this region is maintained at about $10^{-3}$ Torr with a turbomolecular pump. There are many different designs of the ion optical region, but they serve the same function, which is to electrostatically focus the ion beam toward the mass separation device, while stopping photons, particulates, and neutral species from reaching the detector.

The ion beam containing all the analytes and matrix ions exits the ion optics and now passes into the heart of the mass spectrometer—the mass separation device, which is kept at an operating vacuum of approximately $10^{-6}$ Torr with a second turbomolecular pump. An Agilent 8800 QQQ ICP-MS, the instrument that was utilized for this project, has a triple quadrupole system (see below). The goal of the mass separation device of ICP-MS is to allow analyte ions of a particular mass-to-charge ratio through to the detector and to filter out all the nonanalyte, interfering, and matrix ions. Depending on the design of the mass spectrometer, this is either a scanning process, where the ions arrive at the detector in a sequentially manner, or a simultaneous process, where the ions are either sampled or detected at the same time.[58]

### 1.3.3 Quadrupole – A mass filter/selection system for MS analysis

The quadrupole consists of four parallel metal rods. Each opposing rod pair is connected electrically, and an alternating current voltage is applied between one pair of rods and the other at RF. A direct current voltage is then superimposed on the RF voltage, and this causes the ions to adopt an irregular, oscillatory trajectory as they traverse the region bounded by the rods. For a given ratio of voltages, only ions of a specific m/z value reach the detector, and other ions with

unstable trajectories are collisionally annihilated. This allows selected monitoring of ions of a particular m/z value, or a mass spectrum can be obtained by scanning through the m/z range of interest over time.[59]

In a triple quadrupole mass spectrometer, three quadrupoles are placed in tandem (**Figure 7**). The first quadrupole mass filter, Q1, is the primary m/z selector after the sample leaves the ionization source. Any ions that do not have the specific m/z are not allowed to travel to the second quadrupole which serves as a collision cell. Here, ions collide against inert gases such as argon, helium or nitrogen or react with gases such as ammonium or oxygen, depending on the goal of the analysis. The collision cell does not select for specific m/z and all product ions travel into the second quadrupole mass filter, Q2, where m/z selection can occur again.[60,61]


### 1.3.4 Matrix effect on sample stability and the use of the standard reference material during ICP-MS analysis

For the project discussed here, ICP-MS analysis was performed on commercial fish samples. Samples of biological nature are usually digested with an acid or combinations of acids to induce complete mineralization before the analysis. In this context, the matrix is the acid(s) utilized for digestion prior to the analysis (see material and method section of Chapter 4 for more details). Previously published studies have found that trace amounts of mercury can be lost over time possibly due to the ions binding to the walls of the container.[62,63]

Krivan et al. found that at concentrations ranging from 0.05 to 10.0 ng/ml, more than 10% of the dissolved mercury was lost within the first three days and up to 50% of the mercury was lost within nine days after the samples were prepared.[62] Inclusion of hydrochloric acid in the matrix was found to alleviate this problem. Thus, the education protocol was designed to examine the stability of mercury over time by analyzing the samples that have been prepared in two different

matrices, 2% $HNO_3$/0.34% $H_2O_2$ and 2% $HNO_3$/1% $HCl$/0.34% $H_2O_2$, 1 day after the digestion (Day 1) and 8 days after the digestion (Day 8). The results of this comparison will be discussed in Chapter 4.

ICP-MS is a technique primarily used to accurately quantify elements of interest in certain samples. To facilitate method verification among different labs, the National Institute of Standards and Technology (NIST) provides standard reference materials (SRM). SRM is a material whose elemental concentrations has been measured by NIST. Given that the nature of the sample is sufficiently similar, the method used by a particular study is considered verified when the target element's concentration is within ±10% of the value certified by NIST. For example, the mercury concentration of SRM 1947 (Lake Michigan fish tissue)[64] was found to be 0.254 mg/kg (wet-mass basis) by NIST. If the mercury concentration from ICP-MS study conducted here on SRM 1947 arrives between 0.229 and 0.264 mg/kg, the method is considered verified and the analysis results from another sample of similar nature (commercial fish samples from a local supermarket in this case) are assumed to have comparable accuracy.[65-67]

## 1.3.5 Polyatomic interference removal via kinetic energy discrimination

Although HCl inclusion in the matrix is beneficial for long-term mercury storage, Cl⁻ ions can cause polyatomic interference problems. Polyatomic interference is caused by polyatomic species that are isobaric with the target element. For example, $NaCl^+$ (58.44 Da) closely overlaps with $Ni^+$ (58.69 Da) and can hinder accurate determination of Ni concentration. Kinetic energy discrimination (KED)[68] can be used to overcome this problem (**Figure 7**).
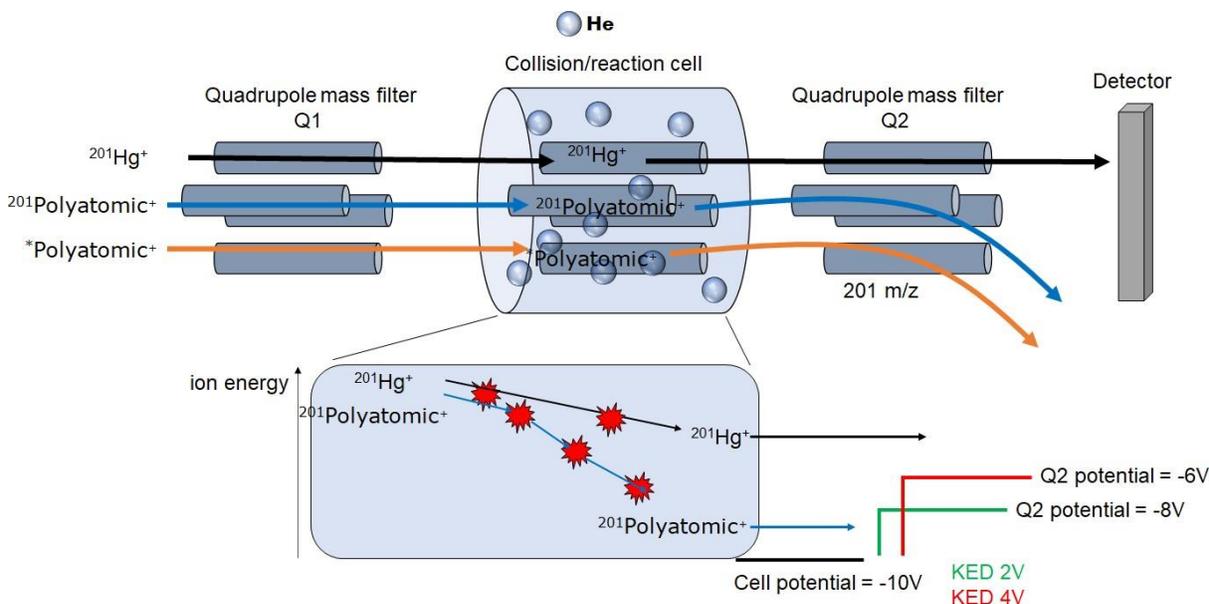


**Figure 7. Polyatomic interference (or isobaric interference) can be attenuated via kinetic energy discrimination.**

KED is defined as the use of a potential energy barrier between the collision/reaction cell and mass analyzer to attenuate isobaric interference from polyatomic species. In a single-quadrupole mode, Quadrupole 1 (Q1) is set to let ions of all m/z pass through. The target element ($^{201}Hg^+$), polyatomic species with the same m/z ($^{201}Polyatomic^+$), and other ions (*$Polyatomic^+$), are introduced into the collision/reaction cell. In no gas mode (not shown), all ions travel unhindered to quadrupole 2 (Q2) and ions with the target m/z value (201 m/z) are selected by Q2. In helium mode, the collision/reaction cell is filled with helium. Polyatomic species that can potentially cause isobaric interference will undergo more collisions with helium compared to $^{201}Hg^+$. The resulting kinetic energy discrepancy can be taken advantage of to filter out $^{201}Polyatomic^+$ by setting up a potential energy barrier, the magnitude of which can be manipulated by adjusting the potential difference between the collision/reaction cell and Q2.

KED takes advantage of the fact that although polyatomic species may have the same mass as the target analyte (isotope $^{201}Hg^+$ in this study), the former may have a larger collision cross section. Thus, as both the polyatomic species and the target analytes travel through the collision

cell, the former will experience greater number of collisions per unit time per unit length, which in turn leads to greater amount of kinetic energy loss. By setting up an appropriate potential barrier before the detector, polyatomic species that have relatively lower kinetic energy can be filtered out. Here, how different levels of potential barrier affect the accuracy of mercury analysis when the collision cell was either filled with helium or no gas was studied with the undergraduate students in an analytical chemistry class. The education goals of the experiment and the students' data will be discussed in detail in Chapter 4.

## 1.4 MS-based protein structural analysis

In the previous sections, fundamentals of the ESI mechanism and how charge state can be manipulated during ESI were discussed. From here, how proteins that were ionized via ESI from a physiological condition can be subjected to MS analysis to gather structural information will be discussed. During native MS (see below), proteins can be subjected to ion mobility to investigate the tertiary structure. In addition, target analyte ions can be fragmented to gather variety of information such as stoichiometry, ligand-induced structural stabilization, and sequence information. However, some protein/protein complexes show significant difference between their solution phase and gas phase structure. Thus, caution is needed when interpreting the gas phase structure as the representation of the solution phase structure.

## 1.4.1 Native MS, a 'snapshot' of solution phase proteins and protein complexes and introduction to ion mobility analysis

Native MS is a MS analysis in which the structure of the target analyte(s) is kept intact if ESI is performed at physiological pH; a protein or protein complex of interest will be in native state prior to ionization. This can result in a 'snapshot' capture of the solution phase structure and

non-covalent interactions.[69] At the most basic level, native MS can provide information on the makeup of protein complexes by measuring molecular weights with high accuracy.

For example, how formylation of the monomer can result in five different proteoforms for an Escherichia coli (E. coli) homotetrameric water channel protein, aquaporin Z (AqpZ) was recently captured.[70] In another example of multimeric complex study, Fan et al. investigated a novel protein called δ protein that is encoded by the RNA genome of the Orsay virus, which is the first virus discovered that infects *Caenorhabditis elegans* naturally. The novel δ protein shares no homology with any other proteins. Thus, the ~420-Å long, pentameric fiber structure, from a combination of electron microscopy (EM), X-ray crystallography, computational and biophysical analyses, was cross verified with native MS approach to confirm that δ protein forms a homopentamer.[71]

Also, target analytes ionized from a physiological condition can be readily subjected to ion-mobility MS (IM-MS)[72] (**Figure 8**). IM-MS utilizes an ion-mobility cell and takes advantage of the fact that while traveling inside a drift tube filled with non-reactive gas such as helium or nitrogen, ions with bigger collision cross section (CCS) will lose greater amount of kinetic energy compared to the smaller ions. Experimental CCS can be compared with theoretical CCS values generated by molecular dynamic (MD) simulations, often using X-ray crystallography and nuclear magnetic resonance (NMR) structures as input for the calculations.[73]

IM-MS has shown great promise as an intact protein separation and analysis methodology to probe higher order structural elements. For example, Bagal et al. used IM-MS to detect disulfide heterogeneity in large (150 kDa) intact immunoglobulin G (IgG) family antibodies.[74] IgG1 and IgG4 hinge core sequences are very similar with two cysteines on each heavy chain involved in inter heavy chain connection, whereas IgG2 is unique in presenting four cysteine residues in the hinge region, notably two consecutive residues, Cys-232 and Cys-233. Two to three gas-phase conformers were observed by ion mobility for IgG2 antibodies and the analysis of redox refolded IgG2s as well as an IgG2 with a Cys to Ser single point mutation.
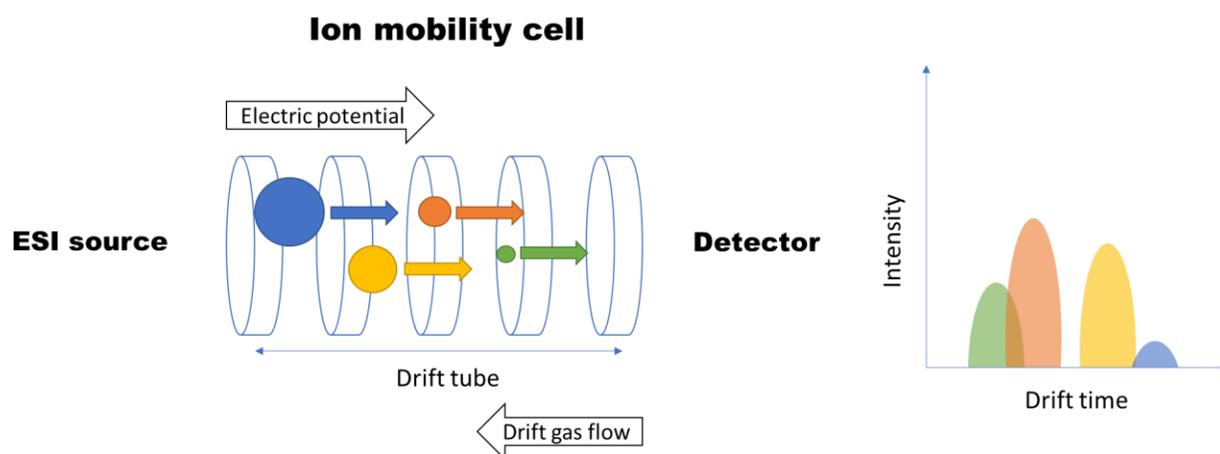


**Figure 8. Schematic representation of core working principle of ion mobility analysis.**

Ionized samples emerging from an ESI source will experience collisions with the inert gas inside the drift tube as they travel down the electric potential. Ions are separated based on their charge and shape: compact ions travel faster than extended ions carrying the same number of charges, whilst ions with a high number of charges travel faster than ions carrying a lower number of charges derived from the same precursor molecules.[72,73]

Although ions that have relatively low charge state are in good agreement with data calculated for the same species based on X-ray crystallography or NMR spectroscopy data, gas-phase CCS values are typically slightly lower than those calculated for solid-phase crystal structures.[73] This discrepancy becomes more evident for larger proteins and/or have extended regions that are prone to gas phase compaction.[75] Thus, interpretation of gas phase structure must be done with caution.

**1.4.2 The correlation between solution phase and gas phase structure**

In 1997, Loo stated there are three camps of opinion on native protein structure upon transition into the gas phase: believers, nonbelievers, and undecided.[76] This cautious approach turned out to be insightful because as structural MS approaches developed, both successful cases and studies that showed gas phase collapse emerged. In this section, a brief review of the studies that support the retention of the native structure during gas phase transition and that report gas phase collapse will be introduced.

Most of the structural information available currently on proteins or protein complexes stem from X-ray crystallography or cryo-electron microscopy (cryo-EM) experiments in which averaging over many molecules assembled into a crystal or over a large ensemble selected from low signal-to noise ratio electron micrographs are used, respectively.[77] However, averaging inevitably takes away distinct conformation information, which is important to understand biomolecules' function in a physiological context. Thus, if a specific proteoform in a specific conformation can be isolated and then imaged, that would provide the necessary details to truly investigate structural change needed for a certain protein/protein complex to fulfill its role.

Compelling evidence for native structure retention has been provided by 'soft-landing' experiments, in which gas phase ions of large protein complexes are mass-selected and then gently decelerated and collected on a grid. Subsequent EM imaging then demonstrated preservation of native-like structures throughout the process of ionization, dehydration, and soft landing.[78,79]

In a study conducted by Longchamp et al., native cytochrome C (CytC), bovine serum albumin (BSA), and hemoglobin (HG) ion beams were generated by electrospray ionization and mass filtering (i.e., quadrupole-based m/z selection).[78] With the soft-landing electrospray ion beam deposition (ES-IBD) system (**Figure 9**), the charge states 5-7+ were selected for CytC, 15-18+

for BSA, and 16-17+ for HG, based on the previous publications that investigated the charge state and the native structure of the proteins.[80-82] The ions were then decelerated to a very low kinetic energy of 2–5 eV per charge, which ensured retention of the native state upon deposition onto ultraclean freestanding graphene. For CytC and BSA, the agreement between the low-energy electron images and the atomic models was almost perfect. For HG, specific orientation of the molecule at the time of the deposition caused disagreement between the averaged atomic model and current data. The result of the study indicates that native structure of proteins can withstand transition into gas phase. In addition, the study hints that if the IM-MS approach can be incorporated into the ES-IBD workflow, proteoform purification based both on mass and conformation could be possible prior to time and cost intensive high-resolution structure study.
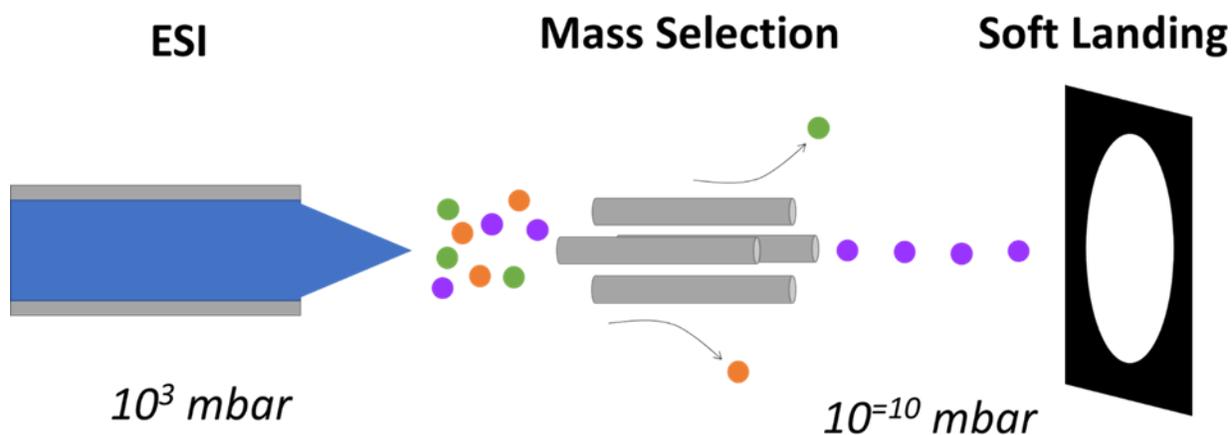


**Figure 9. Schematic representation of soft-landing electrospray ion beam deposition (ES-IBD) system.**

Prior to low energy electron holography analysis for structural characterization,[78] the samples can be purified in the gas phase with the quadrupole. After the mass selection, the specific target analyte of a certain mass can be deposited on to freestanding graphene. During this process, ultrahigh vacuum is maintained. The fact that CytC and BSA structures analyzed with this approach align closely with previously known structure suggests that transition into the gas phase, at least in some cases, does not result in loss of native structure.

Seo et al. used IM-MS in combination with infrared (IR) spectroscopy to probe the tertiary and secondary structure of myoglobin and β-lactoglobulin, respectively.[83] As expected, the collision

cross-sections increase with increasing charge. At low charge states, compact ions are observed that have cross-sections consistent with those expected for native structures. IR spectroscopy analysis of 8+ charge state of both myoglobin and β-lactoglobulin showed that α helical structure and β sheet structure of the proteins were retained in the gas phase, respectively.

However, for non-globular proteins and biomolecules, significant gas phase collapse can take place. Devine et al. conducted ESI-IM-MS analysis on an IgG1 monoclonal antibody, mAb1. [75] The experimentally estimated CCS value of the lowest charge state (68.2 nm$^2$) was significantly lower (32.4%) than the computationally determined CCS (101 nm$^2$) based on the published structure (PDB = 1IGY).[84] Another non-globular protein, a constructed concatamer of five mutant I27 domains ((I27)5), which was originally used for mechanical unfolding studies,[85] was also subjected to ESI-IM-MS study. MD simulation was used to calculate the theoretical CCS of I275. Similar to the mAb1 results, the measured CCS for (I27)5 (39.8 nm$^2$) was lower than both the modeled value predicted for the native structure (63.1 nm$^2$) and the MD simulation end point (49.4 nm$^2$), indicating that significant gas phase collapse was taking place. This study highlights how caution is needed for gas phase-based structural investigation of target analytes that have non-globular structure.


**1.4.3 Higher-order structural characterization of proteins and complexes by top-down CAD MS**

As briefly mentioned in section 1.2.5, various fragmentation approaches can be applied to target analytes that have been ionized through the native MS approach. In this section, how CAD can be utilized for top-down MS analysis will be discussed in more detail. In contrast to the conventional bottom-up MS approach, in which proteins are extracted, chemically or

enzymatically digested, separated by liquid chromatography (LC), ionized via ESI, and analyzed by MS, top-down MS forgoes the digestion process and directly fragments the ionized targets.

For native top-down MS analysis, the solution structure of proteins and protein complexes are preserved during the transfer to the gas phase and then the target analytes are fragmented. Why do we need to fragment proteins and protein complexes after native ESI? There are a number of limitations that native MS has a standalone approach.[48]

- Native MS without breaking up the complexes only provides limited information regarding quaternary structure and is largely blind to subunit connectivity and location of ligand binding within the complexes.

- For unknown complexes, intact mass alone is not enough for determining stoichiometry and composition.

- Proteins' average masses might shift slightly due to natural variations in isotopic abundance, an effect which is able to cause mass shifts greater than the accuracy of modern high-end mass spectrometers. This becomes especially problematic for larger proteins.

- Gentle tuning conditions used to maintain structural integrity of noncovalent complexes can result in insufficient desolvation, peak broadening, and increased uncertainty in mass determination.

- For fragile complexes, the achievable mass resolution may be too low for precisely determining the binding of small ligands.

- Disruption of complexes by partial denaturation in the solution phase with chemical denaturant may lose connectivity information. Also, the protocol has to be refined for each target analyte and can fail for proteins that are resistant to mild denaturants or precipitate easily upon denaturation.

In contrast, CAD analysis can be done to fragment the protein without special optimization for each target analyte (although the amount of energy needed for dissociation will differ for each case). When CAD is applied to a multimer complex, monomer ejection is routinely observed.[86–88] The ejected monomer carries disproportionately higher charge compared to the multimer, which facilitates further CAD analysis. This phenomenon can be exploited to perform complex-down MS analysis, which is a type of top-down MS approach that sequences ejected proteins from native complexes.

Gault et al. applied complex-down MS analysis to a multimeric membrane protein, the outer mitochondrial membrane translocator protein complex, to identify the presence of lipid binding to the protein complex.[89] First the membrane protein-micelle complex was exposed to collisional activation to eject the membrane protein complex. Then a second round of activation was applied to induce monomer ejection. Finally, the monomers were isolated with a quadrupole and then subjected to a third round of CAD to identify the bound lipid, which was found to be phosphatidylethanolamine 34:1. This lipid fit well with the previously established crystal structure.

Similar approaches were utilized to analyze the ammonium channel (AmtB) of *E. coli* bound to phosphatidylglycerol (PG) and cardiolipin (CL). AmtB is a homotrimer channel protein that is necessary to maintain rapid cell growth at low ambient ammonium concentration.[90] Laganowsky et al. found that PG and CL cause significant gas phase structure stabilization for AmtB through ESI-IM-MS analysis.[7] The study conducted collision induced unfolding (CIU) experiments in which they measured the collision energy necessary to induce significant unfolding of AmtB or AmtB-lipid complexes. Both AmtB-PG and AmtB-CL complexes were able to withstand significantly more collision energy before being denaturing compared to apo-AmtB, indicating that these two lipids induce structure stabilization. In Chapter 3, how AmtB-PG and AmtB-CL

were subjected to CAD study to study structure stabilization by lipids and to find lipid-bound fragments will be discussed in more detail.

CAD analysis of multimer complex can also yield stoichiometry information. Aquilina et al. investigated the quaternary structure of the polydisperse mammalian chaperone αB-crystallin, a member of the small heat-shock protein family, with ESI-CAD-MS.[91] The intact assemblies gave rise to mass spectra that are complicated by the overlapping of charge states from the different constituent oligomers. In contrast, after CAD, the spectra reveal a distribution, primarily of oligomers containing 24–33 subunits, the relative populations of which were quantified, to reveal a dominant species being composed of 28 subunits.

### 1.4.4 Inclusion of internal fragments for top-down MS analysis

Up until now, the discussion has been restricted to terminal fragments, i.e., fragments that include the N-terminus (b fragments for CAD) or C-terminus (y fragments for CAD) (see **Figure 4**). However, top-down MS analysis can result in internal fragments, in which fragmentation of initially formed fragments take place and both the N-terminus and C-terminus are lost.

To assign a fragment, mass spectra from a top-down MS experiment must be deconvoluted, i.e., changed from m/z spectra to m spectra. A number of deconvolution algorithms [92–94] exist and the resulting deconvoluted mass list from the raw data is matched against a theoretical terminal fragment list to generate the matched fragment list.

As the number of theoretical internal products that can be generated is significantly greater than the number of theoretical terminal fragments that can be generated, the probability of matching a mass spectral signal to an internal fragment is much larger than the probability of matching it to a terminal fragment. The ambiguity of assigning internal fragments likely scales significantly as the size of the protein increases, thus increasing the false discovery rate.[95] Because of this

issue, only limited number of studies have included internal fragments into the consideration during top-down MS analysis.[95-101] However, including internal fragments can significantly enhance the sequence coverage from a top-down experiment.

To investigate the formation of terminal and internal fragments from different activation methods, Zenaidee et al. analyzed denatured myoglobin that was electrosprayed from acidified solutions and fragmented by CAD, ECD, EID, and UVPD (see section 1.2.5 and **Figure 4**).[95] Using Clips-MS, a computer algorithm that can assign internal fragments generated from top-down MS experiments,[101] how sequence coverage improves once internal fragments are included in the analysis was investigated. For CAD, the inclusion of internal fragments enhanced the sequence information from ~30% to ~75%. Similar trends were observed for ECD (~55% to ~82%), EID (~57% to ~90%), and UVPD (~62% to ~88%). For the membrane protein study discussed here, Clips-MS was utilized for native top-down CAD analysis of AmtB, AmtB-PG complex, and AmtB-CL complex. Clips-MS was also used to detect internal fragments bound with lipids. The study will be discussed in more detail in the following sections and Chapter 3.

**1.4.5 UHMR – a mass spectrometer optimized for native and native top-down MS analysis of large complexes**

Target analytes ionized from a non-denaturing condition tend to have globular structure and low charge state. Thus, to perform native MS and native top-down MS effectively, an instrument that can transmit large analyte ions without compromising their structure while being able to handle detection at high m/z region is necessary. Thermo Scientific's Ultrahigh mass range (UHMR) Orbitrap is one of the instruments that can fulfill this role. The following is a brief summary of the working principle of an Orbitrap mass spectrometer.[102]

During an Orbitrap mass spectrometer operation, ions generated from an ESI source are guided by the quadrupoles to the mass analyzer. The third quadrupole, pressurized to less than $10^{-3}$ Torr with collision gas, acts as an ion accumulator; ion/neutral collisions slow the ions and cause them to pool in an axial potential well at the end of the quadrupole. Then, the ions are injected into the mass analyzer. The Orbitrap's mass analyzer consist of an outer barrel-like electrode which is coaxially positioned with the inner spindlelike electrode. Once injected, the ions will oscillate between the two electrodes in a helix-like path. The oscillation frequency is completely independent of energy and of the spatial spread of the ions and is directly related to the ions' m/z value by the following equation:

$$\omega = \sqrt{(^z/_m) \cdot k} \qquad \text{(Equation 8)}$$

in which $\omega$ is the oscillation frequency along the z-axis and k is a constant. Thus, once the image current data, which is a time domain data generated by the ions' oscillation, is converted into a frequency domain data by Fourier transformation, the mass spectrum can be acquired by a simple arithmetic conversion.

The UHMR Orbitrap system has the capacity to isolate ions up to m/z 25,000 for subsequent fragmentation. Also, in-source trapping allows efficient removal of detergent micelles for accurate native MS native top-down analysis of the AmtB membrane protein complex.[103] Thus, for AmtB and AmtB-lipid complex native top-down analysis discussed here, UHMR was used as the main instrument. The details of this study will be discussed in the next section and Chapter 3.

**1.5 Challenges of membrane protein structural analysis and current status of native MS-based analysis of membrane proteins**

Membrane proteins' interaction with their cognate ligands are responsible for various physiological phenomena such as cell-cell interaction, signal transduction, solute transportation, and energy conversion etc.[104] Membrane proteins account for 60% of known druggable targets in the cell.[105] However, they represent only 4.6% of the PDB despite constituting between 20% to 30% of proteome.[106] They are a difficult class of proteins for structural analysis and high-resolution studies face many challenges.

- Recombinant expression and purification of sufficient amounts of membrane proteins is difficult.

- Since they are hydrophobic proteins, crystallography and EM analysis require the appropriate choice of detergent micelles, lipid cubic phase, or membrane mimetic.[107]

- NMR studies tend to have difficulties due to the detergent micelle formation around the membrane proteins, which effectively doubles the size of the target analyte.[108]

- Detergent micelle's innate heterogeneity makes classification and alignment difficult for accurate image averaging for cryo-EM.[109]

- Membrane proteins are often flexible and unstable, which hinders effective crystallization.[110]

Structural study of membrane protein is both important and necessary but presents unique challenges. Thus, another approach that can overcome these issues and complement conventional structure analysis techniques is necessary. Native and native top-down MS analysis can fulfill this role. In the following sections, seminal studies of membrane protein MS will be introduced. Then, how the current dissertation study fits into the context of membrane protein MS history will be briefly discussed.

**1.5.1 Brief history of native and native top-down analysis of membrane proteins**

In 1998, Whitelegge et al. published the successful ESI-MS analysis of intact membrane proteins.[111] In this seminal study, bacteriorhodopsin, and its apoprotein from Halobacterium, and the D1 and D2 reaction-center subunits from spinach thylakoids were measured within 0.01% of their respective calculated theoretical mass value. Especially for bacteriorhodopsin, which was first subjected to reversed-phase HPLC in aqueous formic acid/isopropanol prior to ESI-MS analysis, chromophore retainment was observed, indicating that structurally intact membrane protein analysis was possible via ESI-MS. The study also hinted why crystallization of plant thylakoid PS2 was difficult. Upon ESI-MS analysis, D2 subunit's heterogeneity was apparent, which was likely to hinder crystal formation. Currently, how native MS can be incorporated into the single-particle cryo-EM workflow by providing a rapid assessment of sample stability and homogeneity has been published by Olinares et al.[112]

In 2002, Sobott et al. introduced a quadrupole time-of-flight mass spectrometer optimized for the transmission and analysis of large macromolecular assemblies.[113] As mentioned above, native MS tend to produce ions of globular nature that leads to low charge state and high m/z. The instrument featured a carefully moderated pressure gradient that provides gentle transmission of ions and modified quadrupoles that were able to mass filter to as high as m/z 22,000. A few years later, the Orbitrap system was introduced as well (see section 1.4.5). These instruments opened the door for native and native top-down MS of membrane protein analysis.

Sharon et al. used the modified quadrupole time-of-flight mass spectrometer to capture intact micelles, reverse micelles, and reverse micelles encapsulating myoglobin in 2007.[114] In the same year, Zabrouskov et al. performed the first true high-res top-down ESI-MS of an integral membrane protein.[115] A year later, the heteromeric adenosine 5´-triphosphate (ATP)–binding cassette transporter BtuC2D2, protected within a n-dodecyl-β-D-maltoside micelle, was detected also with the quadrupole time-of-flight mass spectrometer.[116] In 2010, Ryan et al. introduced a pivotal study in which a top-down high-resolution Fourier transform mass

spectrometry with CAD approach was used to analyze PTMs of bacteriorhodopsin holoprotein. N-terminal propeptide cleavage and conversion of N- terminal Gln-14 to pyrrolidone carboxylic acid and C-terminal removal of Asp-262 were detected.[117]

These studies showed that transition to gas phase can be withstood by membrane protein complexes when they are protected by a micellar structure. In addition, it was clear that both the instrumentation and the sample preparation methods for native top-down analysis of membrane proteins have matured to a point where fragmentation approaches yielded meaningful biological information.

In 2013, a detailed protocol paper, which outlines how to screen detergents for membrane protein MS studies and how to prepare lipids for membrane protein-lipid binding experiment was published.[118] This seminal work paved the way for the "explosion" of membrane protein MS publications in the following years.

- 2014 – IM-MS study of mechanosensitive channel of large conductance (MscL) from Mycobacterium tuberculosis and AqpZ and AmtB from *E. coli* with various lipid interactions reveals how lipids play a crucial role in structural stability of membrane proteins.[7]

- 2016 - Gault et al., using the Orbitrap mass spectrometer, investigates a wide range of membrane proteins' (a monomeric G-protein coupled receptor (CCR5); dimeric glycan transporter (semiSWEET); trimeric channels, the outer membrane porin OmpF and ammonia transporter AmtB; and pentameric ligand-gated ion channel (ELIC)) interaction with small molecules. The study shows the capacity of native MS to capture solution phase interactions for membrane proteins with varying size.[89]

- 2017 – Gupta et al. discovers how oligomerization of LeuT, a sodium symporter that transports small aliphatic amino acids across the bacterial inner membrane, is mediated by interfacial lipids.[119]

- 2017 - Yen et al. publishes a landmark study in which the human purinergic receptor P2Y1R, a class A G-protein coupled receptor (GPCR) is captured with native MS. Both the endogenous ligand-bound form and drug-bound form are detected with high mass accuracy.[120]

- 2018 – Susa et al. finds out that with the submicrometer nano-ESI emitter tips, resolved charge-state distributions of membrane protein ions can be obtained from a solution containing 150 mM NaCl, 25 mM Tris-HCl and 1.1% octyl glucoside. The study highlights how emitter tips' diameter plays a crucial role in ionization efficiency.[121]

- 2018 – Lippens et al. identifies an optimized condition for Fourier transform-ion cyclotron resonance mass spectrometer-based analysis of AqpZ. The study achieves high enough resolution to detect five distinct proteoforms of the protein.[70]

## 1.5.2. Native top-down analysis of AmtB-PG and AmtB-CL complexes

Although significant progress has been made for native MS analysis for membrane proteins, there is a distinct lack of fragmentation-based analysis of membrane protein-lipid complexes. Thus, the dissertation study focused on native top-down analysis of AmtB bound to two different lipids, cardiolipin (CL) and phosphatidylglycerol (PG).

Ammonium is the favored nitrogen source for *E. coli* and in conditions of nitrogen limitation, ammonium uptake is facilitated by AmtB.[90] They are almost ubiquitous among archaea, eubacteria, fungi, and plants, whereas in animals they are represented by the closely related Rhesus family.[122] The channel is a homotrimer and mature *E. coli* AmtB has 11 transmembrane helices with an $N_{out}$ and $C_{in}$ topology that appears to characterize all members of the Amt family. The Amt structures suggest that substrate conductance occurs through a narrow, mainly hydrophobic pore located at the center of each monomer and containing two highly conserved

histidines (H168 and H318), each of which is essential for conductance of the ammonium analogue methylammonium.[122]

Laganowsky et al. observed that lipid binding to AmtB results in a range of stabilizing effects. Among the lipids tested, addition of CL or PG resulted in striking increase in stabilization with cumulative binding of PG increasing protein stability linearly.[7] The crystal structure of the AmtB-PG complex was subsequently screened in this study and hydrogen bonds forming between the phospho-headgroup and K84 was observed along with conformational change of residues 70–81, induced by binding PG.

Native top-down analysis of AmtB-PG and AmtB-CL complexes was conducted to compare the fragmentation pattern of AmtB and AmtB-PG/CL. As observed previously, PG showed more robust binding. AmtB bound up to 2 PG molecules and AmtB bound to 1 CL molecule could readily be isolated and then subjected CAD fragmentation. The fragmentation pattern and internal fragments bound to lipids enabled detailed localization of the structure stabilization effect that PG and CL confers on the AmtB. The study shows how lipid interaction can be maintained in the gas phase and how fragmentation pattern can be used to infer structural effect of lipid interaction. Also, how inclusion of internal fragments into the native top-down MS workflow can provide richer information from a single fragmentation experiment is highlighted.


## 1.6 Summary

MS analysis has emerged as a versatile tool for proteoform characterization. The development of ESI, which is a soft ionization technique that confers multiple charge to the target analyte, has made intact proteoform analysis more approachable. Although ESI is globally used throughout the MS community, the exact mechanism behind the ionization is still an area of ongoing research.

Three major models have been proposed. CRM for relatively large, globular proteins with low charge density, IEM for relatively small, compact proteins with high charge density, and CEM for elongated, denatured protein whose hydrophobic residues are exposed. It is likely that a mixture of these mechanism takes place during ESI of complex samples.

Although CRM-based charge predictions tend to align closely with observed charge states when a relatively simple solvent is utilized, the model falls apart when supercharging agents are included prior to the ESI analysis. Supercharging agents are small molecules with lower volatility and weaker basicity compared to the solvent that can increase the CSDs of the target analytes without causing denaturation. Also, the fact the emitter tip's orifice diameter, and the flow rate of the ESI affect CSDs directly contradicts CRM.

The three-regime view of ESI can account for these apparent contradictions. The core principle of the three-regime view is that there exists an intermediate regime between solution phase and the gas phase at which the volatility and the basicity of the components of the nanodroplet will dictate over the charge distribution during the ion emission from the nanodroplet. To investigate this hypothesis further, physicochemical properties of the non-ionic saccharide-based detergents, a class of detergents commonly used for membrane protein analysis which shows supercharging behavior, along with various amides and nitriles were investigated for their charge manipulating behavior (Chapter 2).

When ESI is performed at physiological pH, a protein or protein complex of interest will be in its native state prior to ionization. When sufficiently gentle conditions are used for the transmission of the ions, the gas phase structure can be interpreted as a snapshot of the solution phase structure, although caution must be taken for non-globular biomolecules. Native MS can be used for stoichiometry investigation, novel proteoforms identification, and ligand binding screening. In addition, native MS paired with IM analysis can provide global structure information such as stabilization effect conferred by a specific ligand interaction. However, for

more specific information such as subunit connectivity and location of ligand binding within the complexes, a fragmentation approach is necessary.

By combining native MS with fragmentation approaches, native top-down MS analysis can be performed to gain structural information. This approach can be especially useful for samples that are difficult to analyze with conventional high-resolution structural analysis techniques such as membrane proteins. The advent of instruments capable of gentle transmission of ions at high m/z and refinement of the protocol for membrane protein preparation for native top-down MS analysis is also helping further development of the technique.

In addition, incorporation of internal fragments (fragments in which both the N-terminus and the C-terminus are lost) into the analysis, with the development of a new software capable of detecting internal fragments, is helping researchers achieve even higher sequence coverage and more robust detection of ligand-bound fragments. Combining all these advances together, AmtB, which is a multimeric channel protein that is known to greatly stabilized by two lipids, CL, and PG, was analyzed with native top-down MS analysis (Chapter 3).

When the analytical goal is to preserve the target analyte's original structure and confer multiple charges in the process, native MS via ESI is preferred. However, when the goal is to quantify specific element of a certain sample (for instance, heavy metal contamination level), ionization technique that atomizes the sample is necessary. ICP-MS uses plasma generated by high RF energy, which is continually transferred to the load coil through the inductive coupling process, to fully atomize the sample during the ionization. However, some polyatomic species that are isobaric to the target analyte might survive the ionization process. In this case, the polyatomic interference can be attenuated by KED, in which the relative kinetic energy loss difference between the target element and the isobaric polyatomic species as they travel through the collision cell, is taken advantage of.

Another problem that can occur is the sample loss due to the matrix utilized for the sample digestion and storage. For example, as much as half of the mercury can be lost after nine days of storage in a pure nitric acid matrix. Thus, careful modulation of matrix composition and KED application is necessary for a successful mercury quantitation with ICP-MS.

To verify the accuracy of a certain method utilized by a lab, SRMs of similar nature to the target sample can be used. NIST provides a wide variety of SRMs of which elemental composition has been quantified. If the observed concentration of a certain element of interest arrives within ±10% of the value certified by NIST, the method is considered verified. In an effort to teach undergraduate students these concepts necessary for accurate quantitation of mercury in commercial fish samples, an education paper was devised that guides students through ICP-MS experimental design (Chapter 4).

# References

1. Pertea, M. & Salzberg, S. L. Between a chicken and a grape: estimating the number of human genes. *Genome Biol.* **11,** 206 (2010).

2. Aebersold, R. et al. How many human proteoforms are there? *Nat. Chem. Biol.* **14**, 206–214 (2018).

3. Smith, L. M. & Kelleher, N. L. Proteoform: A single term describing protein complexity. *Nat. Methods.* **10**, 186–187 (2013).

4. Emmett, M. R. & Caprioli, R. M. Micro-electrospray mass spectrometry: Ultra-high-sensitivity analysis of peptides and proteins. *J. Am. Soc. Mass Spectrom. 57.* **5**, 605–613 (1994).

5. Brown, K. A. et al. Proteomic Analysis of the Functional Inward Rectifier Potassium Channel (Kir) 2.1 Reveals Several Novel Phosphorylation Sites. *Biochemistry. 60.* **44**, 3292–3301 (2021).

6. Yen, H. Y. et al. Ligand binding to a G protein–coupled receptor captured in a mass spectrometer. *Sci. Adv.* **3**, (2017).

7. Laganowsky, A. et al. Membrane proteins bind lipids selectively to modulate their structure and function. *Nature.* **510**, 172–175 (2014).

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

9. Karas, M. & Hillenkamp, F. Laser Desorption Ionization of Proteins with Molecular Masses Exceeding 10 000 Daltons. *Analytical Chemistry vol.* **60**, 2299–2301 (1988).

10. Tanaka, K. et al. Protein and polymer analyses up to m/z 100 000 by laser ionization time-of-flight mass spectrometry. *Rapid Commun. Mass Spectrom.* **2**, 151–153 (1988).

11. Andrej Shevchenko, Matthias Wilm, Ole Vorm, and & Mann*, M. Mass Spectrometric Sequencing of Proteins from Silver-Stained Polyacrylamide Gels. *Anal. Chem.* **68**, 850– 858 (1996).

12. Susa, A. C., Xia, Z. & Williams, E. R. Small Emitter Tips for Native Mass Spectrometry of Protein and Protein Complexes from Nonvolatile Buffers That Mimic the Intracellular Environment. *Anal. Chem.* **89**, 3116–3122 (2017).

13. Y. L. & Cole*, R. B. Shifts in Peptide and Protein Charge State Distributions with Varying Spray Tip Orifice Diameter in Nanoelectrospray Fourier Transform Ion Cyclotron Resonance *Mass Spectrometry. Anal. Chem.* **75**, 5739–5746 (2003).

14. Konermann, L., Metwally, H., Duez, Q. & Peters, I. Charging and supercharging of proteins for mass spectrometry: Recent insights into the mechanisms of electrospray ionization. *Analyst vol.* **144**, 6157–6171 (2019).

15. Dole, M. et al. Molecular Beams of Macroions. *J. Chem. Phys.* **49**, 2240–2249 (1968).

16. Iribarne, J. V. & Thomson, B. A. On the evaporation of small ions from charged droplets. *J. Chem. Phys.* **64**, 2287–2294 (1976).

17. Ahadi, E. & Konermann, L. Modeling the behavior of coarse-grained polymer chains in charged water droplets: Implications for the mechanism of electrospray ionization. *J. Phys. Chem. B.* **116**, 104–112 (2012).

18. Konermann, L., Rodriguez, A. D. & Liu, J. On the formation of highly charged gaseous ions from unfolded proteins by electrospray ionization. *Anal. Chem.* **84**, 6798–6804 (2012).

19. Iavarone, A. T. & Williams, E. R. Mechanism of charging and supercharging molecules in electrospray ionization. *J. Am. Chem. Soc.* **125**, 2319–2327 (2003).

20. Fernandez De La Mora, J. Electrospray ionization of large multiply charged species proceeds via Dole's charged residue mechanism. *Anal. Chim. Acta.* **406**, 93–104 (2000).

21. Reading, E. et al. The Role of the Detergent Micelle in Preserving the Structure of Membrane Proteins in the Gas Phase. *Angew. Chemie Int. Ed.* **54**, 4577–4581 (2015).

22. Aliyari, E. & Konermann, L. Formation of Gaseous Proteins via the Ion Evaporation Model (IEM) in Electrospray Mass Spectrometry. *Anal. Chem.* **92**, 10807–10814 (2020).

23. Donor, M. T., Ewing, S. A., Zenaidee, M. A., Donald, W. A. & Prell, J. S. Extended Protein Ions Are Formed by the Chain Ejection Model in Chemical Supercharging Electrospray Ionization. *Anal. Chem.* **89**, 5107–5114 (2017).

24. Loo, R. R. O., Dales, N. & Andrews, P. C. The Effect of Detergents on Proteins Analyzed by Electrospray Ionization. *Protein and Peptide Analysis by Mass Spectrometry.* Vol **61**, 141–160 (Humana Press, 1996).

25. Hardy, D., Desuzinges Mandon, E., Rothnie, A. J. & Jawhari, A. The yin and yang of solubilization and stabilization for wild-type and full-length membrane protein. *Methods vol.* **147**, 118–125 (2018).

26. Iavarone, A. T., Jurchen, J. C. & Williams, E. R. Supercharged protein and peptide ions formed by electrospray ionization. *Anal. Chem.* **73**, 1455–1460 (2001).

27. Ogorzalek Loo, R. R., Lakshmanan, R. & Loo, J. A. What Protein Charging (and Supercharging) Reveal about the Mechanism of Electrospray Ionization. *J. Am. Soc. Mass Spectrom.* **25**, 1675–1693 (2014).

28. Sterling, H. J., Prell, J. S., Cassou, C. A. & Williams, E. R. Protein conformation and supercharging with DMSO from aqueous solution. *J. Am. Soc. Mass Spectrom.* **22**, 1178–1186 (2011).

29. Sterling, H. J. et al. Supercharging Protein Complexes from Aqueous Solution Disrupts their Native Conformations. *J. Am. Soc. Mass Spectrom.* **23**, 191 (2012).

30. Metwally, H., McAllister, R. G., Popa, V. & Konermann, L. Mechanism of Protein Supercharging by Sulfolane and m-Nitrobenzyl Alcohol: Molecular Dynamics Simulations of the Electrospray Process. *Anal. Chem.* **88**, 5345–5354 (2016).

31. Kebarle, P. & Verkcerk, U. H. Electrospray: From Ions in solution to Ions in the gas phase, what we know now. *Mass Spectrom. Rev.* **28**, 898–917 (2009).

32. Iribarne, J. V. & Thomson, B. A. On the evaporation of small ions from charged droplets. *J. Chem. Phys.* **64**, 2287 (1976).

33. Nguyen, S. & Fenn, J. B. Gas-phase ions of solute species from charged droplets of solutions. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 1111–1117 (2007).

34. Dole, M. et al. Molecular Beams of Macroions. *J. Chem. Phys.* **49**, 2240–2249 (1968).

35. Konermann, L., Ahadi, E., Rodriguez, A. D. & Vahidi, S. Unraveling the Mechanism of Electrospray Ionization. *Anal. Chem.* **85**, 2–9 (2012).

36. Cassou, C. A., Sterling, H. J., Susa, A. C. & Williams, E. R. Electrothermal supercharging in mass spectrometry and tandem mass spectrometry of native proteins. *Anal. Chem.* **85**, 138–146 (2013).

37. Flick, T. G. & Williams, E. R. Supercharging with trivalent metal ions in native mass spectrometry. *J. Am. Soc. Mass Spectrom.* **23**, 1885–1895 (2012).

38. Brahim, B., Alves, S., Cole, R. B. & Tabet, J.-C. Charge Enhancement of Single-Stranded DNA in Negative Electrospray Ionization Using the Supercharging Reagent Meta- nitrobenzyl Alcohol. *J. Am. Soc. Mass Spectrom.* **24**, (2013).

39. Lomeli, S. H., Yin, S., Ogorzalek Loo, R. R. & Loo, J. A. Increasing Charge While Preserving Noncovalent Protein Complexes for ESI-MS. *J. Am. Soc. Mass Spectrom.* **20**, 593–596 (2009).

40. Li, Y., Cole, R.B.: Shifts in peptide and protein charge state distributions with varying spray tip orifice diameter in nanoelectrospray Fourier transform ion cyclotron resonance mass spectrometry. *Anal. Chem.* **75**, 5739– 5746 (2003)

41. Mansoori, B. A., Volmer, D. A., Boyd, R. K. & Cole, R. B. 'Wrong-way-round' Electrospray Ionization of Amino Acids. *Rapid Commun. Mass Spectrom.* **11**, 1120–1130 (1997).

42. Zenaidee, M. A. & Donald, W. A. Electron capture dissociation of extremely supercharged protein ions formed by electrospray ionisation. *Anal. Methods.* **7**, 7132–7139 (2015).

43. Madsen, J. A. & Brodbelt, J. S. Comparison of infrared multiphoton dissociation and collision-induced dissociation of supercharged peptides in ion traps. *J. Am. Soc. Mass Spectrom.* **20**, 349–358 (2009)

44. Zhurov, K. O., Fornelli, L., Wodrich, M. D., Laskay, Ü. A. & Tsybin, Y. O. Principles of electron capture and transfer dissociation mass spectrometry applied to peptide and protein structure analysis. *Chem. Soc. Rev.* **42**, 5014–5030 (2013).

45. Brodbelt, J. S., Morrison, L. J. & Santos, I. Ultraviolet Photodissociation Mass Spectrometry for Analysis of Biological Molecules. *Chem. Rev.* **120**, 3328–3380 (2019).

46. Voinov, V. G., Bennett, S. E. & Barofsky, D. F. Electron-Induced Dissociation of Peptides in a Triple Quadrupole Mass Spectrometer Retrofitted with an Electromagnetostatic Cell. *J. Am. Soc. Mass Spectrom.* **26**, 752 (2015).

47. Cook, S. L. et al. Comparison of CID, ETD and metastable atom-activated dissociation (MAD) of doubly and triply charged phosphorylated tau peptides. *J. Mass Spectrom.* **47**, 786-794 (2012).

48. Zhou, M. et al. Higher-order structural characterisation of native proteins and complexes by top-down mass spectrometry. *Chemical Science. vol.* **11**, 12918–12936 (2020).

49. Yin, S. & Loo, J. A. Top-down mass spectrometry of supercharged native protein–ligand complexes. *Int. J. Mass Spectrom.* **300**, 118–122 (2011).

50. Lomeli, S. H., Peng, I. X., Yin, S., Ogorzalek Loo, R. R. & Loo, J. A. New Reagents for Increasing ESI Multiple Charging of Proteins and Protein Complexes. *J. Am. Soc. Mass Spectrom.* **21**, 127–131 (2010).

51. DiTusa, C. A., Christensen, T., McCall, K. A., Fierke, C. A. & Toone, E. J. Thermodynamics of Metal Ion Binding. 1. Metal Ion Binding by Wild-Type Carbonic Anhydrase. *Biochemistry*. **40**, 5338–5344 (2001).

52. Saito, R., Sato, T., Ikai, A. & Tanaka, N. Structure of bovine carbonic anhydrase II at 1.95 A resolution. *Acta Crystallogr. D. Biol. Crystallogr.* **60**, 792–795 (2004).

53. Keener, J. E. et al. Chemical Additives Enable Native Mass Spectrometry Measurement of Membrane Protein Oligomeric State within Intact Nanodiscs. *J. Am. Chem. Soc.* **141**, 1054–1061 (2019).

54. Denisov, I. G., Grinkova, Y. V., Lazarides, A. A. & Sligar, S. G. Directed Self-Assembly of Monodisperse Phospholipid Bilayer Nanodiscs with Controlled Size. *J. Am. Chem. Soc.* **126**, 3477–3487 (2004).

55. Going, C. C. & Williams, E. R. Supercharging with m-Nitrobenzyl Alcohol and Propylene Carbonate: Forming Highly Charged Ions with Extended, Near-Linear Conformations. *Anal. Chem.* **87**, 3973–3980 (2015).

56. Abaye, D. A., Agbo, I. A. & Nielsen, B. V. Current perspectives on supercharging reagents in electrospray ionization mass spectrometry. *RSC Adv.* **11**, 20355–20369 (2021).

57. Montaser, A. & Golighty, D. W. Inductively Coupled Plasmas in Analytical Atomic Spectrometry, 2nd Revised and Enlarged Edition. (Wiley, 1996).

58. Thomas, R. Practical Guide to ICP-MS. (CRC Press, 2013).

59. Amais, R. S., Amaral, C. D. B., Fialho, L. L., Schiavo, D. & Nóbrega, J. A. Determination of P, S and Si in biodiesel, diesel and lubricating oil using ICP-MS/MS. *Anal. Methods.* **6**, 4516–4520 (2014).

60. Douglas, D. J. Linear quadrupoles in mass spectrometry. *Mass Spectrom. Rev.* **28**, 937–960 (2009).

61. Hopfgartner, G. et al. Triple quadrupole linear ion trap mass spectrometer for the analysis of small molecules and macromolecules. *J. Mass Spectrom.* **39**, 845–855 (2004).

62. Krivan, V. & Haas, H. F. Prevention of loss of mercury(II) during storage of dilute solutions in various containers. *Fresenius' Zeitschrift für Anal. Chemie.* **332**, 1–6 (1988).

63. Louie, H., Wong, C., Huang, Y. J. & Fredrickson, S. A study of techniques for the preservation of mercury and other trace elements in water for analysis by inductively coupled plasma mass spectrometry (ICP-MS). *Anal. Methods.* **4**, 522-529 (2012)

64. NIST. SRM Order Request System SRM 1947 - Lake Michigan Fish Tissue. (2020). at <https://www-s.nist.gov/srmors/view_detail.cfm?srm=1947>

65. Liaw, M.-J., Jiang, S.-J. & Li, Y.-C. Determination of mercury in fish samples by slurry sampling electrothermal vaporization inductively coupled plasma mass spectrometry. *Spectrochim. Acta Part B At. Spectrosc.* **52**, 779–785 (1997).

66. Nixon, D. E., Burritt, M. F. & Moyer, T. P. Determination of mercury in whole blood and urine by inductively coupled plasma mass spectrometry. *Spectrochim. acta, Part B At. Spectrosc.* **54**, 1141–1153 (1999).

67. Allibone, J., Fatemian, E. & Walker, P. J. Determination of mercury in potable water by ICP-MS using gold as a stabilising agent. *J. Anal. At. Spectrom.* **14**, 235–239 (1999).

68. Yamada, N. Kinetic energy discrimination in collision/reaction cell ICP-MS: Theoretical review of principles and limitations. *Spectrochimica Acta - Part B Atomic Spectroscopy vol.* **110** 31–44 (2015).

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

70. Lippens, J. L. et al. Fourier Transform-Ion Cyclotron Resonance Mass Spectrometry as a Platform for Characterizing Multimeric Membrane Protein Complexes. *J. Am. Soc. Mass Spectrom.* **29**, 183–193 (2018).

71. Fan, Y. et al. Structure of a pentameric virion-associated fiber with a potential role in Orsay virus entry to host cells. *PLOS Pathog.* **13**, e1006231 (2017).

72. Gabelica, V. et al. Recommendations for reporting ion mobility Mass Spectrometry measurements. *Mass Spectrom. Rev.* **38**, 291–320 (2019).

73. Lanucara, F., Holman, S. W., Gray, C. J. & Eyers, C. E. The power of ion mobility-mass spectrometry for structural characterization and the study of conformational dynamics. *Nature Chem.* **6***,* 281–294 (2014)

74. Bagal, D., Valliere-Douglass, J. F., Balland, A. & Schnier, P. D. Resolving disulfide structural isoforms of IgG2 monoclonal antibodies by ion mobility mass spectrometry. *Anal. Chem.* **82**, 6751–6755 (2010).

75. Devine, P. W. A. et al. Investigating the Structural Compaction of Biomolecules Upon Transition to the Gas-Phase Using ESI-TWIMS-MS. *J. Am. Soc. Mass Spectrom.* **28**, 1855 (2017).

76. Loo, J. A. Studying noncovalent protein complexes by electrospray ionization mass spectrometry. *Mass Spectrom. Rev.* **16**, 1–23 (1997).

77. Callaway, E. The revolution will not be crystallized: A new method sweeps through structural biology. *Nature* **525**, 172–174 (2015).

78. Longchamp, J. N. et al. Imaging proteins at the single-molecule level. *Proc. Natl. Acad. Sci. U. S. A.* **114**, 1474–1479 (2017).

79. Mikhailov, V. A., Mize, T. H., Benesch, J. L. P. & Robinson, C. V. Mass-selective soft-landing of protein assemblies with controlled landing energies. *Anal. Chem.* **86**, 8321– 8328 (2014).

80. Clemmer, D. E. et al. Naked Protein Conformations: Cytochrome c in the Gas Phase. *J. Am. Chem. Soc.* **117**, 10141–10142 (2002).

81. Clarke, D. J. & Campopiano, D. J. Desalting large protein complexes during native electrospray mass spectrometry by addition of amino acids to the working solution. *Analyst* **140**, 2679–2686 (2015).

82. Light-Wahl, K. J., Schwartz, B. L. & Smith, R. D. Observation of the Noncovalent Quaternary Associations of Proteins by Electrospray Ionization Mass Spectrometry. *J. Am. Chem. Soc.* **116**, 5271–5278 (2002).

83. Seo, J. et al. Retention of Native Protein Structures in the Absence of Solvent: A Coupled Ion Mobility and Spectroscopic Study. *Angew. Chem. Int. Ed. Engl.* **55**, 14173–14176 (2016).

84. Harris, L. J., Skaletsky, E. & McPherson, A. Crystallographic structure of an intact IgG1 monoclonal antibody. *J. Mol. Biol.* **275**, 861–872 (1998).

85. Brockwell, D. J. et al. The Effect of Core Destabilization on the Mechanical Resistance of I27. *Biophys. J.* **83**, 458–472 (2002).

86. Benesch, J. L. P. Collisional Activation of Protein Complexes: Picking Up the Pieces. *J. Am. Soc. Mass Spectrom.* **20**, 341–348 (2009).

87. Bornschein, R. E. & Ruotolo, B. T. Ion mobility-mass spectrometry of charge-reduced protein complexes reveals general trends in the collisional ejection of compact subunits. Analyst 140, 7020–7029 (2015).

88. Rostom, A. A. et al. Detection and selective dissociation of intact ribosomes in a mass spectrometer. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 5185 (2000).

89. Gault, J. et al. High-resolution mass spectrometry of small molecules bound to membrane proteins. *Nat. Methods* **13**, 333–336 (2016).

90. Kim, M. et al. Need-based activation of ammonium uptake in *Escherichia coli*. *Mol. Syst. Biol.* **8**, 616 (2012).

91. Aquilina, J. A., Benesch, J. L. P., Bateman, O. A., Slingsby, C. & Robinson, C. V. Polydispersity of a mammalian chaperone: Mass spectrometry reveals the population of oligomers in αB-crystallin. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 10611–10616 (2003).

92. Marty, M. T. et al. Bayesian deconvolution of mass and ion mobility spectra: From binary interactions to polydisperse ensembles. *Anal. Chem.* **87**, 4370–4376 (2015).

93. Jeong, K. et al. FLASHDeconv: Ultrafast, High-Quality Feature Deconvolution for Top- Down Proteomics. *Cell Syst.* **10**, 213-218.e6 (2020).

94. Bern, M. et al. Parsimonious Charge Deconvolution for Native Mass Spectrometry. *J. Proteome Res.* **17**, 1216–1226 (2018).

95. Zenaidee, M. A. et al. Internal Fragments Generated from Different Top-Down Mass Spectrometry Fragmentation Methods Extend Protein Sequence Coverage. *J. Am. Soc. Mass Spectrom.* **32**, 1752–1758 (2021).

96. Durbin, K. R., Skinner, O. S., Fellers, R. T. & Kelleher, N. L. Analyzing internal fragmentation of electrosprayed ubiquitin ions during beam-type collisional dissociation. *J. Am. Soc. Mass Spectrom.* **26**, 782–787 (2015).

97. Zenaidee, M. A. et al. Internal fragments generated by electron ionization dissociation enhance protein top-down mass spectrometry. *J. Am. Soc. Mass Spectrom.* **31**, 1896– 1902 (2020).

98. Rolfs, Z. & Smith, L. M. Internal Fragment Ions Disambiguate and Increase Identifications in Top-Down Proteomics. *J. Proteome Res.* **20**, 5412–5418 (2021)

99. Lyon, Y. A., Riggs, D., Fornelli, L., Compton, P. D. & Julian, R. R. The Ups and Downs of Repeated Cleavage and Internal Fragment Production in Top-Down Proteomics. *J. Am. Soc. Mass Spectrom.* **29**, 150–157 (2018).

100. Schmitt, N. D., Berger, J. M., Conway, J. B. & Agar, J. N. Increasing Top-Down Mass Spectrometry Sequence Coverage by an Order of Magnitude through Optimized Internal Fragment Generation and Assignment. *Anal. Chem.* **93**, 6355–6362 (2021).

101. Lantz, C. et al. ClipsMS: An Algorithm for Analyzing Internal Fragments Resulting from Top-Down Mass Spectrometry. *J. Proteome Res.* **20**, 1928–1935 (2021).

102. Hu, Q. et al. The Orbitrap: a new mass spectrometer. *J. Mass Spectrom.* **40**, 430–443 (2005).

103. Thermo Scientific. Q Exactive UHMR Hybrid Quadrupole-Orbitrap Mass Spectrometer System - US. (2018) at <https://www.thermofisher.com/us/en/home/industrial/mass-spectrometry/liquid-chromatography-mass-spectrometry-lc-ms/lc-ms-systems/orbitrap-lc- ms/q-exactive-orbitrap-mass-spectrometers/q-exactive-uhmr.html>

104. Engel, A. & Gaub, H. E. Structure and Mechanics of Membrane Proteins. *Annurev. Biochem.* **77**, 127–148 (2008).

105. Overington, J. P., Al-Lazikani, B. & Hopkins, A. L. How many drug targets are there? *Nat. Rev. Drug Discov.* **5**, 993–996 (2006).

106. Allen, K. N., Entova, S., Ray, L. C. & Imperiali, B. Monotopic Membrane Proteins Join the Fold. *Trends in Biochemical Sciences vol.* **44**, 7–20 (2019).

107. Hendrickson, W. A. Atomic-level analysis of membrane-protein structure. *Nat. Struct. Mol. Biol.* **23**, 464–467 (2016)

108. Fernández, C., Hilty, C., Wider, G., Güntert, P. & Wüthrich, K. NMR Structure of the Integral Membrane Protein OmpX. *J. Mol. Biol.* **336**, 1211–1221 (2004).

109. Thonghin, N., Kargas, V., Clews, J. & Ford, R. C. Cryo-electron microscopy of membrane proteins. *Methods* **147**, 176–186 (2018).

110. Carpenter, E. P., Beis, K., Cameron, A. D. & Iwata, S. Overcoming the challenges of membrane protein crystallography. *Current Opinion in Structural Biology vol.* **18**, 581–586 (2008).

111. Whitelegge, J. P., Gundersen, C. B. & Faull, K. F. Electrospray-ionization mass spectrometry of intact intrinsic membrane proteins. *Protein Sci.* **7**, 1423 (1998).

112. Olinares, P. D. B. et al. Native Mass Spectrometry-Based Screening for Optimal Sample Preparation in Single-Particle Cryo-EM. *Structure.* **29**, 186-195 (2020)

113. Sobott, F., Hernández, H., McCammon, M. G., Tito, M. A. & Robinson, C. V. A tandem mass spectrometer for improved transmission and analysis of large macromolecular assemblies. *Anal. Chem.* **74**, 1402–1407 (2002).

114. Sharon, M., Ilag, L. L. & Robinson, C. V. Evidence for micellar structure in the gas phase. *J. Am. Chem. Soc.* **129**, 8740–8746 (2007).

115. Zabrouskov, V. & Whitelegge, J. P. Increased coverage in the transmembrane domain with activated-ion electron capture dissociation for top-down Fourier-transform mass spectrometry of integral membrane proteins. *J. Proteome Res.* **6**, 2205–2210 (2007).

116. Barrera, N. P., Di Bartolo, N., Booth, P. J. & Robinson, C. V. Micelles protect membrane complexes from solution to vacuum. *Science.* **321**, 243–246 (2008).

117. Ryan, C. M. et al. Post-translational modifications of integral membrane proteins resolved by top-down Fourier transform mass spectrometry with collisionally activated dissociation. *Mol. Cell. Proteomics* **9**, 791–803 (2010).

118. Laganowsky, A., Reading, E., Hopper, J. T. S. & Robinson, C. V. Mass spectrometry of intact membrane protein complexes. *Nat. Protoc.* **8**, 639–51 (2013).

119. Gupta, K. et al. The role of interfacial lipids in stabilizing membrane protein oligomers. *Nature* **541**, 421–424 (2017).

120. Yen, H.-Y. et al. Ligand binding to a G protein-coupled receptor captured in a mass spectrometer. Sci. Adv. 3, e1701016 (2017).

121. Susa, A. C. et al. Submicrometer Emitter ESI Tips for Native Mass Spectrometry of Membrane Proteins in Ionic and Nonionic Detergents. *J. Am. Soc. Mass Spectrom.* **29**, 203–206 (2018).

122. Conroy, M. J. et al. The crystal structure of the Escherichia coli AmtB–GlnK complex reveals how GlnK regulates the ammonia channel. *Proc. Natl. Acad. Sci.* **104**, 1213–1218 (2007).

**Effects of Nonionic Saccharide Detergents, Amides, and Nitriles on ESI Charge State Distributions**

Wonhyeuk Jung†, Wenzhe Chen†, Muhammad Zenaidee†, Boyu Zhao†, Pascal Egea‡, Rachel R. Ogorzalek Loo† and Joseph A. Loo†‡

†Department of Chemistry and Biochemistry, University of California, Los Angeles, California 90095, United States

‡Department of Biological Chemistry, David Geffen School of Medicine, University of California, Los Angeles, California 90095, United States

The following is the preliminary manuscript draft being prepared for a submission.

## Abstract

The advent of the soft ionization technique, electrospray ionization (ESI), has made it possible to ionize large biomolecules and transition from the solution phase to gas phase without harming their native structure and non-covalent complexes while conferring multiple charging. In turn, these complexes can be subjected to various fragmentation approaches, which can lead to novel proteoform identification, stoichiometry information, and structural insights. However, when ESI is performed from a solution close to physiological pH, biomolecules with globular structure tend to acquire less charge, which reduced the efficiency for subsequent fragmentation analysis. Supercharging, an increase in charge state induced by small molecule

addition to the analyte solution, can counter this problem, but the mechanism behind supercharging is still unclear. To gain further insights, charge state shifts caused by non-ionic saccharide-based detergents, which is a class of detergent commonly used for membrane protein solubilization, and amides and nitriles with varying gas phase basicity were examined.

The alkyl chain length and the critical micelle concentration of the detergents did not have significant impact on the charge state of a 3x-Flag-tag peptide and the protein lysozyme. In contrast, the polar head group affected the resulting charge state, with glucoside-based detergents having the strongest supercharging effect. Screening of amides and nitriles showed that gas phase basicity does not have clear correlation with charge state. The result of this study implies that Brønsted basicity of solution and the intermediate phase, which is the phase that bridge the solution phase and gas phase, is one of the primary driving factors behind supercharging.

## Introduction

Electrospray ionization (ESI) is an ionization technique that has become the gold standard for mass spectrometry (MS) of intact proteins and protein complexes.[1] ESI is considered a "soft" ionization technique, i.e., the ionization process induces little to no fragmentation when intact proteins and protein complexes are transferred from solution into the gas phase. ESI produces multiply charged ions and this aspect of the technique has proven to be instrumental in analyzing large analytes.[2]

Multiply charged ions produced by ESI can be fragmented with various activation approaches to gather structural and sequence information.[3–6] For the analysis of protein complexes and proteoforms, the ability to enhance the charge on the ion can result in higher efficiency

51

dissociation and thus sequence coverage.[7] Conventionally, protein ions are dissolved into a denaturing solution containing acid that elongates the protein, allowing for more charge to be deposited on the protein. In contrast, ions are generated from solutions at physiological pH during native ESI, which typically leads to less average charge and less efficient fragmentation.[8] However, the latter approach can conserve non-covalent interactions of interest and solution-phase structure in the gas phase.[9–11] Manipulation of charge state for native ESI, therefore, can be highly beneficial, especially for structure-focused studies.

More than 30 years after its development, the debate about the ESI mechanism is still actively ongoing. For native ESI, especially for large, globular proteins, it is widely accepted that transition from the solution phase to the gas phase occurs via the charge residue model (CRM).[12] The analytes contained inside the droplets produced during ESI gain charge as the solvent completely evaporates during the desolvation process.[13,14] The major experimental proof for this model is that ionized globular proteins' charge states tends to align closely with the Rayleigh charge of protein-sized water droplets.[15] However, challenges to the CRM has been accumulating.

Since the droplet will undergo multiple fission cycles before complete desolvation during CRM-based ESI, nascent droplet size should not affect the final charge state of the target analyte. However, as the emitter's diameter decreases, analyte charge state tends to increase for both proteins and peptides.[16–18] When non-ionic saccharide-based detergents are added to the sample prior to ESI, it can cause charge increase under both denaturing and native conditions.[19] This phenomenon contradicts the CRM because decrease in surface tension caused by detergent treatment should lead to decrease in charge state under the model. Finally, the addition of m-nitrobenzyl alcohol (m-NBA) to aqueous solutions results in increase in charge state. The increase happens without harming non-covalent interactions while decreasing surface tension.[10]

Iavarone and Williams, who initially discovered that glycerol and m-NBA can be added to ESI solutions to increase the charge state,[20] named this phenomenon as 'supercharging' and the mechanism has been under debate for last 20 years.[13,21–28] Currently, there are four major physicochemical properties that are believed to be important for supercharging capacity of a reagent: volatility, surface tension, Brønsted basicity, and the dipole moment.[22,23,28–31] Supercharging agents that are less volatile compared to the solvent (e.g., water and methanol) will be enriched during the Coulombic fission/desolvation process, which amplifies their physicochemical properties' effect on the analyte charging.

In the surface tension model,[32] the supercharging effect results from the relatively high surface tension of solution additives (>38 mN m−1).[28] For a water-acetic acid solvent, water evaporates more rapidly than acetic acid. Thus, the nanodroplets in the final stages of ESI are likely to be enriched with acetic acid and thus should have lower surface tensions than pure water.[33] Supercharging agents with lower volatility and larger surface tensions than acetic acid should increase the surface tension of the nanodroplet, which in turn will increase the average charge state (ACS) of the analyte, according to the CRM. In the Brønsted basicity model, as the charge inside the final nanodroplet gets distributed among the solvent, the analytes, and the supercharging agents, i.e., molecules that have lower Brønsted basicity and thus more negative $pK_a$ values of the conjugate acids, will yield more charge to the analytes.[23] In the dipole moment model, the relatively high electric dipole moments of supercharging agents can decrease the Coulombic repulsion caused by charge density inside the nanodroplets, which increases the extent of charge nanodroplets can harbor.[28,34]

Throughout this debate, the term 'supercharging' has expanded to include charge increase resulting from high capillary voltage and/or capillary temperature induced denaturation.[24] Generally, these studies were performed in denaturing conditions. Here, we explore how supercharging takes place not only in a denaturing condition but also in a native solution

53

condition. In addition, we limit the definition of supercharging as the charge increase caused by the chemical additive that was spiked in at relatively low concentration prior to ESI that does not lead to significant denaturation in solution.[10]

As mentioned above, it was observed that non-ionic saccharide-based detergents can increase the charge state of soluble proteins and peptides.[19] This class of detergents is widely used to solubilize membrane proteins for native MS analysis and the correlation between the detergents and the charge state observed for membrane proteins have been closely studied.[35] Despite extensive use of these detergents for membrane protein analysis, the mechanism of how these detergents affect the charge state of electrosprayed proteins has not been explored in depth.

The role in enhancing the charge state of soluble proteins and peptides by detergents commonly used in membrane protein analysis should answer fundamental questions about the mechanism of charging by these additives.

Although the utility of manipulating charge states of the target analyte is great for mass spectrometry applications, the lack of a straightforward model to identify potential supercharging agents and gauge the extent of supercharging to be expected has hindered more widespread application of the technique. In this study, we probe supercharging effects of six glucoside-based detergents (n-hexyl, heptyl, octyl, nonyl, decyl and dodecyl-β-D glucosides), two maltoside-based detergents (n-hexyl, nonyl-β-D-maltosides), two cyclohexyl maltoside-based detergents (5-cyclohexyl-1-pentyl-β-D-maltoside, 6-cyclohexyl-1-hexyl-β-D-maltoside), and 18 different amides and nitriles.

**Materials and Methods**

**Detergents, peptides, and protein solution preparation for MS analysis.** 3x-Flag-tag

peptide (Sigma-Aldrich, F4799) was dissolved in 49% acetonitrile, 50% water, 1% formic acid at

20 µM concentration and was used without further purification. Lysozyme (Sigma-Aldrich,

L6876) was dissolved with 200 mM NH4OAc at 20 µM concentration and was used without

further purification for detergent supercharging experiment. For conventional supercharging

agents, amides, and nitrile testing, lysozyme was dissolved in 50% water, 49.9% acetonitrile,

0.1% formic acid and then used without further purification. All detergents (n-hexyl, heptyl, octyl,

nonyl, decyl and dodecyl-β-D glucosides, n-hexyl, nonyl-β-D-maltosides, 5-cyclohexyl-1-pentyl-

β-D-maltoside, 6-cyclohexyl-1-hexyl-β-D-maltoside) were purchased from Anatrace and were

dissolved into 49% acetonitrile, 50% water, 1% formic acid or 200 mM NH4OAc at 5 to 20-fold

of their respective CMCs, except for dodecyl-β-D glucosides, which was dissolved at 1X CMC

because of its limited solubility.

**Preparation and treatment of supercharging agents, amides, and nitriles.** 3-Nitrobenzyl

alcohol (m-NBA, Sigma, 73148), 1, 2-butylene carbonate (BC, TCI America, B3321), and

sulfolane (Sigma, T22209) were directly spiked into the 20 µM lysozyme dissolved in 50%

water, 49.9% acetonitrile, 0.1% formic acid. All amides were dissolved in 50% water, 49.9%

acetonitrile, 0.1% formic acid to a stock concentration of 2% w/v and then spiked in at an

appropriate volume to the final concentration of 0.2% w/v.

**Synapt G2-Si mass spectrometry.** 20 µM Lysozyme and 20 µM 3x-flag-tag peptide with or

without respective detergents were loaded into platinum-coated borosilicate capillaries (Thermo

Fisher Scientific, ES387) and sprayed at a flow rate of 10-40 nL/min through a nanospray ion

source and analyzed with a Synapt G2-Si mass spectrometer (Waters, Milford, MA). Key

parameters for detergent treated samples were: cone 40 V, trap collision 20-50 V, transfer collision 20-30 V, capillary 1 kV-1.2 kV, and trap gas flow at 4 ml/min. 20 µM Lysozyme treated with m-NBA, 1,2-butylene carbonate, sulfolane, amides, and nitriles were loaded into borosilicate capillaries (Warner Instruments, 30-0042) that was pulled in-house with a Sutter P-1000 Micropipette Puller and then coated with either gold or platinum for 4 minutes (Anatech, Hummer 6.2 Sputtering System). Key parameters for supercharging agents, amides, and nitriles treated samples were: cone 40 V, trap collision 20-50 V, transfer collision 20-30 V, capillary 1 kV- 1.2 kV, and trap gas flow at 4 ml/min.

One hundred spectra were averaged, and all spectra were externally calibrated with cesium iodide. All measurements were done in triplicate and average chare states were calculated by intensity-weighted average of all peaks that had signal-to-noise ratio larger than 5. For ion mobility studies, the trap and TWIM devices were filled with nitrogen. The T-wave settings were: trap 300 ms-1/2.0 V, IMS 300 ms-1/20–24 V, and transfer 100 ms-1/10 V. Source temperature was kept at 100 °C across all experiments.

**Circular dichroism spectroscopy of myoglobin.** 10 µM myoglobin was dissolved in 200 mM ammonium acetate with 9.6 mM to 48 mM OG added or 0.4% formic acid. The samples were analyzed with a Jasco J-175 spectropolarimeter in a 1mm cuvette. The degree of ellipticity at 222 nm was recorded.

## Results and Discussion

**Charge increase effect of non-ionic saccharide-based detergents are dependent on polar head group.** Non-ionic saccharide-based detergents, especially glucoside- and maltoside-based detergents, are commonly utilized for membrane protein solubilization.[36-38] Interestingly,

for soluble proteins and peptides, these detergents were able to generate protein and peptide ions with higher charge state distributions than protein and peptide ions without these detergents.[19] Although enhancing charge has many advantages for top-down MS analysis, the mechanism of how these detergents impart charge on proteins is not well understood.

In a study conducted by Kundlacz et al.,[38] the effect of three commonly used detergents, 1-O-(n-Octyl)-tetraethyleneglycol (C8E4), n-octyl β-d-glucoside (OG), and N,N-Dimethyl-n-dodecylamine-N-oxide (LDAO), on charge state distributions (CSDs) of soluble proteins were screened. C8E4 had a primarily charge reducing effect and OG had a charge increasing effect. LDAO, in contrast, appeared to affect the number of charges only when a transmembrane domain is present. The study provides interesting insights but does not delve into how the physicochemical properties of the detergents could affect the charge manipulation properties.

Here, 3x-flag-tag peptide and native lysozyme were used as model systems to probe the effect of chain length on supercharging for six glucoside-based detergents (n-hexyl, heptyl, octyl, nonyl, decyl and dodecyl-β-D glucosides), two maltoside-based detergents (n-hexyl, nonyl-β-D-maltosides) and two cyclohexyl maltoside-based detergents (5-cyclohexyl-1-pentyl-β-D-maltoside, 6-cyclohexyl-1-hexyl-β-D-maltoside) (**Figure 1**).

For the 3x-flag-tag peptide, the chain length of the detergents had no correlation with the charge increase effect. For example, for hexyl glucoside, the average charge was 5.62±0.12 and for nonyl glucoside, the average charge was 5.51±0.04 (**Supplementary Table 1**). Similarly, for native lysozyme, the chain length of the detergents had no clear correlation with the charge increase effect. For example, for hexyl glucoside, the average charge was 5.62±0.12 and for nonyl glucoside, the average charge was 5.51±0.04. The data presented here suggests little correlation between the alkyl chain length and the extent of charging observed, which is in good agreement with previous results by Foley et. al., showing that for carbonates, the average charge state does not correlate with alkyl chain length.[28]
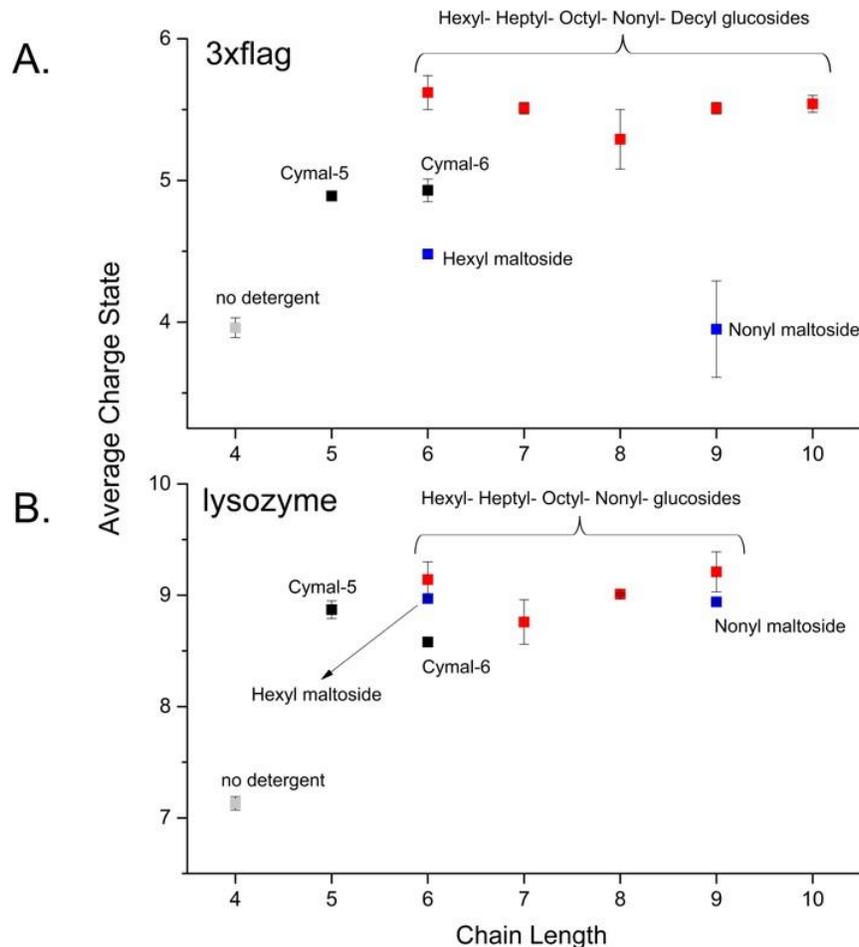
**Figure 1. Non-ionic saccharide-based detergent's effect on average charge state of 3x-flag-tag peptide and lysozyme.** A. Average charge state shift of 3x-flag-tag peptide in denaturing condition and B. lysozyme in native condition induced by six glucoside-based detergents of varying alkyl chain lengths (red) was compared to two maltoside-based detergents (blue) and two cyclohexyl maltoside-based detergents (black). Error bar indicates standard deviation from triplicate experiments. (Cymal-5: 5-cyclohexyl-1-pentyl-β-D-maltoside, Cymal-6: 6-cyclohexyl-1-hexyl-β-D-maltoside) All detergents were added to a 12mM concentration.

Interestingly, for both 3x-flag-tag peptide and lysozyme, glucoside-based detergents induced more charge increase compared to maltoside-based detergents. For 3x-flag-tag peptide, the hexyl-glucoside caused ACS shift from 3.96±0.07 to 5.62±0.12 compared to the hexyl maltoside, which increased ACS to only 4.48±0.03. Similar results were observed for lysozyme, although the charge increase is not as pronounced. These results indicate that the polar head group may be responsible for the mechanism of charging for these soluble proteins.

Surprisingly, when comparing to additives with similar polar head groups, maltoside to cymal-6, the extent of charge observed was not significantly different. These results can be rationalized by the polar head groups being responsible for the extent of charge on a protein with glucoside detergents imparting more charge than maltoside based detergents.

These results indicate that the polar head groups and analyte identity affect supercharging capacity rather than the alkyl chain length for non-ionic saccharide-based detergents. This can be rationalized by examining the two main proposed physicochemical properties linked to supercharging: (i) the surface tension of the additive, and (ii) the dipole moment of the additive. As the chain length of the additive increases, the surface tension of the additive should decrease. Our results show that the average charge state of the proteins does not change drastically with increasing tail length, suggesting that the supercharging effect observed here is not related to the surface tension mechanism. In contrast, we see significant differences in average charge states when comparing glucoside-based detergents with maltoside based detergents. Glucoside based detergents generally have lower dipole moments than maltoside based detergents, which suggests that the results observed here are not in agreement with the dipole moment model. Based on this, it suggests that the bulk solution is not responsible for extending charge on a protein ion in the gas phase.

In the gas phase, protein ions can donate protons to basic compounds in proton-transfer reactions. As glucosides have lower basicities than the maltosides, the data shown here agrees with the proton-transfer reactions related to charging of protein ions.[39]

**Charge increase effect of detergents are independent of micelle formation.** A key use of glucoside and maltoside based detergents is the solubilization of membrane proteins, allowing for the transfer of membrane proteins from solution to the gas phase.[40] Solubilization of

membrane proteins prior to native MS analysis typically takes place at 2X the critical micelle

concentration (CMC).[37] The CMC represents the concentration at which micelle formation

begins to take place. Adding detergents beyond this concentration does not increase the

monomer concentration in solution.[40] Thus, the detergent monomer concentration is effectively

dictated by their respective CMC value. Charge states of membrane proteins vary widely with

the detergents utilized for their solubilization.[35] However, little is known about the correlation

between the charge states of the membrane proteins and the detergents' CMC values. Here,

the charge increase effect of the detergents as CMC concentration increases from 0.5X to 2X

CMC was screened with native lysozyme (**Figure 2**).
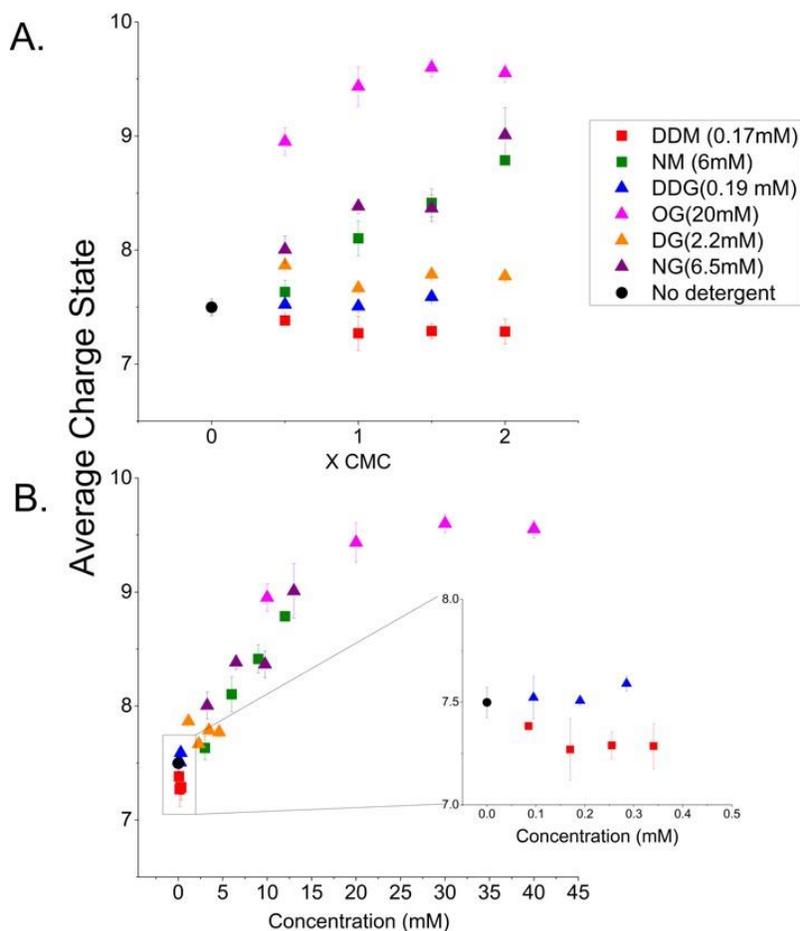


**Figure 2. Correlation between average charge state of native lysozyme and the concentrations of the detergents treated.** A. Average charge shift of native lysozyme induced by four glucoside-based detergents and two maltoside-based detergents of varying chain lengths was screened. All detergents were treated at 0.5X, 1X, 1.5X and 2X CMC, except for DDG, which was insoluble at 2X CMC (each detergent's respective CMC is mentioned inside

the bracket). Error bars indicate standard deviation from triplicate experiments. (DDM: dodecyl-β-D-maltosides, DDG: dodecyl-β-D glucosides, DG: decyl-β-D glucosides, NM: nonyl-β-D-maltosides, NG: nonyl-β-D-glucosides, OG: octyl-β-D-glucosides) B. Figure A was re-plotted with the x-axis representing the absolute concentration. The inset shows detergents with especially low CMCs (DDG and DDM).

Detergents with longer alkyl chain and thus relatively lower CMCs (DDG, DG and DDM) did not increase the charge of native lysozyme as their concentration climbed from 0.5X to 2X CMC. In contrast, detergents with shorter alkyl chain and higher CMCs (OG, NG and NM) showed concentration dependent supercharging behavior (**Figure 2A**). When the data is replotted with the x-axis representing the absolute concentration (**Figure 2B**), a linear charging behavior of proteins in detergents with relatively higher CMCs becomes more apparent, which is observed approximately up to 20 mM. The result indicates that the absolute concentration of non-ionic saccharide-based detergents, rather than the CMC, determines the supercharging capacity of non-ionic saccharide-based detergents with relatively short alkyl chains. In other words, the number of monomers present in the solution, which increases as CMC increases, dictates over the extent of supercharging conferred by the detergents.

Another proposed mechanism linked to supercharging is that supercharging additives unfold the protein during the ESI process.[31,41,42] ESI of unfolded proteins tend to accommodate more charge, which leads to higher and broader charge state distributions, while folded proteins typically exhibit lower and narrower charge state distributions.[43] Thus, if detergents are causing unfolding and denaturation of the soluble proteins with increasing concentration, the concentration dependent supercharging behavior can be explained.

To investigate whether non-ionic saccharide-based detergents cause denaturation, octyl-β-D glucoside (OG), which showed the strongest supercharging capacity during native ESI (**Figure 3**), was added to solutions of myoglobin. The myoglobin secondary structure was screened with circular dichroism. Virtually no secondary structure change was observed as OG concentration

61

increased from 0.4X CMC (9.6 mM) to 2.0X CMC (48 mM), indicating that OG treatment does not harm secondary structure (**Supplementary Figure 1**).

Also, the ion mobility drift times of the two major charge states of native lysozyme were measured in the presence of glucoside-based detergents of different alkyl chain length (**Supplementary Figure 2**). The drift time (the time it takes for the analyte to travel through a drift tube filled with non-reactive gas such as nitrogen) from native ion mobility MS study is a representation of the overall tertiary structure and depends on the charge and the collision cross section.[43] If denaturation is taking place for lysozyme in the presence of glucoside-based detergents and thus causing an increase in ACS, an increase in drift time should be observed as well. However, no significant drift time increase was observed for hexyl-, heptyl-, nonyl-, and octyl-β-D glucosides. Thus, detergent-based supercharging is not caused by denaturation. Rather the monomer population density, which is decided by the CMC of the respective detergents, and the basicity of the polar head group seem to be the deciding factor for supercharging capacity.

**Screening of 18 amides and nitriles' effect on denatured lysozyme ACS compared to 3 conventional supercharging agents. Supplementary Table 2** lists the gas phase basicities of the amides and nitriles screened[45] and **Supplementary Table 3** shows the chemical structures of the conventional supercharging agents that were used as a positive control, and amides/nitriles with various functional groups that were screened for their sub/supercharging capacity. These reagents' effect on intact lysozyme in denaturing condition (see Material and Methods for details) was screened (**Figure 3**).
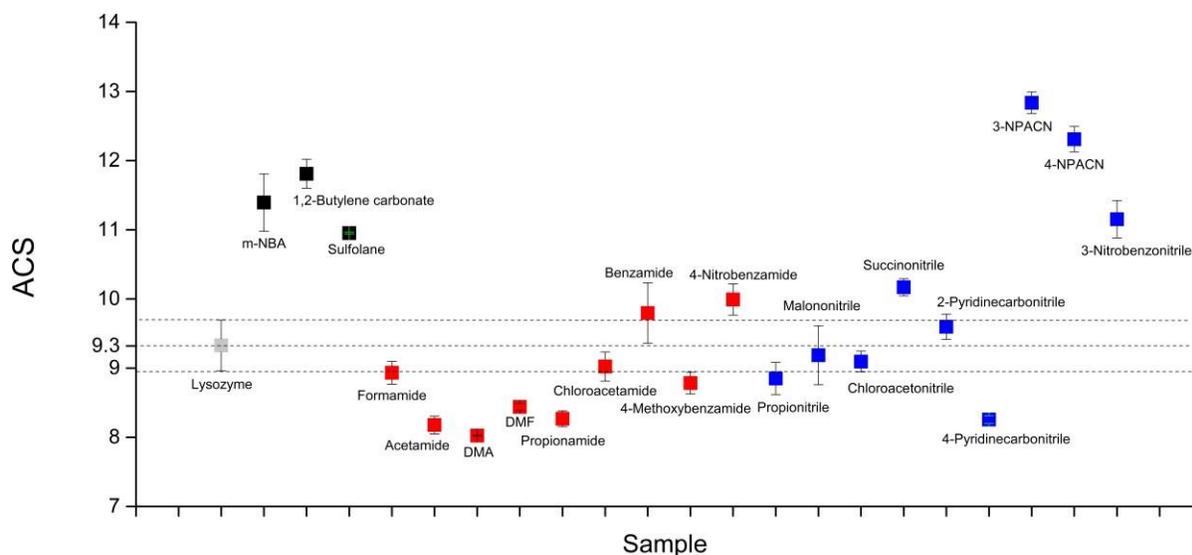
**Figure 3. ACS of intact lysozyme in denaturing condition after treatment with conventional supercharging agents (black), amides (red), and nitriles (blue).** Lysozyme was solubilized in 50% water, 49.9% acetonitrile, 0.1% formic acid. Every reagent was added to a level of 0.2% w/v. ACSs are the average charge state values from triplicate measurements and the error bars indicate a standard deviation from the triplicate measurements. Along with the three conventional supercharging agents (m-NBA, 1,2-butylene carbonate, and sulfolane), 4-nitrobenzamide, succinonitrile, 3-NPACN, 4-NPACN, and 3- nitrobenzonitrile show supercharging behavior. Acetamide, DMA, DMF, propionamide, and 4- pyridinecarbonitrile show subcharging behavior. (m-NBA: 3-nitrobenzylalcohol, DMA: N, N- dimethylacetamide, DMF: N, N-dimethylformamide, 3-NPACN: 3-nitrophenylacetonitrile, 4- NPACN: 4-nitrophenylacetonitrile)

Benzamide shows slight supercharging behavior while having a gas phase basicity of 861.2 kJ/mol. However, DMA, with a higher gas phase basicity (877.0 kJ/mol) decreases ACS. DMF has a lower gas phase basicity (856.6 kJ/mol) while also showing subcharging. Chloroacetonitrile (715.1 kJ/mol) shows no effect on ACS. Although propionitrile (694.1 kJ/mol) with significantly less gas phase basicity also shows no effect, 3-nitrobenzonitrile (750.7 kJ/mol) shows clear supercharging behavior. A gas phase basicity increase for nitriles do not clearly result in supercharging either since 2-pyridine carbonitrile (841.0 kJ/mol) does not have significant effect on the ACS. Thus, no clear connection between the gas phase basicity and the supercharging capacity can be drawn.

## Conclusion

To further gain insights into the supercharging mechanism, non-ionic saccharide-based detergents' effects on CSDs of 3x-flag-tag peptide and lysozyme were screened. In line with the previously published results,[19] in which detergents show primarily supercharging behavior, most of the detergents screened showed supercharging in both denaturing and native conditions.

Upon further inspection of the correlation between the physicochemical properties of the detergents and their charge altering capacity, two major conclusions could be drawn: (i) supercharging capacity of non-ionic saccharide-based detergents primarily depends on polar head group rather than the alkyl chain length; (ii) for non-ionic saccharide-based detergents with relatively higher CMCs, the supercharging capacity scales linearly with detergent concentration.

The implication of the first conclusion is that the surface tension model and the dipole model of the supercharging mechanism cannot explain the detergents' supercharging behavior. Given that glucoside-based detergents consistently show stronger supercharging compared to maltoside-based detergents, and that glucosides have lower basicities than the maltosides, the Brønsted basicity model seems to be the primary driver of the supercharging of non-ionic saccharide-based detergents.

Also, to examine the correlation between the gas phase basicity and supercharging, 19 different amides and nitriles were added to lysozyme under denaturing conditions. The data clearly indicates that gas phase basicity cannot explain the observed ACS altering capacity of the chemical agents tested here. This further strengthens the argument that solution phase and intermediate phase[23] basicity dictates over the charge altering effect when no significant structural rearrangement is induced during ESI.

## References

1. Fenn, J. B. Electrospray wings for molecular elephants (Nobel lecture). in *Angewandte Chemie - International Edition* vol. **42**, 3871–3894 (2003).

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

3. 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. U. S. A.* **101**, 9528–9533 (2004).

4. Zubarev, R. A. et al. Electron capture dissociation for structural characterization of multiply charged protein cations. *Anal. Chem.* **72**, 563–573 (2000).

5. 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).

6. Hogan, J. M. & McLuckey, S. A. Charge state dependent collision-induced dissociation of native and reduced porcine elastase. *J. Mass Spectrom.* **38**, 245–256 (2003).

7. Zenaidee, M. A. & Donald, W. A. Electron capture dissociation of extremely supercharged protein ions formed by electrospray ionisation. *Anal. Methods* **7**, 7132–7139 (2015).

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

9. Pacholarz, K. J., Garlish, R. A., Taylor, R. J. & Barran, P. E. Mass spectrometry based tools to investigate protein–ligand interactions for drug discovery. *Chem. Soc. Rev.* **41**, 4335–4355 (2012).

10. Lomeli, S. H., Yin, S., Ogorzalek Loo, R. R. & Loo, J. A. Increasing Charge While Preserving Noncovalent Protein Complexes for ESI-MS. *J. Am. Soc. Mass Spectrom.* **20**, 593–596 (2009).

11. Wyttenbach, T. & Bowers, M. T. Structural Stability from Solution to the Gas Phase: Native Solution Structure of Ubiquitin Survives Analysis in a Solvent-Free Ion Mobility– Mass Spectrometry Environment. *J. Phys. Chem. B* **115**, 12266–12275 (2011).

12. Konermann, L., Ahadi, E., Rodriguez, A. D. & Vahidi, S. Unraveling the Mechanism of Electrospray Ionization. *Anal. Chem.* **85**, 2–9 (2012).

13. Iavarone, A. T. & Williams, E. R. Mechanism of charging and supercharging molecules in electrospray ionization. *J. Am. Chem. Soc.* **125**, 2319–2327 (2003).

14. Kebarle, P. & Verkcerk, U. H. Electrospray: From Ions in solution to Ions in the gas phase, what we know now. *Mass Spectrom. Rev.* **28**, 898–917 (2009).

15. Fernandez De La Mora, J. Electrospray ionization of large multiply charged species proceeds via Dole's charged residue mechanism. *Anal. Chim. Acta* **406**, 93–104 (2000).

16. Li, Y. & Cole, R. B. Shifts in Peptide and Protein Charge State Distributions with Varying Spray Tip Orifice Diameter in Nanoelectrospray Fourier Transform Ion Cycltron Resonance Mass Spectrometry. *Anal. Chem.* **75**, 5739–5746 (2003).

17. Yuill, E. M., Sa, N., Ray, S. J., Hieftje, G. M. & Baker, L. A. Electrospray Ionization from Nanopipette Emitters with Tip Diameters of Less than 100 nm. *Anal. Chem.* **85**, 8498– 8502 (2013).

18. Nesatyy, V. J. Mass spectrometry evaluation of the solution and gas-phase binding properties of noncovalent protein complexes. *Int. J. Mass Spectrom.* **221**, 147–161 (2002).

19. Loo, R. R., Dales, N. & Andrews, P. C. The effect of detergents on proteins analyzed by electrospray ionization. *Methods Mol. Biol.* **61**, 141–160 (1996).

20. Iavarone, A. T., Jurchen, J. C. & Williams, E. R. Supercharged protein and peptide ions formed by electrospray ionization. *Anal. Chem.* **73**, 1455–1460 (2001).

21. Konermann, L., Metwally, H., Duez, Q. & Peters, I. Charging and supercharging of proteins for mass spectrometry: Recent insights into the mechanisms of electrospray ionization. *Analyst.* **144**, 6157–6171 (2019).

22. Lomeli, S. H., Peng, I. X., Yin, S., Ogorzalek Loo, R. R. & Loo, J. A. New Reagents for Increasing ESI Multiple Charging of Proteins and Protein Complexes. *J. Am. Soc. Mass Spectrom.* **21**, 127–131 (2010).

23. Ogorzalek Loo, R. R., Lakshmanan, R. & Loo, J. A. What Protein Charging (and Supercharging) Reveal about the Mechanism of Electrospray Ionization. *J. Am. Soc. Mass Spectrom.* **25**, 1675–1693 (2014).

24. Cassou, C. A., Sterling, H. J., Susa, A. C. & Williams, E. R. Electrothermal supercharging in mass spectrometry and tandem mass spectrometry of native proteins. *Anal. Chem.* **85**, 138–146 (2013).

25. Zenaidee, M. A. & Donald, W. A. Extremely supercharged proteins in mass spectrometry: profiling the pH of electrospray generated droplets, narrowing charge state distributions, and increasing ion fragmentation. *Analyst* **140**, 1894–1905 (2015).

26. Peters, I., Metwally, H. & Konermann, L. Mechanism of Electrospray Supercharging for Unfolded Proteins: Solvent-Mediated Stabilization of Protonated Sites During Chain Ejection. *Anal. Chem.* **91**, 6943–6952 (2019).

27. Going, C. C., Xia, Z. & Williams, E. R. New supercharging reagents produce highly charged protein ions in native mass spectrometry. *Analyst* **140**, 7184–7194 (2015).

28. Foley, E. D. B. et al. On the mechanism of protein supercharging in electrospray ionisation mass spectrometry: Effects on charging of additives with short- and long-chain alkyl constituents with carbonate and sulphite terminal groups. *Anal. Chim. acta X* **1**, 100004 (2018).

29. Grimm, R. L. & Beauchamp, J. L. Evaporation and Discharge Dynamics of Highly Charged Multicomponent Droplets Generated by Electrospray Ionization†. *J. Phys. Chem. A* **114**, 1411– 1419 (2009).

30. Teo, C. A. & Donald, W. A. Solution additives for supercharging proteins beyond the theoretical maximum proton-transfer limit in electrospray ionization mass spectrometry. *Anal. Chem.* **86**, 4455–4462 (2014).

31. Sterling, H. J. et al. Effects of supercharging reagents on noncovalent complex structure in electrospray ionization from aqueous solutions. *J. Am. Soc. Mass Spectrom.* **21**, 1762– 1774 (2010).

32. Iavarone, A. T. & Williams, E. R. Mechanism of Charging and Supercharging Molecules in Electrospray Ionization. *J. Am. Chem. Soc.* **125**, 2319–2327 (2003).

33. Siu, K. W. M., Guevremont, R., Le Blanc, J. C. Y., O'Brien, R. T. & Berman, S. S. Is droplet evaporation crucial in the mechanism of electrospray mass spectrometry? *Org. Mass Spectrom.* **28**, 579–584 (1993).

34. Douglass, K. A. & Venter, A. R. Investigating the role of adducts in protein supercharging with sulfolane. *J. Am. Soc. Mass Spectrom.* **23**, 489–497 (2012).

35. Reading, E. et al. The Role of the Detergent Micelle in Preserving the Structure of Membrane Proteins in the Gas Phase. *Angew. Chemie Int. Ed.* **54**, 4577–4581 (2015).

36. Hardy, D., Desuzinges Mandon, E., Rothnie, A. J. & Jawhari, A. The yin and yang of solubilization and stabilization for wild-type and full-length membrane protein. *Methods vol.* **147**, 118–125 (2018).

37. Laganowsky, A., Reading, E., Hopper, J. T. S. & Robinson, C. V. Mass spectrometry of intact membrane protein complexes. *Nat. Protoc.* **8**, 639–51 (2013).

38. Kundlacz, T., Bender, J. & Schmidt, C. Effects of non-ionic and zwitterionic detergents on soluble proteins during native mass spectrometry experiments. *Int. J. Mass Spectrom.* **468**, 116652 (2021).

39. Schnier, P. D., Gross, D. S. & Williams, E. R. On the maximum charge state and proton transfer reactivity of peptide and protein ions formed by electrospray ionization. *J. Am. Soc. Mass Spectrom.* **6**, 1086–1097 (1995).

40. Barrera, N. P., Di Bartolo, N., Booth, P. J. & Robinson, C. V. Micelles protect membrane complexes from solution to vacuum. *Science.* **321**, 243–246 (2008).

41. Donor, M. T., Ewing, S. A., Zenaidee, M. A., Donald, W. A. & Prell, J. S. Extended Protein Ions Are Formed by the Chain Ejection Model in Chemical Supercharging Electrospray Ionization. *Anal. Chem.* **89**, 5107–5114 (2017).

42. Sterling, H. J. & Williams, E. R. Origin of Supercharging in Electrospray Ionization of Noncovalent Complexes from Aqueous Solution. *J. Am. Soc. Mass Spectrom.* **20**, 1933– 1943 (2009).

43. Loo, J. A., Loo, R. R. O., Udseth, H. R., Edmonds, C. G. & Smith, R. D. Solvent-induced conformational changes of polypeptides probed by electrospray-ionization mass spectrometry. *Rapid Commun. Mass Spectrom.* **5**, 101–105 (1991).

44. Smith, D. P. et al. Deciphering Drift Time Measurements from Travelling Wave Ion Mobility Spectrometry-Mass Spectrometry Studies. *Eur. J. Mass Spectrom.* **15**, 113–130 (2009).

45. Hunter, E. P. L. & Lias, S. G. Evaluated Gas Phase Basicities and Proton Affinities of Molecules: An Update. *J. Phys. Chem. Ref. Data* **27**, 413 (2009).
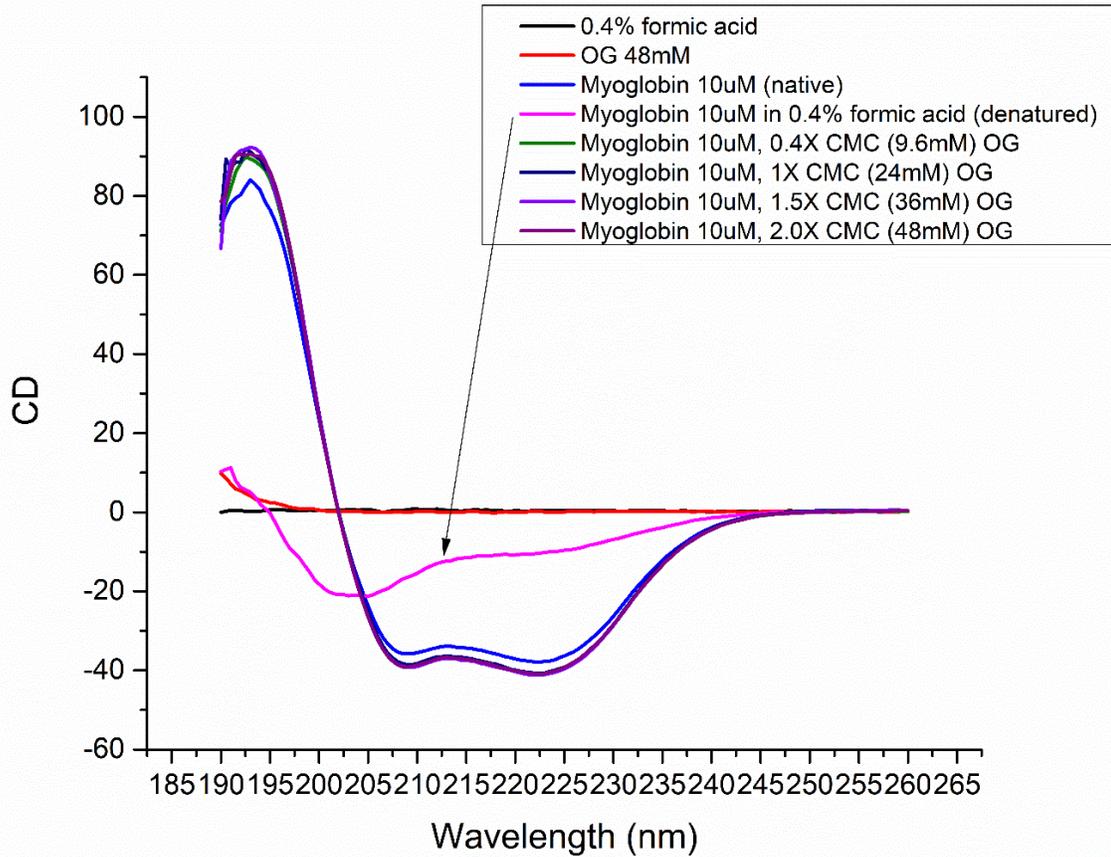
**Supplementary Figure 1. Circular dichroism screening of myoglobin treated with OG.**
10µM myoglobin was dissolved in 200mM ammonium acetate with 9.6mM to 48mM OG or 0.4% formic acid. The samples were analyzed with a Jasco J-175 spectropolarimeter in a 1mm cuvette. The degree of ellipticity at 222nm was recorded. Recorded curves of myoglobin treated with OG are virtually indistinguishable from native myoglobin's curve, which further validates that supercharging induced by detergents are not caused by denaturation.
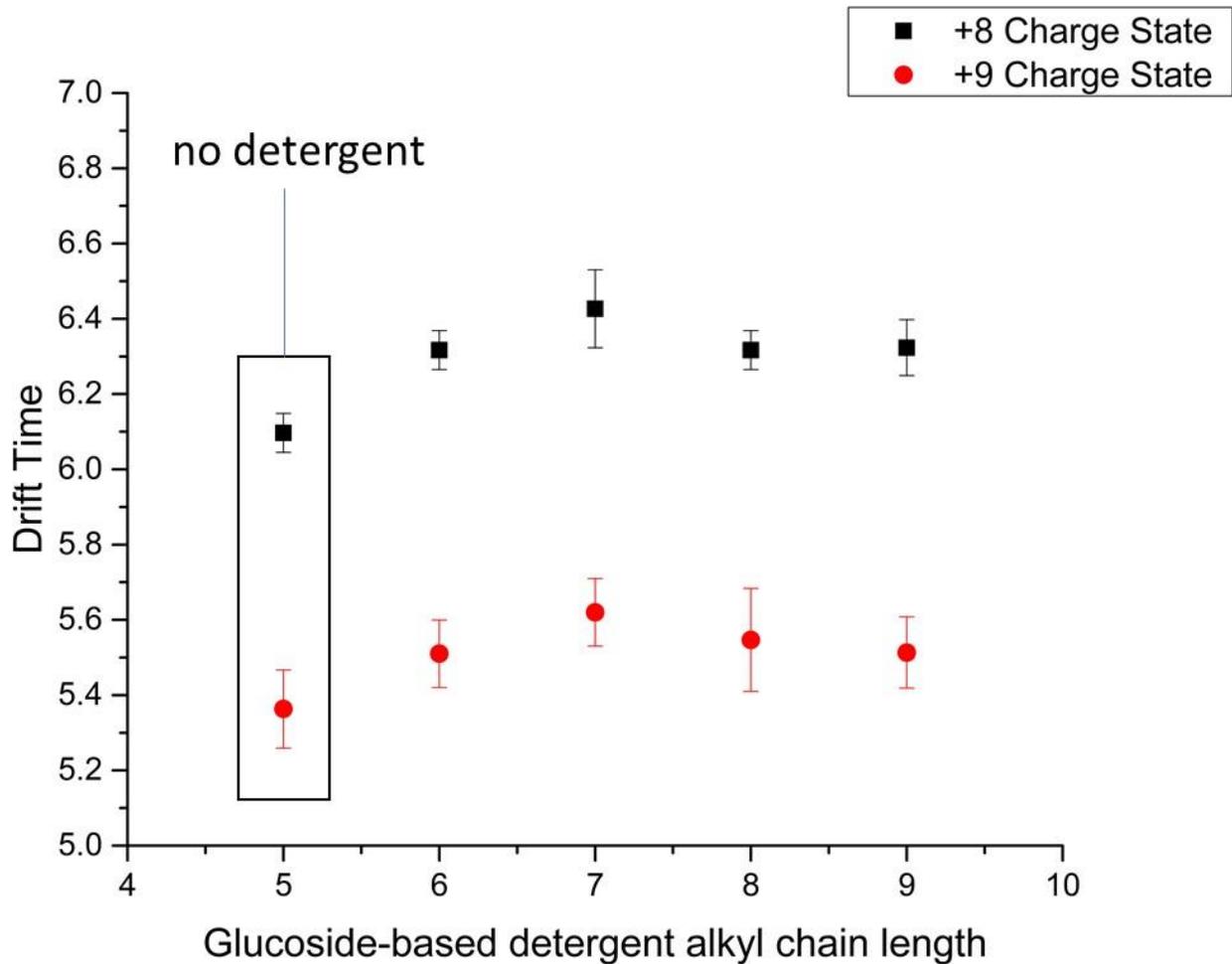
**Supplementary Figure 2. Native lysozyme drift time vs alkyl chain length of the glucoside-based detergents.**

Drift times of two major charge states of native lysozyme observed during glucoside-based detergent supercharging experiments were recorded. All measurements were done in triplicate and the error bars indicate the standard deviation. No significant change in drift time was detected, which implies that the charge increase is not caused by denaturation of the protein.
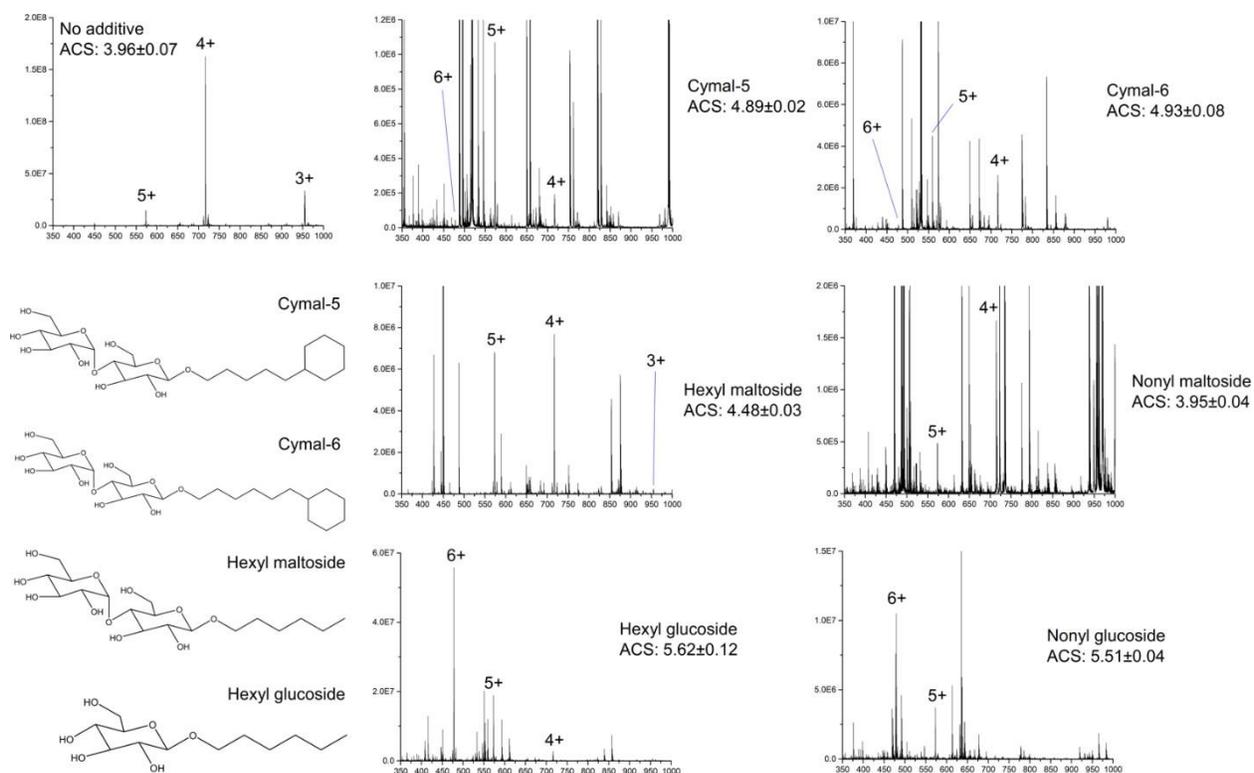
**Supplementary Figure 3. Representative spectra of 3X-flag-tag peptide treated with non-ionic saccharide-based detergents and chemical structures of the detergents screened for their charge shifting capacity.**

3x-Flag-tag peptide in a denaturing condition was treated with non-ionic saccharide-based detergents of varying alkyl chain length and polar head groups to investigate how chemical properties of the detergent affect the resulting average charge state. Representative spectra (from triplicate measurements) are shown here along with the major charge states the peptide. Representative chemical structures of the detergents screened in this study are also shown. Glucoside-based detergents tend to show better supercharging capacity and result in less interference with the target analyte ionization. (See **Figure 1** in the main text for more details.)
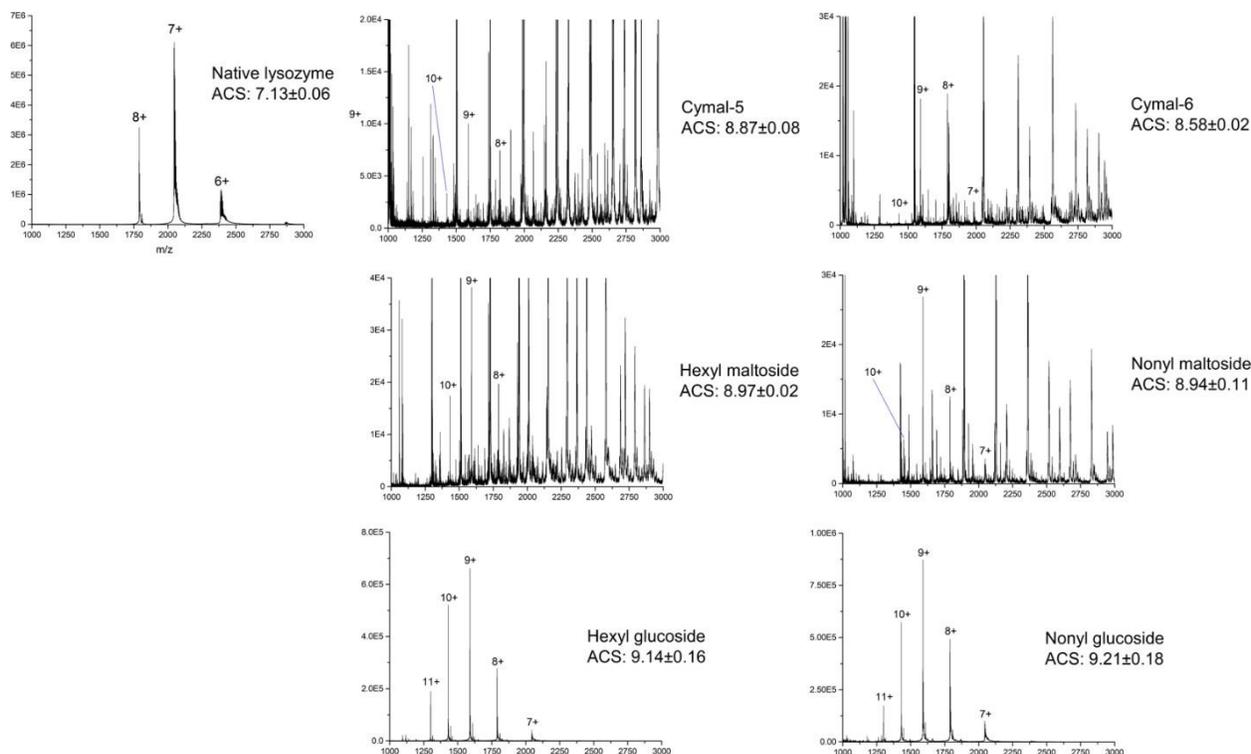
**Supplementary Figure 4. Representative spectra of lysozyme treated with non-ionic saccharide-based detergents.**

Lysozyme in a native condition was treated with non-ionic saccharide-based detergents of varying alkyl chain length and polar head groups to investigate how chemical properties of the detergent affect the resulting average charge state. Representative spectra out of triplicate measurements are shown here along with the major charge states of native lysozyme. Similar to the 3x-flag-tag peptide, glucoside-based detergents tend to show better supercharging capacity and result in less interference with the target analyte ionization. (See **Figure 1** in the main text for more details.)

| | 3x-Flag-tag peptide | | | Lysozyme | | |
|---|---|---|---|---|---|---|
| Additive | Highest observed charge state | Most abundant charge state | ACS | Highest observed charge state | Most abundant charge state | ACS |
| None (ctrl) | 5+ | 4+ | 3.96±0.07 | 8+ | 7+ | 7.13±0.06 |
| Cymal-5 | 6+ | 5+ | 4.89±0.02 | 10+ | 9+ | 8.87±0.08 |
| Cymal-6 | 6+ | 5+ | 4.93±0.08 | 10+ | 9+ | 8.58±0.02 |
| Hexyl Maltoside | 5+ | 5+ | 4.48±0.03 | 10+ | 9+ | 8.97±0.02 |
| Nonyl Maltoside | 5+ | 5+ | 3.95±0.34 | 10+ | 9+ | 8.94±0.11 |
| Hexyl Glucoside | 6+ | 6+ | 5.62±0.12 | 12+ | 9+ | 9.14±0.16 |
| Heptyl Glucoside | 6+ | 6+ | 5.51±0.04 | 11+ | 9+ | 8.76±0.2 |
| Octyl Glucoside | 6+ | 5+ | 5.29±0.21 | 11+ | 9+ | 9.01±0.01 |
| Nonyl Glucoside | 6+ | 6+ | 5.51±0.04 | 12+ | 9+ | 9.21±0.18 |
| Decyl Glucoside | 6+ | 6+ | 5.54±0.06 | *N/A | N/A | N/A |

**Supplementary Table 1. Highest charge state, most abundant charge state, and average charge state observed for 3x-flag-tag peptide and lysozyme solubilized in a denaturing condition and native condition, respectively.**

*Decyl glucoside is insoluble at 12mM with 200mM ammonium acetate. Acidic condition necessary for the experiment. Native lysozyme supercharging experiment is not possible with DG at 12mM concentration.

| Additive | Gas phase basicity (kJ/mol) |
|---|---|
| Acetamide | 832.6 |
| Benzamide | 861.2 |
| N, N-dimethyl acetamide | 877.0 |
| formamide | 791.2 |
| 4-Methoxybenzamide | 869.4 |
| 4-Nitrobenzamide | 814.4 |
| Propionamide | 845.3 |
| Chloroacetamide | N/A |
| N, N-dimethyl foramide | 856.6 |
| Chloroacetonitrile | 715.1 |
| 3-Nitrobenzonitrile | 750.7 |
| 3-Nitrophenylacetonitrile | N/A |
| 4-Nitrophenylacetonitrile | N/A |
| Propionitrile | 763.0 |
| Succinonitrile | N/A |
| Malonitrile | 694.1 |
| 4-pyridine carbonitrile | 848.8 |
| 2-pyridine carbonitrile | 841.0 |

**Supplementary Table 2. Gas phase basicity of amides and nitriles screened.** (See ref 45 of the main text for more details)

| Name | Structure | ACS |
|---|---|---|
| **3-nitrobenzylalcohol (m-NBA)** |  | **11.39** |

| | | |
|---|---|---|
| **1,2-butylene carbonate** | | **11.81** |
| **Sulfolane** | | **10.95** |
| **Formamide** | | **8.93** |
| **Acetamide** | | **8.18*** |
| **N,N-dimethylacetamide (DMA)** | | **8.03** |
| **N,N-dimethylformamide (DMF)** | | **8.44** |
| **Propionamide** | | **8.27** |
| **Chloroacetamide** | | **9.02** |
| **Benzamide** | | **9.80** |
| **4-methoxybenzamide** | | **8.78** |
| **4-nitrobenzamide** | | **9.99**** |

| | | |
|---|---|---|
| **Propionitrile** | | **8.85** |
| **Malononitrile** | | **9.19** |
| **Chloroacetonitrile** | | **9.10** |
| **Succinonitrile** | | **10.17**\*\* |
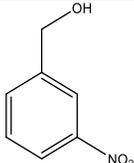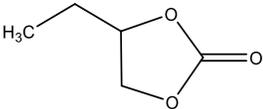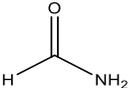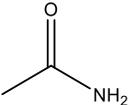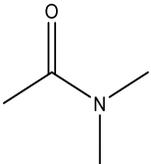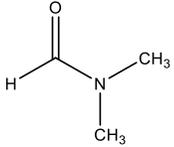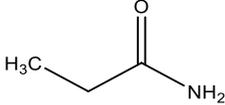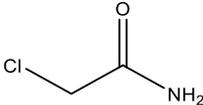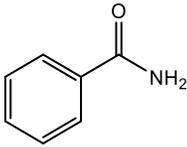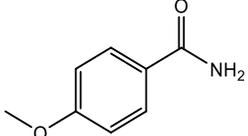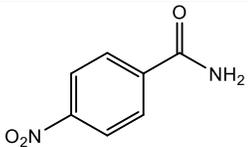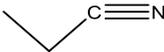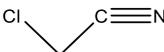
**Supplementary Table 3. Chemical structure of conventional supercharging agents, amides, and nitriles tested for their sub/supercharging capacity and average charge state of lysozyme after their treatment.**

All chemical agents were treated at 0.2% w/v. Lysozyme was solubilized in 50% water, 49.9% acetonitrile, 0.1% formic acid. ACSs are the average values from triplicate measurements (see Material and Methods for details).

**Chapter 3**

**Native top-down MS analysis of membrane protein-lipid complexes**

Wonhyeuk Jung†, Pascal Egea‡, Rachel R. Ogorzalek Loo† and Joseph A. Loo†‡

†Department of Chemistry and Biochemistry, University of California, Los Angeles, California 90095, United States

‡Department of Biological Chemistry, David Geffen School of Medicine, University of California, Los Angeles, California 90095, United States

The following is the preliminary manuscript draft being prepared for a submission.

**Abstract**

Membrane proteins are a unique class of biomolecules that are responsible for a variety of physiological phenomena and are also important drug targets. However, it is difficult to perform structural studies of membrane proteins using conventional techniques such as x-ray crystallography, nuclear magnetic resonance spectroscopy, and cryo-electron microscopy due to their hydrophobicity. With the advent of electrospray ionization, which is an ionization technique that does not fragment the target analyte, and mass spectrometers capable of transmitting large biomolecules without harming their structure, native mass spectrometry (MS) has recently emerged as a complementary structural analysis technique for membrane proteins.

For native MS analysis of membrane proteins solubilized in a detergent micelle, the proteins can be ejected out of the micelles by inducing collisional activation with a non-reactive gas such as

nitrogen. The ejected membrane proteins, in turn, can be subjected to further fragmentation analysis to gather structural information. This approach, termed native top-down MS analysis, has successfully been applied to large proteins and protein complexes. Here, an ammonium channel (AmtB) of *Escherichia coli* (*E. coli*), a homotrimeric channel protein, and its lipid-bound forms are subjected to native top-down MS analysis. The study demonstrates how fragmentation pattern differences between AmtB and AmtB-lipid complexes can be used to investigate structural effects of lipids on the proteins. The study also highlights how lipid-bound fragments can directly provide information on the lipid binding sites of membrane proteins.

## Introduction

Membrane proteins' interaction with their cognate ligands are responsible for various physiological phenomena such as cell-cell interaction, signal transduction, solute transportation, and energy conversion etc.[1] Membrane proteins account for 60% of known druggable targets in the cell.[2] However, they represent only 4.6% of the PDB despite constituting between 20% to 30% of proteome.[3] They are a difficult class of proteins for structural analysis and high-resolution studies face many challenges.

Recombinant expression and purification of sufficient amounts of membrane proteins is difficult.[4] Since they are hydrophobic proteins, crystallography and electron microscopy (EM) analysis require the appropriate choice of detergent micelles, lipid cubic phase, or membrane mimetic.[5] Nuclear magnetic resonance (NMR) studies tend to have difficulties due to the detergent micelle formation around the membrane proteins, which effectively doubles the size of the target analyte.[6] Detergent micelle's innate heterogeneity makes classification and alignment difficult for accurate image averaging for cryo-EM.[7] Membrane proteins are often flexible and unstable,

which hinders effective crystallization.[4] Thus, there is a need for analytical techniques that can handle difficult samples while complementing other conventional methods for high-resolution structural studies. Native mass spectrometry (MS), in which the proteins or protein complexes of interest are ionized via electrospray ionization (ESI) from a solution at physiological pH,[8] can potentially fulfill this role.

With the advent of ESI, which is a gentle ionization technique that can confer multiple charges onto the target analytes without analyte fragmentation,[9] and mass spectrometers that are able to transmit large biomolecules without harming analyte structures,[10–12] native MS has successfully been applied to membrane proteins.[13] During native MS of detergent-solubilized membrane proteins, the complex is activated in the source region via collisions (in-source activation) with a non-reactive gas such as nitrogen. The activation level experienced by the micelle-membrane protein complex can be modulated to eject out the membrane proteins without harming their structure.[14,15] The ejected membrane proteins, in turn, can be directly mass analyzed or be subjected to another round of activation inside the collision cell, where collisionally activated dissociation (CAD) can occur to gain primary structure information. For the Q Exactive UHMR Hybrid Quadrupole-Orbitrap MS System (from here the instrument will be simply referred to as 'UHMR') used for work presented here, a specific kind of CAD called higher-energy C-trap dissociation (HCD)[16] can be applied (**Supplementary Figure 1**). The approach of fragmenting target analytes after ionization under native conditions has been termed as native top-down MS analysis.[17]

Native top-down MS has successfully been applied to identify regions of multimeric proteins responsible for noncovalent interactions, sites of post-translational modifications (PTMs), and ligand binding sites.[18–21] For membrane proteins, after a detailed protocol for membrane protein and lipid sample preparation for MS has been established,[13] successful analysis of the structural effect of lipid binding,[22] endogenous ligand and drug binding to G-protein coupled receptors,[23]

global structural change to channels during gating process,[24] interplay of nucleotides, lipids, and drugs binding to a multidrug resistance efflux pump,[25] and copper centers of membrane-bound methane monooxygenase[18] has been performed. However, to date, the fragmentation pattern difference between the membrane protein and membrane protein-lipid complexes and whether lipid-bound fragments (induced in the mass spectrometer) can be used to directly identify lipid-binding sites has not been examined in detail.

The ammonium channel (AmtB) of *Escherichia coli* (*E. coli*) is a homotrimer channel protein that is necessary to maintain rapid cell growth at low ambient ammonium concentration.[26] Laganowsky et al. found through ion mobility analysis that phosphatidylglycerol (PG) and cardiolipin (CL) cause significant gas phase structure stabilization for AmtB.[22] The study conducted a collision induced unfolding (CIU) experiment in which they compared the collision energy necessary to induce unfolding of AmtB or AmtB-lipid complexes. Both AmtB-phosphatidylglycerol (PG) and AmtB-cardiolipin (CL) complexes were able to withstand significantly more collision energy before denaturing compared to apo-AmtB, indicating that these two lipids induce structure stabilization. These membrane protein-lipid complexes make attractive targets for a proof-of-concept study since crystal structures are available for AmtB-PG (PDB 4NH2),[22] and AmtB-TopFluor CL (PDB 6B21)[27]. Thus, cross-verification of the structural data acquired from native top-down MS analysis can be performed.

In this study, AmtB and AmtB-PG complex were subjected to native top-down MS analysis. The apo-form and 1- and 2-lipid bound forms were isolated via quadrupole mass filter and then subjected to HCD fragmentation; the resulting fragmentation pattern differences between AmtB and AmtB-PG complex were investigated. The resulting data show the lipid interaction sites that confers regional gas phase stability. Specifically, Val 236, Ala 237, Val 281, Met 328, and Phe 332, enjoyed protective effect from PG binding, which was in good agreement with the potential lipid binding sites suggested by the crystal structure of AmtB-PG. Interestingly, Leu 243, Phe

247, Gly 248, and Ala 248 also was protected by PG interaction. However, the crystal structure did not indicate any lipid binding to this particular region of the protein.

The native top-down MS analysis workflow involving CAD typically involves matching theoretical mass lists generated from the sequence of the protein of interest to the observed mass list generated from the fragmentation spectra. Conventionally, only the terminal fragments, i.e., the fragments that contains the either N-terminus or the C-terminus, are considered for data interpretation. However, it is known that including internal fragments into the assignments can increase the sequence coverage and the total number of ions assigned from the raw spectra.[28,29] Recently, a software for internal fragment detection, Clips-MS has been developed in our lab,[30] and was utilized to interpret the fragmentation spectra of three test proteins, cytochrome c, myoglobin, and carbonic anhydrase II. In this study, Zenaidee et al. observed that the inclusion of internal fragments in the analysis resulted in approximately 15−20% more sequence coverage, with no less than 85% sequence coverage obtained.[29]

To take advantage of these recent developments, Clips-MS was used to analyze the AmtB and AmtB-PG complex fragmentation spectra. Before including internal fragments, about 41% of the raw spectra deconvoluted mass list was able to be assigned to either N-terminus or C-terminus fragments. The proportion of the mass list assigned expanded to 55% upon inclusion of internal fragments, highlighting how internal fragment interpretation can increase the amount of information gained from a single native top-down MS experiment. In addition, internal fragments with lipids bound were found. This is in good agreement with the previously published study that found that after CAD induced fragmentation, some non-covalent interactions can survive in the gas phase, leading to direct confirmation of the ligand binding sites.[31] Here, a total of 11 PG-bound internal fragments were identified; 9 were found to be in the region of encompassing residues 200-275, indicating that this region is also where lipid interactions are primarily taking place. The results from this study highlight how native top-down MS can be applied to

80

membrane proteins and membrane protein-lipid complexes to understand structural effects of lipid binding and to identify lipid-binding sites.

## Materials and Methods

**Membrane protein preparation for MS analysis.** AqpZ was expressed and purified following the previously published protocol.[32] The stock solution of ~120 µM of AqpZ solubilized in 200 mM NH4OAc, 40 mM OG was diluted to 4 µM with the same solvent and was used without further purification. AmtB was expressed and purified following the previously published protocol.[33] In brief, the MBP-AmtB fusion protein was expressed and purified by NiNTA chromatography; the expression strain was *E. coli* C41 (DE3) and Terrific broth was used for cell growth with protein expression induced at an OD600 of 1.0. For proteolytic removal of the MBP moiety from the MBP-AmtB fusion protein the peak fractions from the NiNTA column were pooled and 2-mercaptoethanol was added to 5 mM. His-tagged TEV protease was added, and the solution was dialyzed against 150 mM NaCl, 10% glycerol, 40 mM OG, and 50 mM Tris, pH 7.4 overnight at 4°C. The following day, the solution was centrifuged and the supernatant decanted and diluted 1:1 with SEC Buffer (200 mM NH4OAc pH 6.9, 40 mM OG). The sample was concentrated with a 100 kDa MWCO concentrator to ~2.5 mL and injected onto a HiLoad 16/600 Superdex 200 column (GE Healthcare Life Sciences) equilibrated in SEC Buffer. Peak fractions were analyzed by SDS-PAGE and fractions containing pure AmtB were pooled. The resulting stock solution of ~150 µM of AmtB solubilized in 200 mM NH4OAc, 40 mM OG was diluted to 4 µM with the same solvent and was used without further purification.

**Lipid preparation for MS analysis and complex formation with membrane proteins.** Stock lipid solution was prepared with 14:0 phosphatidylglycerol (sodium salt) (Avanti Polar Lipids,

840445) and 14:0 cardiolipin (ammonium salt) (Avanti Polar Lipids, 710332) following the previously published protocol.[13] The final stock solutions was prepared by resuspending with 200mM NH4OAc, 40mM OG. The stock solutions were directly spiked into the membrane protein solution at appropriate volume to achieve 1:20 protein: lipid ratio for PG, and 1:30 protein: lipid ratio for CL. After the lipid treatment, the solution was incubated for an hour in ice before MS analysis.

**UHMR mass spectrometry.** 4 μM AmtB and AqpZ with/without PG and CL treated were loaded into borosilicate capillaries (Warner Instruments, 30-0042) that was pulled in house with (Sutter Instrument, P-1000 Micropipette Puller) and then coated with either gold or platinum for 4 minutes (Anatech, Hummer 6.2 Sputtering System). The nanoESI (nESI) was performed at a flow rate of 10-40 nL/min through a nanospray ion source and analyzed with Q Exactive UHMR Hybrid Quadrupole-Orbitrap MS System (Thermo Fisher Scientific). Key parameters during membrane protein and membrane protein-lipid complex native MS were: spray voltage at 1.3 kV to 1.8 kV, in-source CID at 20 eV, capillary temperature at 100 °C, source DC offset at 25 V, in-source trapping on with desolvation voltage at -10 V to -25 V, injection flatpole RF amplitude at 550, bent flatpole RF amplitude at 550, C-Trap RF amplitude at 2100, injection flatpole DC at 0V, inter flatpole lens at 14 V, bent flatpole DC at 8 V, and trapping gas pressure at 7. Key parameters for membrane protein and membrane protein native top-down analysis were: spray voltage at 1.3 kV to 1.8 kV, in-source CID at 20 eV, capillary temperature at 100 °C, source DC offset at 25 V, in-source trapping on with desolvation voltage at -10 V to -25 V, injection flatpole RF amplitude at 610, bent flatpole RF amplitude at 550, C-Trap RF amplitude at 2100, injection flatpole DC at 0 V, inter flatpole lens at 15 V, bent flatpole DC at 8 V, HCD CE at 150 V to 300 V, (see main text and **Supplementary Figure 2** for more details) and trapping gas pressure at 2. All measurements were done in triplicates and 100 scans were averaged per spectra.

**Data Analysis.** Unidec was used for native MS spectra deconvolution.[34] Thermo's BioPharma Finder Software was used for native top-down MS spectra deconvolution and the major parameters were: S/N threshold at 4, relative abundance threshold at 1%, and mass accuracy threshold at 7.5 ppm. For fragmentation spectra, 1,000 – 5,000 m/z region was targeted for deconvolution. The deconvoluted mass list was then fed into the Clips-MS software. The error for matching was set at 7.5ppm, FASTA sequences of AmtB and AqpZ (provided in the supplementary document) were entered. For AqpZ, formylation at the N-terminus was accounted for. For AmtB, there was no modification. The fragments searched for were a-fragment, x-fragment, b-fragment, y-fragment, and by-fragment. For unlocalized modification, PG's molecular weight was entered as 665.44Da, and CL's molecular weight was entered as 1238.822Da. All internal fragments and internal fragments bound with PG or CL were manually verified.

## Results and Discussion

### Native MS of AmtB and AmtB-PG complexes

AmtB is an ammonium channel protein of *E. coli* that is commonly used as a test membrane protein for native MS analysis.[13,14,27] It is well established that cognate lipid interactions are crucial for the structural stability of this homotrimeric protein.[22,35,36] Among the lipids tested by a previously published study, PG was the only lipid that conferred increasing stability, as multiple binding took place.[22] Thus, whether structural differences between the AmtB and AmtB-PG complex can be probed through a CAD approach was investigated.

Native MS spectra of AmtB and AmtB-PG complex acquired with UHMR is shown in **Figure 1.**

AmtB was solubilized in 200mM NH4OAc and 40mM octyl-β-D-glucoside (OG). OG is a

commonly used non-ionic saccharide-based detergent that tends to confer relatively higher

charge to the analytes for ESI-MS when added to the solution.[14,37] The higher charge states in

turn, can enhance the effectiveness of the fragmentation analysis.[38]

Once the instrument parameters for effective AmtB detection was established, the optimal

protein-to-lipid ratio was probed (data not shown). Too little lipid treatment results in minimal

AmtB-PG complex formation and too much lipid treatment hinders efficient ionization. An

optimal protein-lipid ratio at which 2 lipids bound from could be reliably detected at high enough

intensity for subsequent MS-isolation and fragmentation was determined to be 1 to 20 (AmtB

4µM, PG 80µM) (**Figure 1B**).


**Native top-down MS analysis of AmtB and AmtB-PG complexes**

To probe the correlation between HCD energy and the sequence coverage during native top-

down MS analysis of AmtB, AmtB was exposed to series of HCD energies from 150V to 300V

(**Supplementary Figure 2**). At 150V, the lowest major charge state, 20+, yields less than 2.5%

sequence coverage, making subsequent analysis suboptimal. At 275V, the sequence coverage

of the base peak, 21+, starts to decrease, indicating that fragmentation of fragments (i.e.,

internal fragment generation) is starting to predominate. Thus, fragmentation experiments for

AmtB and AmtB-PG complex were performed with HCD energies of 175 V, 200 V, 225 V, and

250 V.

**Figure 2** shows the example fragmentation spectra of AmtB-PG complex (1 lipid bound form,

HCD 225V). After the deconvolution (transition from mass-to-charge ratio to just mass), 125

fragments in total were detected. Out of 125 fragments, 51 could be assigned to either N-

terminus or C-terminus fragments, suggesting that there is a significant amount of potential

information being lost. To investigate whether deeper interpretation of the spectra is possible by

including internal fragments into the analysis, Clips-MS[30] was utilized.
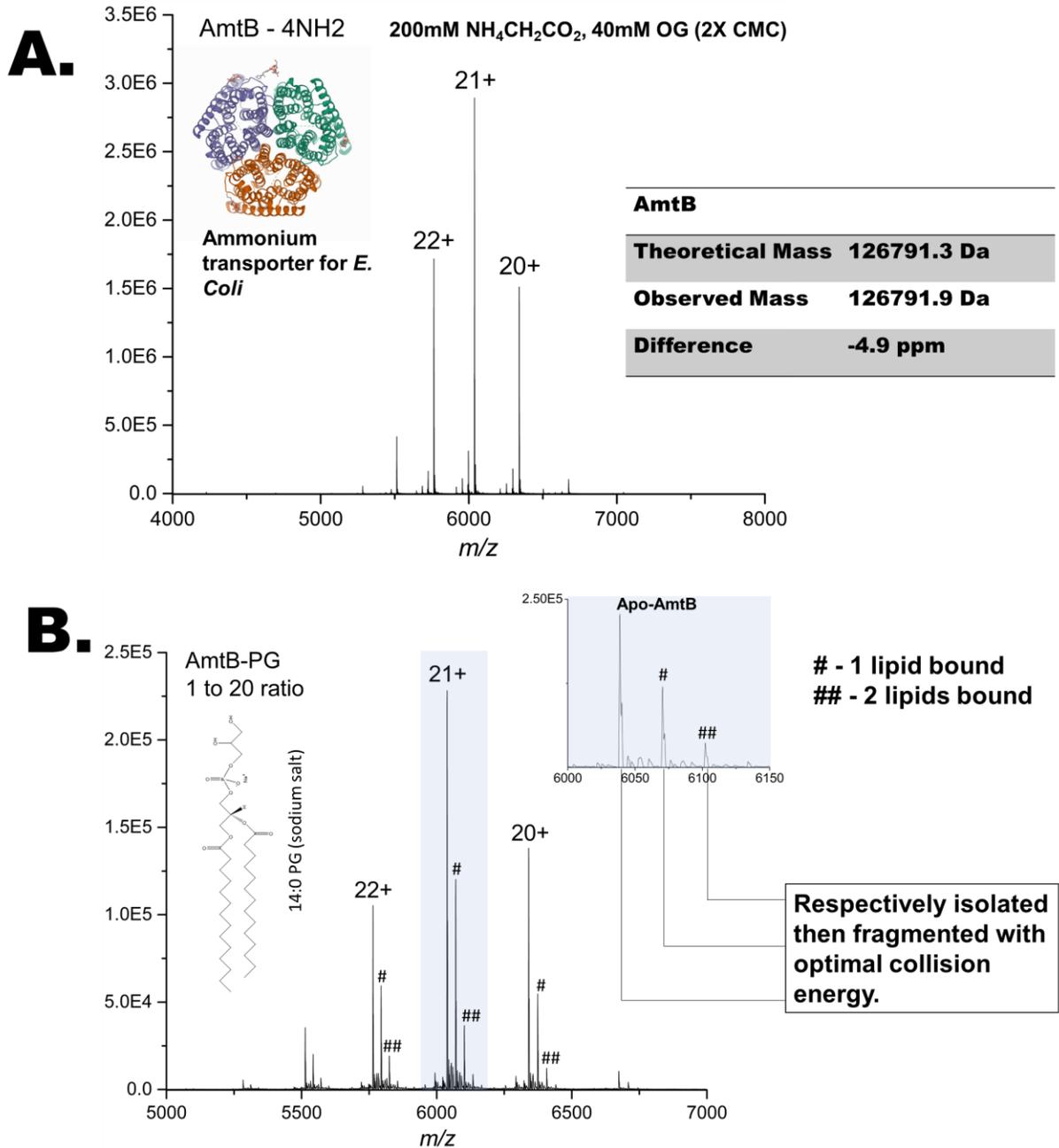


**Figure 1. Native MS analysis of AmtB and AmtB-PG complex.** A. Native MS spectrum of AmtB and its crystal structure is shown. Three major charge states are observed at a mass accuracy better than 5ppm. The protein was solubilized in 200mM NH4OAc and 40mM OG. B. Native MS spectrum of AmtB-PG complex and the chemical structure of PG is shown. Protein-

to-lipid ratio was at 1 to 20 and the three major charge states and their respective lipid bound forms were detected. Up to 3 lipids bound forms could be observed but the signal intensity for 3 lipids bound forms were too weak for subsequent analysis. Thus, 1 lipid bound form and 2 lipids bound from were isolated and subjected to fragmentation analysis.
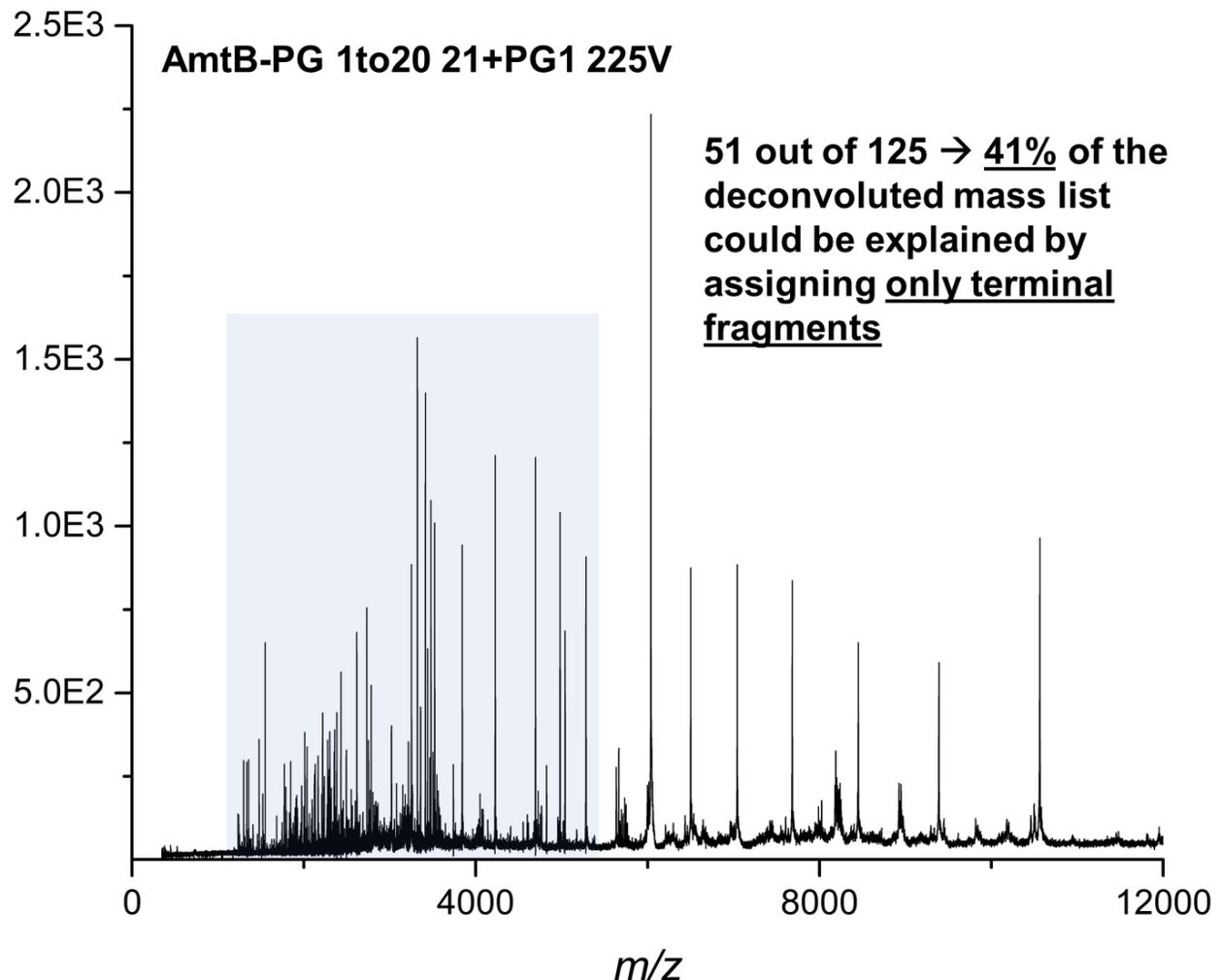


**Figure 2. The fragmentation spectra of AmtB-PG 1 lipid bound complex.** AmtB-PG complex was fragmented with energies ranging from 175V to 250V. At 225V, all three major charge states showed sequence coverage ranging from 10% to 15% when only the terminal fragments were accounted for. Deconvolution of the fragmentation spectra resulted in a total of 125 fragments of which 51 were matched to either N-terminus fragments or C-terminus fragments.

**Figure 3** shows the output of Clips-MS analysis. 45 fragments were assigned as internal

fragments in addition to the original 51 terminal fragments. Upon manual verification of the

86

internal fragments, 18 were confirmed to be true (**Supplementary Figure 3**). Thus, the

assignment increased from 51 out of 125 (41%) to 69 out 125 (55%). This increase in sequence

coverage is in good agreement with the previously published research by Zenaidee et al.[29] in

which inclusion of internal fragments resulted in approximately 15−20% more sequence

coverage for soluble proteins. Out of the 18 manually verified internal fragments, 11 were lipid-

bound fragments. When these lipid-bound fragments were separately mapped, it was clear that



**Fragment location**
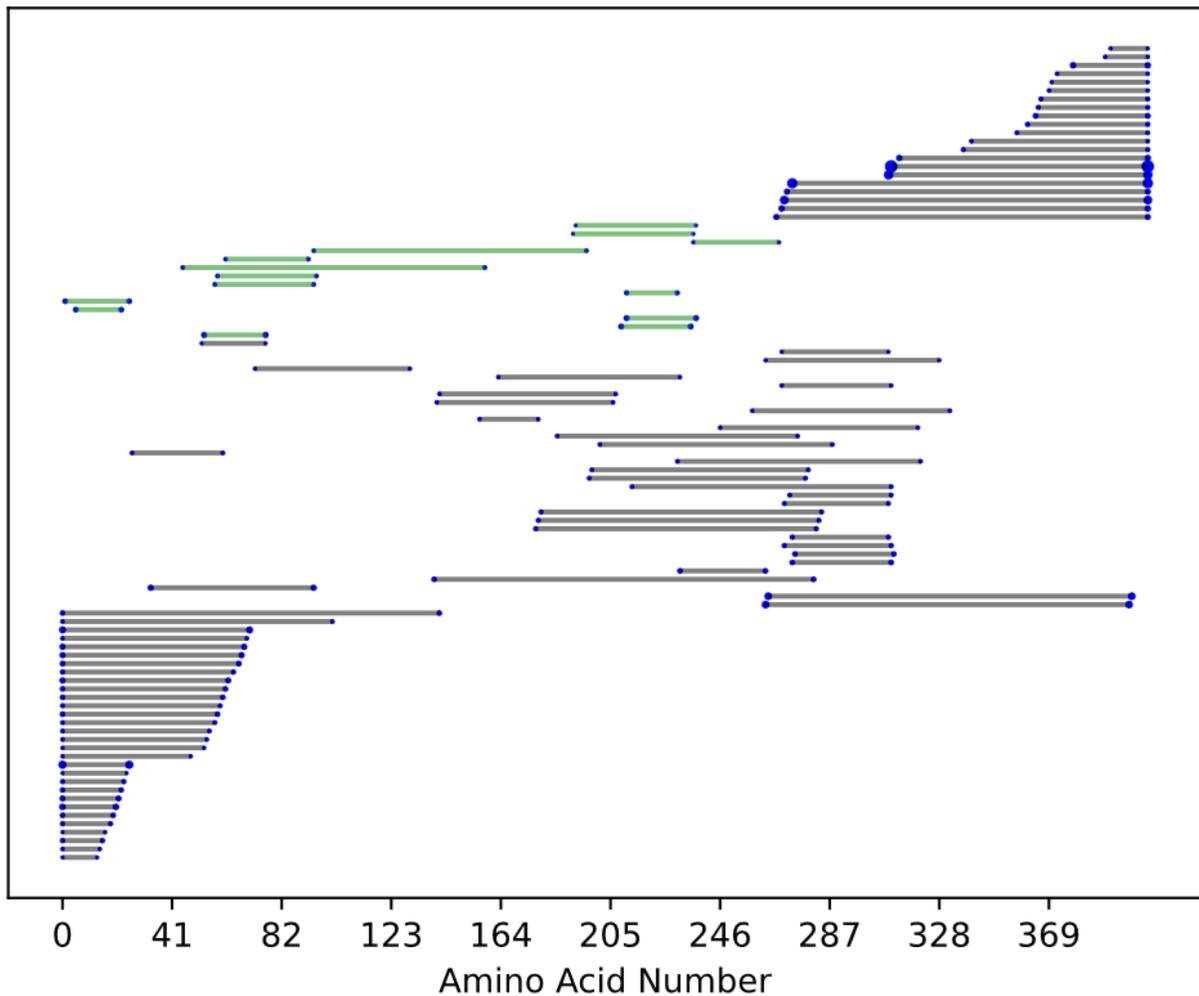
lipid-bound fragments were concentrated around the region spanned by residues 200-275

(**Figure 4**).

**Figure 3. Clips-MS analysis results of AmtB-PG 1 lipid bound complex.** Each line
represents a fragment. Out of the 125 fragments detected after the initial deconvolution, 51
were assigned as terminal fragments. Clips-MS analysis indicated that 45 additional fragments

could be assigned as internal fragments of which 14 were designated as lipid-bound fragments (highlighted in green). Upon manual verification of these internal fragments, 18 were confirmed to be true. Out of the 18 verified internal fragments, 11 were lipid-bound fragments.
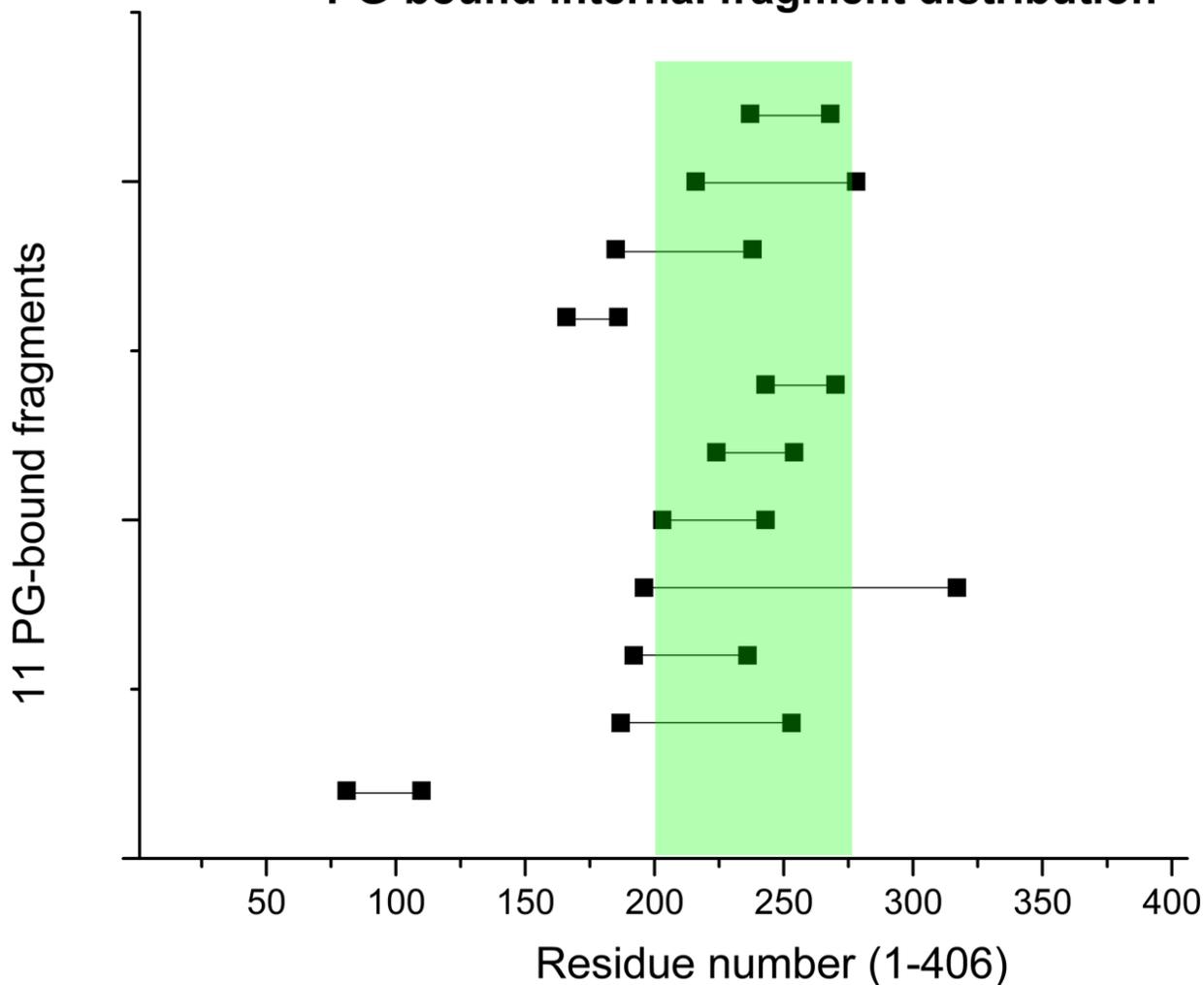


**Figure 4. Lipid-bound fragments are concentrated along residues 200-275.** Each line represents lipid-bound internal fragments that were manually verified. A total of 11 lipid-bound internal fragments were confirmed. Out of 11 fragments, 9 were mapped from residues 200 to 275, indicating that this zone is likely to be in direct contact with PG.

### 3D mapping of the native top-down MS analysis results

The fragmentation patterns observed for AmtB and AmtB-PG complexes were mapped onto the

crystal structure of AmtB-PG complex (PDB 4NH2). As shown in **Figure 5**, significant

fragmentation pattern differences were observed for 2 lipid binding sites out of 6 potential lipid

binding sites suggested by the crystal structure. Internal fragmentations that were observed at

Val 236, Ala 237, Val 281, Met 328, and Phe 332 (highlighted in blue), which runs along the

groove of the AmtB, did not take place for AmtB-PG. Val 50, Thr 51, Leu 54, Cys 56, and Leu

58, in contrast, underwent terminal fragmentation (highlighted in red) for both AmtB and AmtB-

PG, which validates that region that are minimally affected by the lipid binding.

Interestingly, Leu 243, Phe 247, Gly 248, and Ala 248, also underwent internal fragmentation,

but only for apo-AmtB. However, the crystal structure did not indicate that lipid interaction is

taking place at that region. As mentioned above, 11 lipid-bound internal fragments were

identified and 9 out of the 11 fragments were within the residue 200-275 region. Thus, native

top-down MS analysis suggests that lipid interaction is taking place at this region, although the

crystal structure does not indicate lipid binding.
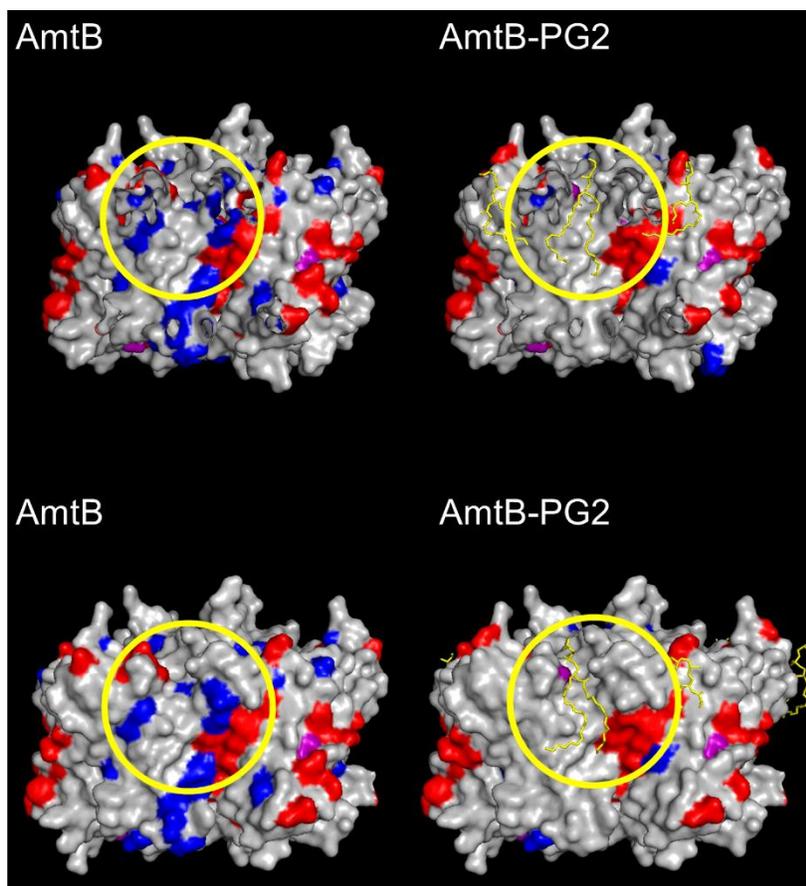


**Figure 5. MS fragmentation pattern mapped on to the crystal structure of AmtB-PG complex.** AmtB-PG complex crystal structure (PDB 4NH2) is colored red where terminal fragmentation takes place and blue where internal fragmentation is located. The PG molecules are represented in yellow. There are a total of 6 potential binding sites that were suggested by the crystal structure. Out of 6 binding sites, 2 showed significant fragmentation pattern difference, indicating that lipids are likely to preferably bind these 2 sites.

## Conclusion

The advancement of protein MS in general, and specifically the development of ESI, in which large biomolecules are ionized without fragmenting, and the instruments that can transmit these ions without harming their structure has led to the development of native MS. Native MS approaches can be combined with fragmentation techniques to gather structural information of the target analytes of interest. Termed native top-down MS analysis, this approach has been used to successfully analyze proteins and protein complexes.

Here, native top-down MS analysis was performed on AmtB and AmtB-PG complexes. The fragmentation pattern difference indicates that out of 6 potential lipid binding sites suggested by the crystal structure, 2 sites confer significant protective effect from top-down MS fragmentation. Specifically, internal fragmentations that were observed at Val 236, Ala 237, Val 281, Met 328, and Phe 332 for apo-AmtB did not take place for AmtB-PG, indicating that these are major sites of interaction. Interestingly, Leu 243, Phe 247, Gly 248, and Ala 248, also underwent internal fragmentation only for the apo-AmtB. However, these sites were not indicated by the crystal structure as a lipid interaction site. When the lipid-bound internal fragments were screened, it was observed that majority of the lipid-bound fragments are positioned along residues 200-275. Another membrane protein, aquaporin-Z, a homotetramer water channel protein, and its complex with lipids were also measured by native top-down MS (**Supplementary Figure 4**); the data thus far is consistent with the AmtB-PG work, i.e., it highlights how native top-down MS analysis of membrane proteins can be used to cross-verify crystal structures and gain new insights into the non-covalent interactions that stabilize membrane protein structures.

## References

1. Engel, A. & Gaub, H. E. Structure and Mechanics of Membrane Proteins. *Annu. Rev. Biochem.* **77**,127-48 (2008)

2. Overington, J. P., Al-Lazikani, B. & Hopkins, A. L. How many drug targets are there? *Nat. Rev. Drug Discov.* **5**, 993–996 (2006).

3. Allen, K. N., Entova, S., Ray, L. C. & Imperiali, B. Monotopic Membrane Proteins Join the Fold. *Trends in Biochemical Sciences vol.* **44**, 7–20 (2019).

4. Carpenter, E. P., Beis, K., Cameron, A. D. & Iwata, S. Overcoming the challenges of membrane protein crystallography. *Current Opinion in Structural Biology vol.* **18**, 581–586 (2008).

5. Hendrickson, W. A. Atomic-level analysis of membrane-protein structure. *Nat. Struct. Mol. Biol.* **23**, 464–467 (2016).

6. Fernández, C., Hilty, C., Wider, G., Güntert, P. & Wüthrich, K. NMR Structure of the Integral Membrane Protein OmpX. *J. Mol. Biol.* **336**, 1211–1221 (2004).

7. Thonghin, N., Kargas, V., Clews, J. & Ford, R. C. Cryo-electron microscopy of membrane proteins. *Methods vol.* **147**, 176–186 (2018).

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

9. Fenn, J. B. Electrospray wings for molecular elephants (Nobel lecture). in *Angewandte Chemie - International Edition* vol. **42**, 3871–3894 (2003).

10. Hu, Q. et al. The Orbitrap: a new mass spectrometer. *J. Mass Spectrom.* **40**, 430–443 (2005).

11. Sobott, F., Hernández, H., McCammon, M. G., Tito, M. A. & Robinson, C. V. A tandem mass spectrometer for improved transmission and analysis of large macromolecular assemblies. *Anal. Chem.* **74**, 1402–1407 (2002).

12. Lippens, J. L. et al. Fourier Transform-Ion Cyclotron Resonance Mass Spectrometry as a Platform for Characterizing Multimeric Membrane Protein Complexes. *J. Am. Soc. Mass Spectrom.* **29**, 183–193 (2018).

13. Laganowsky, A., Reading, E., Hopper, J. T. S. & Robinson, C. V. Mass spectrometry of intact membrane protein complexes. *Nat. Protoc.* **8**, 639–51 (2013).

14. Reading, E. et al. The Role of the Detergent Micelle in Preserving the Structure of Membrane Proteins in the Gas Phase. *Angew. Chemie Int. Ed.* **54**, 4577–4581 (2015).

15. Barrera, N. P., Di Bartolo, N., Booth, P. J. & Robinson, C. V. Micelles protect membrane complexes from solution to vacuum. *Science.* **321**, 243–246 (2008).

16. Olsen, J. V. et al. Higher-energy C-trap dissociation for peptide modification analysis. Nat. Methods **4**, 709–712 (2007)

17. Lermyte, F., Tsybin, Y. O., O'Connor, P. B. & Loo, J. A. Top or Middle? Up or Down? Toward a Standard Lexicon for Protein Top-Down and Allied Mass Spectrometry Approaches. *J. Am. Soc. Mass Spectrom.* **30**, 1149–1157 (2019).

18. Ro, S. Y. et al. Native top-down mass spectrometry provides insights into the copper centers of membrane-bound methane monooxygenase. *Nat. Commun.* **10**, 1–12 (2019).

19. Li, H., Nguyen, H. H., Loo, R. R. O., Campuzano, I. D. G. & Loo, J. A. An integrated native mass spectrometry and top-down proteomics method that connects sequence to structure and function of macromolecular complexes. *Nat. Chem.* **10**, 139–148 (2018).

20. 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).

21. Konijnenberg, A. et al. Top-down mass spectrometry of intact membrane protein complexes reveals oligomeric state and sequence information in a single experiment. *Protein Sci.* **24**, 1292–1300 (2015).

22. Laganowsky, A. et al. Membrane proteins bind lipids selectively to modulate their structure and function. *Nature* **510**, 172–175 (2014).

23. Yen, H. Y. et al. Ligand binding to a G protein–coupled receptor captured in a mass spectrometer. *Sci. Adv.* **3**, (2017).

24. Konijnenberg, A. et al. Global structural changes of an ion channel during its gating are followed by ion mobility mass spectrometry. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 17170–17175 (2014).

25. Marcoux, J. et al. Mass spectrometry reveals synergistic effects of nucleotides, lipids, and drugs binding to a multidrug resistance efflux pump. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 9704–9709 (2013).

26. Kim, M. et al. Need-based activation of ammonium uptake in *Escherichia coli. Mol. Syst. Biol.* **8**, 616 (2012).

27. Patrick, J. W. et al. Allostery revealed within lipid binding events to membrane proteins. *Proc. Natl. Acad. Sci. U. S. A.* **115**, 2976–2981 (2018).

28. Schmitt, N. D., Berger, J. M., Conway, J. B. & Agar, J. N. Increasing Top-Down Mass Spectrometry Sequence Coverage by an Order of Magnitude through Optimized Internal Fragment Generation and Assignment. *Anal. Chem.* **93**, 6355–6362 (2021).

29. Zenaidee, M. A. et al. Internal Fragments Generated from Different Top-Down Mass Spectrometry Fragmentation Methods Extend Protein Sequence Coverage. *J. Am. Soc. Mass Spectrom.* **32**, 1752–1758 (2021).

30. Lantz, C. et al. ClipsMS: An Algorithm for Analyzing Internal Fragments Resulting from Top-Down Mass Spectrometry. *J. Proteome Res.* **20**, 1928–1935 (2021).

31. Yin, S. & Loo, J. A. Top-down mass spectrometry of supercharged native protein–ligand complexes. Int. *J. Mass Spectrom.* **300**, 118–122 (2011).

32. Savage, D. F., Egea, P. F., Robles-Colmenares, Y., III, J. D. O. & Stroud, R. M. Architecture and Selectivity in Aquaporins: 2.5 Å X-Ray Structure of Aquaporin Z. *PLoS Biol.* **1**, e72 (2003).

33. Cong, X., Liu, Y., Liu, W., Liang, X. & Laganowsky, A. Allosteric modulation of protein-protein interactions by individual lipid binding events. *Nat. Commun.* **8**, (2017).

34. Marty, M. T. et al. Bayesian deconvolution of mass and ion mobility spectra: From binary interactions to polydisperse ensembles. *Anal. Chem.* **87**, 4370–4376 (2015).

35. Gupta, K. et al. The role of interfacial lipids in stabilizing membrane protein oligomers. *Nature* **541**, 421–424 (2017).

36. Gupta, K. et al. Identifying key membrane protein lipid interactions using mass spectrometry. *Nat. Protoc.* **13**, 1106–1120 (2018).

37. Loo, R. R., Dales, N. & Andrews, P. C. The effect of detergents on proteins analyzed by electrospray ionization. *Methods Mol. Biol.* **61**, 141–160 (1996).

38. Madsen, J. A. & Brodbelt, J. S. Comparison of infrared multiphoton dissociation and collision-induced dissociation of supercharged peptides in ion traps. *J. Am. Soc. Mass Spectrom.* **20**, 349–358 (2011).
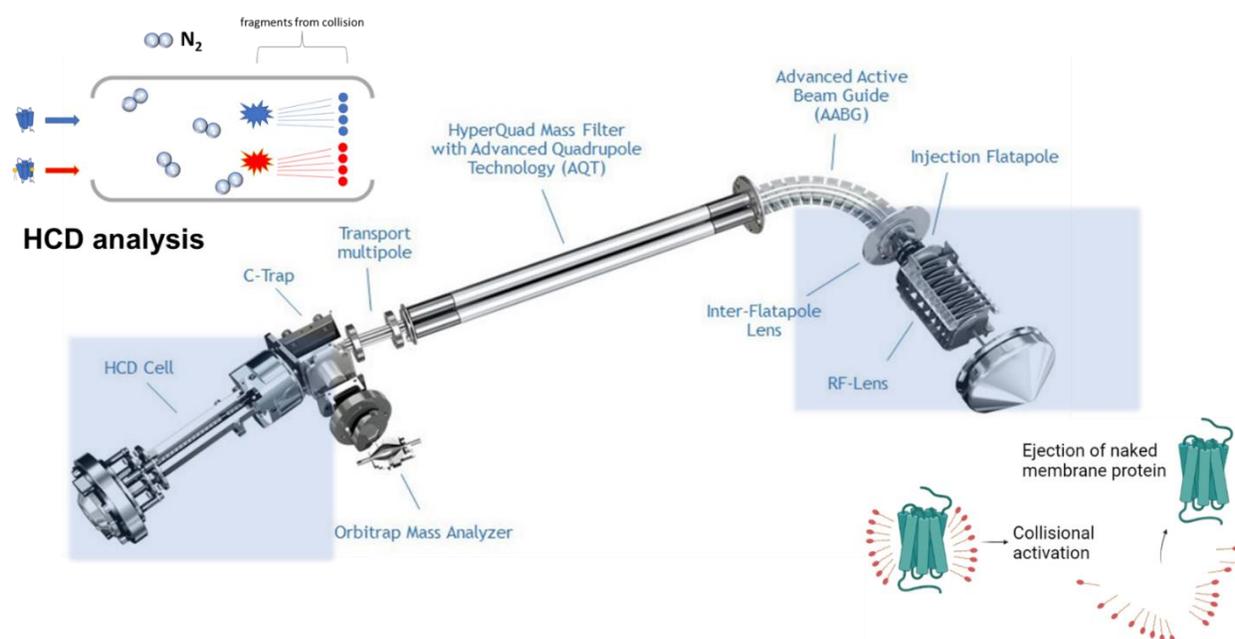
**Supplementary Figure 1. Native top-down analysis of membrane proteins by UHMR.** The UHMR from ThermoFisher is a mass spectrometer specialized for native and native top-down MS analysis of large biomolecules. For membrane protein analysis, the source region (highlighted in blue) is used to activate the membrane protein-micelle complex via collisions with a non-reactive gas such as nitrogen. The collisional activation experienced by the molecule can be modulated to eject the membrane protein out without causing significant structural perturbations. The quadrupole system, then, can transmit and select ions of interest by filtering by the analyte's mass-to-charge ratio. When the target analytes of interest have been transmitted to the HCD cell (highlighted in blue), collisional activation can be induced again to cause fragmentation. The resulting fragmentation spectrum is deconvoluted (conversion from mass-to-charge ratio to mass) to be interpreted for structural information.
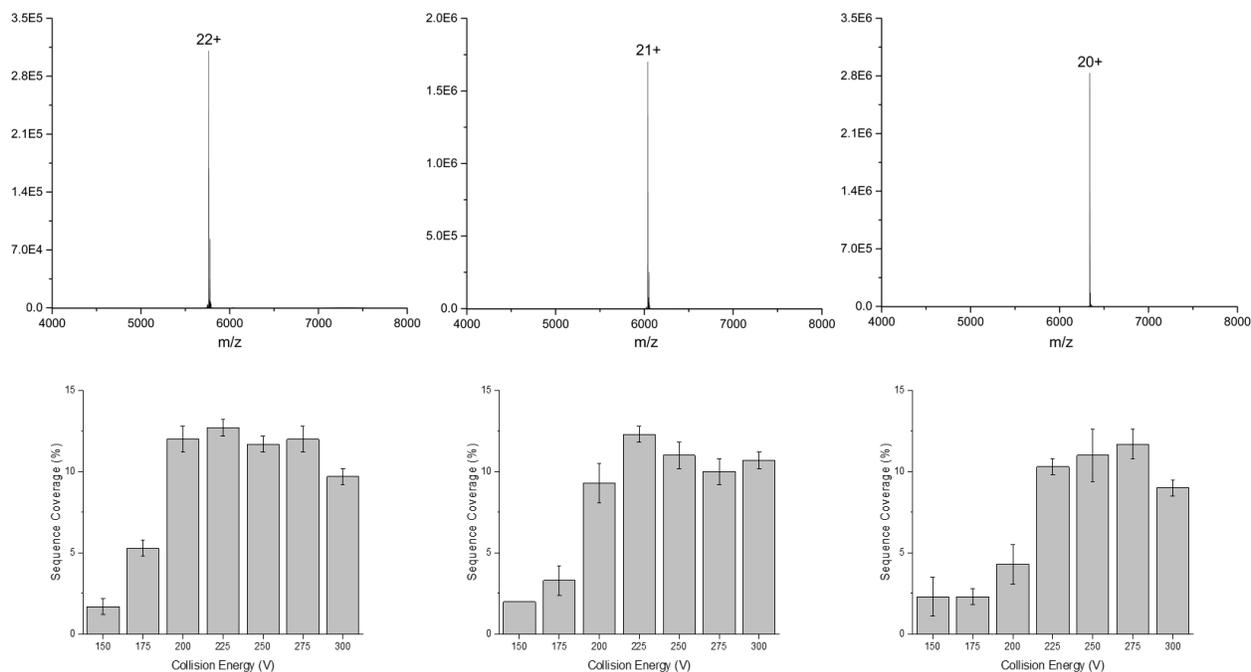
**Supplementary Figure 2. Correlation between the collision energy and sequence coverage obtained for AmtB.** To investigate the collisional energy levels to be used for HCD analysis of the AmtB and AmtB-PG complexes, apo-AmtB was subjected to a series of collisional energy levels from 150V to 300V. Three major charge states' sequence coverage was investigated after subjecting each charge state to 7 different levels of fragmentation energy in triplicates. The sequence coverage was less than 2.5% for the lowest charge state for 150V and hindered subsequent structural analysis. Activation level over 275V started to result in decline of sequence coverage for the 21+ and 22+ charge state. Thus, energy levels of 175V, 200V, 225V, and 250V were used for the native top-down MS analysis.

Total of 11 internal fragments bound with PG was verified through raw spectra screening.

Internal fragment 196-317
bound to PG
4+ charge state
0.3 ppm off from theoretical
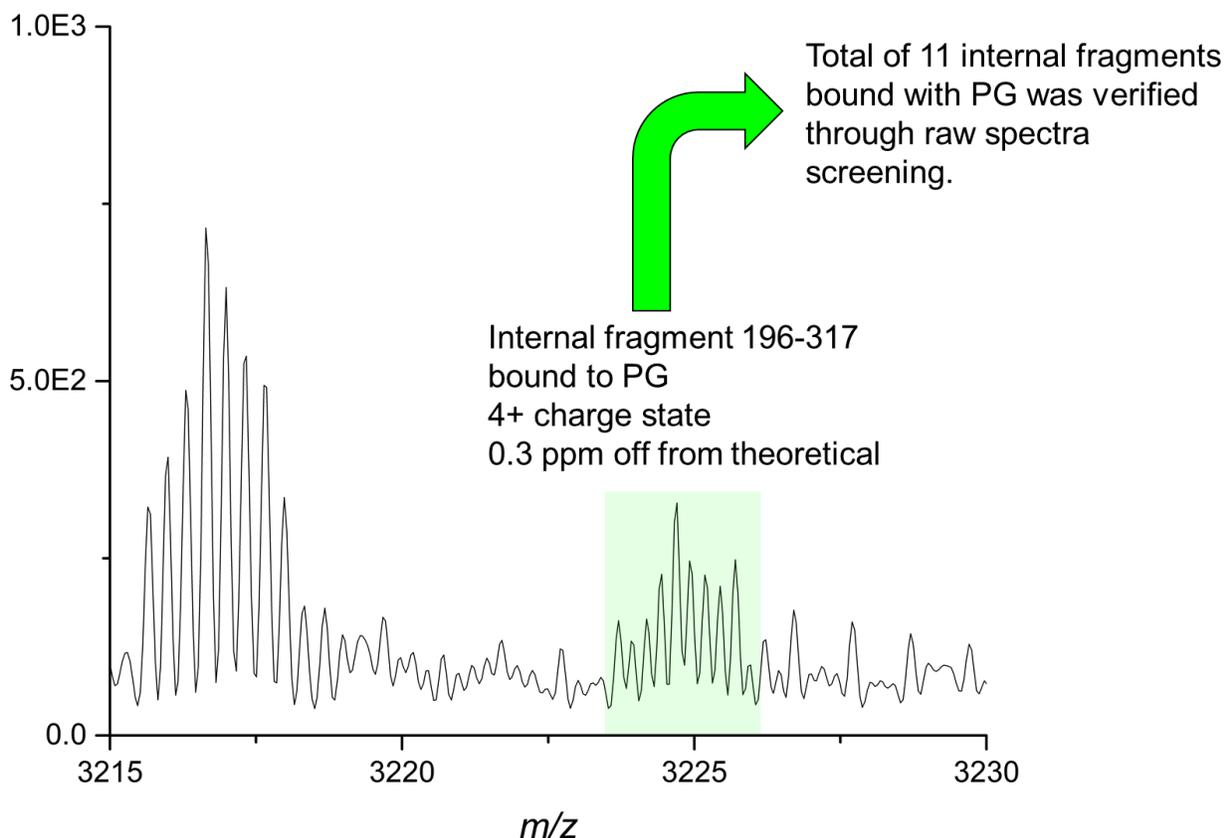
**Supplementary Figure 3. Example of manual verification process for internal fragments.**
Internal fragments that were detected by Clips-MS analysis were subjected to manual verification. The raw spectrum was screened to look for correct isotopic distribution as depicted in the figure. The example here is for the internal fragment residue 196-317 bound to 1 PG molecule. The isotopic distribution had the spacing of 0.25 m/z, indicating that the charge state is 4+. The observed mass was 0.3 ppm off from the theoretical mass value. A total of 11 internal fragments bound with PG were verified with this approach.
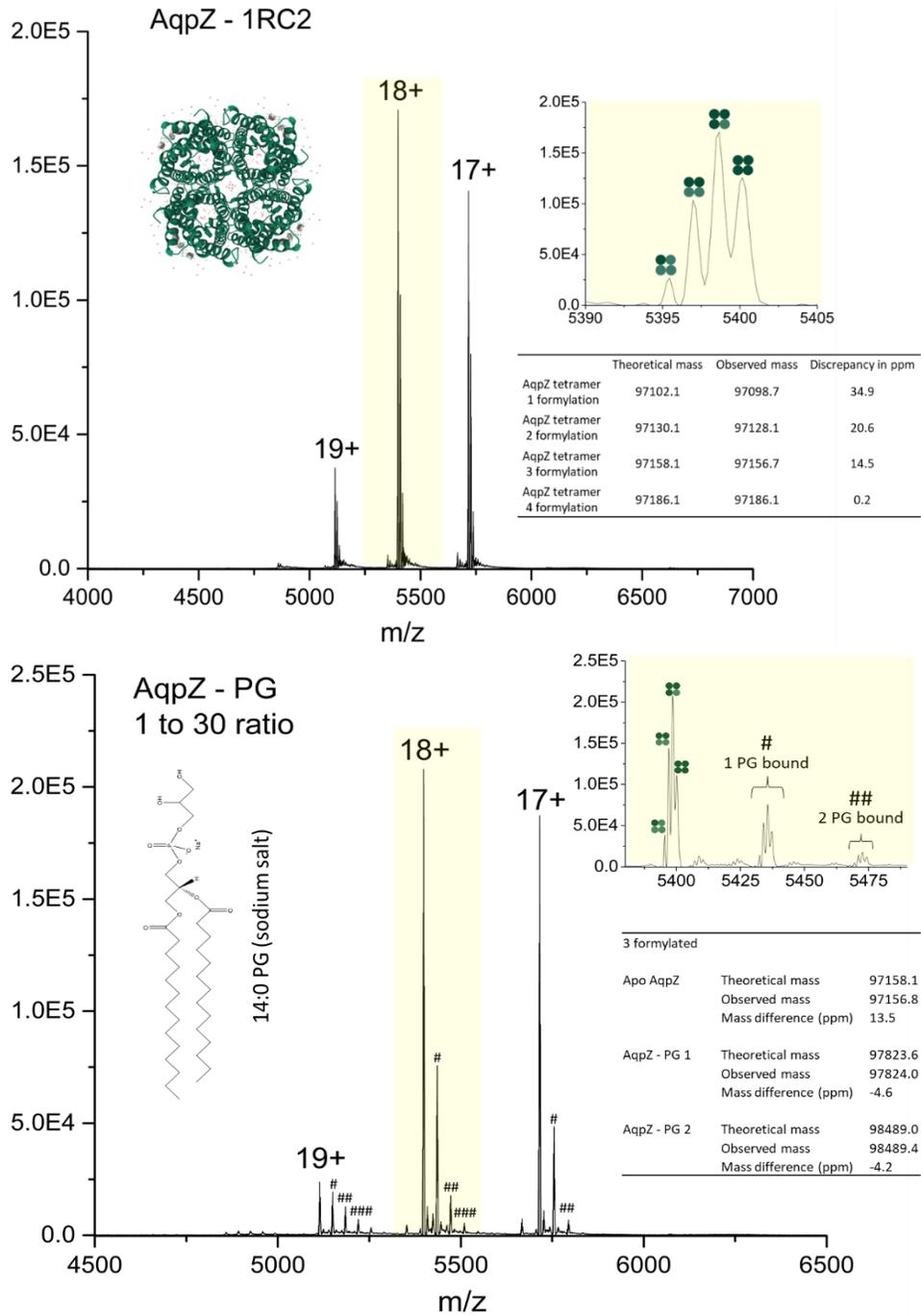
**Supplementary Figure 4. Native MS spectrum of AqpZ and AqpZ PG complex.** Although not discussed in the main text. Aquaporin Z, another channel protein for *E. coli*, and AqpZ-PG complex was subjected to native top-down MS analysis. The crystal structure of the protein and the 5 different proteoforms due to the N-terminal formylation of the monomer is shown. When PG is added, up to 3 lipids were found to bind readily. Native top-down MS data of AqpZ and AqpZ-PG complexes is consistent with the results found for AmtB-PG.

**Supplementary Figure 5. Native MS spectrum of AmtB-CL complex.** Although not discussed in the main text. AmtB-CL was also subjected to native top-down MS analysis. AmtB is also greatly stabilized by CL interactions and native top-down analysis was performed to understand the structural effects of the CL binding.

**AmtB sequence**

GASVADKADNAFMMICTALVLFMTIPGIALFYGGLIRGKNVLSMLTQVTVTFALVCILWVVYGYS
LAFGEGNNFFGNINWLMLKNIELTAVMGSIYQYIHVAFQGSFACITVGLIVGALAERIRFSAVLIFV
VVWLTLSYIPIAHMVWGGGLLASHGALDFAGGTVVHINAAIAGLVGAYLIGKRVGFGKEAFKPH
NLPMVFTGTAILYIGWFGFNAGSAGTANEIAALAFVNTVVATAAAILGWIFGEWALRGKPSLLGA
CSGAIAGLVGVTPACGYIGVGGALIVGVVAGLAGLWGVTMLKRLLRVDDPCDVFGVHGVCGIV
GCIMTGIFAASSLGGVGFAEGVTMGHQLLVQLESIAITIVWSGVVAFIGYKLADLTVGLRVPEEQ
EREGLDVNSHGENAYNA

**AqpZ sequence**

MFRKLAAECFGTFWLVFGGCGSAVLAAGFPELGIGFAGVALAFGLTVLTMAFAVGHISGGHFN
PAVTIGLWAGGRFPAKEVVGYVIAQVVGGIVAAALLYLIASGKTGFGAAASGFASNGYGEHSPG
GYSMLSALVVELVLSAGFLLVIHGATDKFAPAGFAPIAIGLALTLIHLISIPVTNTSVNPARSTAVAI
FQGGWALEQLWFFWVVPIVGGIIGGLIYRTLLEKRDGTLVPR

**Chapter 4**


**ICP-MS Analysis of Mercury in Fish: Exploration of Method Validation, Matrix Effect, and Kinetic Energy Discrimination**

Wonhyeuk Jung*†, Christopher S. Dunham*†, Katie A. Perrotta†, Yu Chen‡, James K. Gimzewski†##, Joseph A. Loo†#


†Department of Chemistry and Biochemistry, University of California, Los Angeles, California 90095, United States

‡Molecular Instrumentation Center, University of California, Los Angeles, California 90095, United States

#Department of Biological Chemistry, David Geffen School of Medicine, University of California, Los Angeles, California 90095, United States

##California NanoSystems Institute, University of California, Los Angeles, California, 90095, United States

*These authors contributed equally to this work.

The following is the manuscript submitted to ChemRxiv.

## Abstract

Inductively coupled plasma-mass spectrometry (ICP-MS) is a powerful analytical technique that can quantify elements of interest at parts-per-trillion concentrations. In this laboratory class, students performed ICP-MS analysis to quantify mercury concentration of standard reference material (SRM) 1947 (Lake Michigan fish tissue) and canned tuna from a local supermarket.

These two samples were digested in two different matrices ($HNO_3$/ $H_2O_2$ or $HNO_3$/HCl/$H_2O_2$) and then analyzed using no-gas mode or helium mode with two different kinetic energy discrimination voltages (2 V or 4 V). The inclusion of HCl in the matrix produced more accurate results and stabilized mercury over the 8-day period after the digestion. Based on their analysis, the students were asked to draw their own conclusions about what they perceived to be the most accurate representation of the true mercury concentration of the tuna samples. This laboratory class provides students with a wide range of scientific concepts to explore such as method verification with SRM, kinetic energy discrimination, matrix effect, and trace metal stability over time.

**Graphical Abstract**

## ICP-MS Analysis

**Mercury in my fish?**
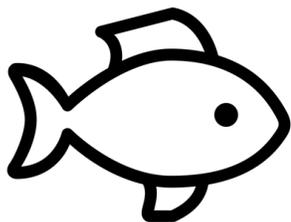
- ❑ **Different matrices?**
- ❑ **No gas mode or helium mode?**
- ❑ **KED parameter?**
- ❑ **Sample stability over time?**

**Keywords**

**Introduction**

Mercury (Hg) is found in the earth's crust at approximately 0.5 parts per million (ppm) as elemental mercury or as a sulfide.[1] Outgassing from rock, volcanic activity, coal burning, and mining can release mercury into the atmosphere. Additionally, mercury can be introduced to marine ecosystems as a result of industrial runoff or discharge, such as in the events that occurred in Minamata Bay, Japan, in the 1950s and in commercial gold mining operations.[2-4] As elemental or inorganic mercury enters the water cycle, it is converted into an organic compound, such as methylmercury, by microorganisms.[5] Once these microorganisms are consumed by smaller creatures and eventually by larger fish near or at the top trophic levels (e.g., tuna, swordfish, or shark, etc.), mercury concentration in the tissue can climb to significant levels.[1] Human consumption of these contaminated marine species can lead to accumulation of mercury in the body and result in adverse health effects, including neurological and behavioral disorders, neuromuscular effects, headaches, cognitive and motor dysfunction, kidney damage, and potentially death.[6]

The Food and Drug Administration (FDA) closely monitors Hg levels in commercial fish, provides guidelines on fish consumption, and prohibits the sale of fish that have methylmercury levels higher than an action level of 1 ppm.[7-10] Action levels represent the limit at which FDA will take legal action to remove products from the market. The FDA regularly updates the elemental analysis manual for food and related products and inductively coupled plasma-mass

spectrometry (ICP-MS) is one of the primary methods for the quantification of mercury in fish by the FDA.[11]

ICP-MS is a sensitive tool for elemental quantification with diverse applications such as environmental sample analysis, water quality control, and food analysis.[12–16] In the Journal of Chemical Education, the general principle of ICP-MS has been reviewed.[17] Additionally, previous publications in this journal investigated analysis of trace metals in sinus wash and apple juice, explored parameter optimization through simulation, and examined the effects of polyatomic interference.[18–22] Also, an undergraduate experiment analyzing mercury in fish samples using cold-vapor atomic absorbance has been published[23]; however, to date, mercury analysis of solid foodstuffs with ICP-MS in a class setting has not been published. Thus, in this educational paper, we pursued the following learning objectives while performing mercury ICP-MS analysis on fish samples:

- Method validation with standard reference material (SRM).
- Matrix effect: how different acid composition affects accuracy of the analysis and sample stability over time.
- Kinetic energy discrimination (KED) effect on accuracy of the analysis as KED voltage is varied.

In ICP-MS studies, it is common practice to use standard reference materials (SRMs) in order to validate the accuracy of the method.[24] The reference material has a known concentration of the element(s) you are designing the workflow for, in a matrix similar to the unknown sample. In this experiment, SRM 1947 (Lake Michigan Fish Tissue - certified by NIST)[25] was incorporated so that students could determine which combinations of parameters resulted in the most accurate results. Then, students were asked to translate this information into estimates about mercury concentrations in the canned tuna samples. The average level of mercury in canned tuna

samples, as determined by the FDA, was provided so that students can make conclusions about the accuracy of their unknown analysis.

Use of $HNO_3$ as the matrix, rather than HCl, is generally recommended for ICP-MS analysis to avoid isobaric interference caused by Cl-based polyatomic species.[26] However, Agilent reported detecting parts per trillion level of Hg with He or $O_2$ mode with 0.5 to 1.0% HCl used as the matrix.[26-27] Also, HCl has been found to promote stable storage of trace amounts of Hg, especially in combination with $HNO_3$ and $H_2O_2$.[28–30] Thus, with a correct implementation of the collision/reaction cell, the addition of HCl into the matrix for trace level Hg can be beneficial in terms of both ICP-MS analysis accuracy and long-term sample stability. To examine this aspect of the matrix effect, SRM 1947 and canned tuna from a local supermarket were digested with either $HNO_3/H_2O_2$ or $HNO_3/HCl/H_2O_2$ and then analyzed 1 day after and 8 days after the digestion. Stability over time is an important metric in an undergraduate experiment, since funds and time restrictions may only allow for one sample digestion that must last the duration of the class.

The main instrument of this study is Agilent 8800 ICP-MS, which utilizes the collision/reaction cell between two quadrupoles to attenuate isobaric interference.[31] The collision/reaction cell can be disabled (no-gas mode) or filled with either non-reactionary, e.g. He, or reactionary gases, e.g. $H_2$ or $NH_3$.[32]
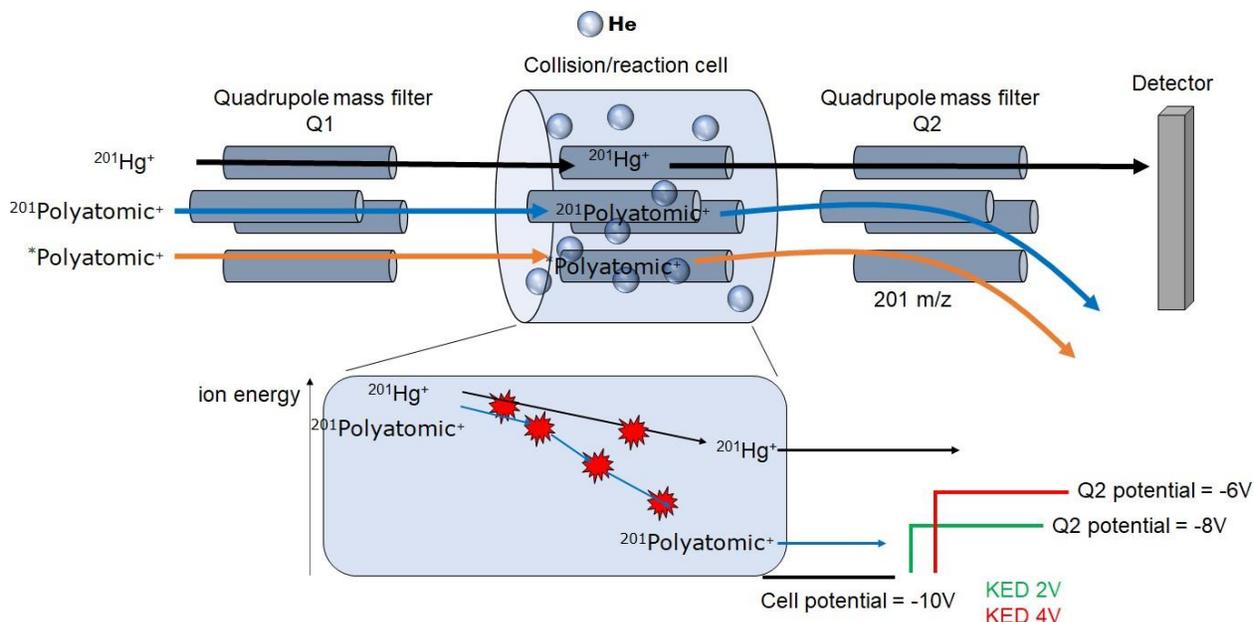
**Figure 1. Polyatomic interference (or isobaric interference) can be attenuated via kinetic energy discrimination.** Kinetic energy discrimination is defined as the use of a potential energy barrier between the collision/reaction cell and the mass analyzer to attenuate isobaric interference from polyatomic species. In a single-quadrupole mode, Quadrupole 1 (Q1) is set to let ions of all m/z pass through. The target element (201Hg+), polyatomic species with the same m/z (201Polyatomic+), and other ions (*Polyatomic+) are introduced into the collision/reaction cell. In no-gas mode (not shown), all ions travel unhindered to quadrupole 2 (Q2) and ions with the target m/z value (201 m/z) are selected by Q2. In helium mode, the collision/reaction cell is filled with helium and polyatomic species that can potentially cause isobaric interference undergo more collisions compared to 201Hg+. The resulting kinetic energy discrepancy can be taken advantage of to filter out 201Polyatomic+ by setting up a potential energy barrier, the magnitude of which can be manipulated by adjusting the potential difference between the collision/reaction cell and Q2.

When the collision/reaction cell is filled with non-reactive gas such as helium, the atomic ions (analyte ions) and the isobaric polyatomic ions (interfering ions) undergo multiple collisions with gas atoms. The former loses less kinetic energy than the latter during this process, resulting in energy difference between the two ion species of the same mass. By setting up a potential barrier of an appropriate magnitude, interfering polyatomic species can be filtered out while minimizing the loss of the target analyte, as shown in **Figure 1**. The efficiency of this suppression of isobaric interference, referred to as kinetic energy discrimination or KED, depends on a potential barrier that is neither too high nor too low.[34] Students were encouraged

105

to investigate the effect of this parameter by conducting the experiment at two different KED voltages, 2 V and 4 V.

This experiment incorporates many key concepts of analytical chemistry research including method verification, matrix effect examination, parameter optimization, and quantification using external calibration curves. This paper focuses on the analysis of Hg in commercial fish products in a classroom setting. However, the general methodology can be applied to any sample that can be digested with $HNO_3$ or HCl and with any element susceptible to polyatomic interference during ICP-MS analysis.

## Materials and Methods

**Reagents.** Hydrochloric acid 37%, ≥99.999% trace metals basis (Sigma-Aldrich, Cat. no. 339253), nitric acid 70%, purified by redistillation, ≥99.999% trace metals basis (Sigma-Aldrich, Cat. no. 225711), hydrogen peroxide 30%, for trace analysis (Sigma-Aldrich, Cat. no. 95321), and water, ultra-trace elemental analysis grade (Fisher, Cat. no. W9-1) were used to generate matrices for digestion and analysis. Scandium stock solution 100 µg/mL, 7% $HNO_3$ (Inorganic Ventures, part no. CGSC10-125ML) and Yttrium stock solution 100 µg/mL, 2% $HNO_3$ (Inorganic Ventures, part no. MSY-100PPM-125ML) was diluted to 8 ng/mL and 4 ng/mL, respectively, with either 2% $HNO_3$/0.34% $H_2O_2$ or 2% $HNO_3$/1% HCl/0.34% $H_2O_2$ matrix to generate internal standard (ISTD) solutions (Sc 8 ng/mL, Y 4 ng/mL). Mercury standard 10 µg/mL, 5% $HNO_3$ (Agilent, part no.

5190-8575) was diluted with either 2% $HNO_3$/0.34% $H_2O_2$ or 2% $HNO_3$/1% HCl/0.34% $H_2O_2$ matrix to generate 10 ng/mL stock solution, which was used to make calibration solutions (see Calibration Curve Solutions Preparation for details). ICP-MS tuning solution 10 µg/mL Ce, Co, Li, Mg, TI, Y (Agilent, part no. 5190-0465) was diluted 1,000-fold to 10 µg/L with 2% $HNO_3$ to

conduct plasma warm-up runs and auto-tuning (see Instrument Parameters & Auto Tuning section for details). SRM 1947 (Lake Michigan Fish Tissue, certification date 9/3/2020, expiration date 12/31/2026) was purchased from NIST.[25] Canned tuna (albacore tuna in water, Thailand) was purchased from a local supermarket. After the digestion, (see Microwave Digestion & Sample Preparation section for details) the samples were stored in acid-resistant Nalgene™ Narrow-Mouth Bottles Made of Teflon™ PFA (Fisher, Cat. no. DS1630-0001).

**Microwave Digestion & Sample Preparation.** All samples were digested via microwave digestion. Microwave digestion was performed at the ICP-MS facility within the UC Center for Environmental Implications of Nanotechnology at UCLA. Samples (SRM, tuna) were measured into clean Teflon vessels for acid digestion. The digestion was carried out using either a mixture of 7% $HNO_3$ and 3% $H_2O_2$ or 7% $HNO_3$, 3.7% HCl, and 3% $H_2O_2$ (see Table 1. for details) at 190 °C for 20 minutes in a microwave digestion system (Titan MPS, PerkinElmer). Once the samples were cooled to room temperature, it was subsequently diluted to the final volume of 50 mL by adding filtered deionized water. Samples were then stored in a -4 °C freezer overnight and then analyzed the next morning (Day 1). After the Day 1 analysis, the samples were stored in a 4 °C refrigerator for 7 days and then re- analyzed (Day 8).

**Table 1. Microwave digestion sample parameters.**

| Initial Mineralization In | Sample weight (g) | Final sample volume (mL) |
|---|---|---|
| 5 mL of 70% $HNO_3$, 2 mL of 30% $H_2O_2$ | N/A | 50 |
| 4 mL of 70% $HNO_3$, 1 mL of 37% HCl, 2 mL of 30% $H_2O_2$ | N/A | 50 |
| 5 mL of 70% $HNO_3$, 2 mL of 30% $H_2O_2$ | 1.183 | 50 |
| 4 mL of 70% $HNO_3$, 1 mL of 37% HCl, 2 mL of 30% $H_2O_2$ | 1.169 | 50 |
| 5 mL of 70% $HNO_3$, 2 mL of 30% $H_2O_2$ | 1.251 | 50 |
| 4 mL of 70% $HNO_3$, 1 mL of 37% HCl, 2 mL of 30% $H_2O_2$ | 1.346 | 50 |

Prior to ICP-MS analysis, 2 mL of each sample (method blank, tuna, SRM) was transferred to a rinsed (rinse protocol described below) 15 mL conical tube and subsequently diluted using 5 mL of ultra-trace elemental analysis grade water (see Reagents section for details). All samples

were prepared in triplicate, such that a total of 9 vials (3x method blank, 3x tuna, 3x SRM) would be analyzed per run.

Rinse Protocol: The exterior, interior, and caps of 15 mL conical tubes were rinsed thrice with deionized water. Then, the interior of each tube was rinsed twice with either 2% $HNO_3$ or 2% $HNO_3$/1% HCl, depending on the final matrix.

**Calibration Curve Solutions Preparation.** Calibration curve solutions were prepared by first generating a 10 ng/mL, 50 mL stock solution of mercury in an acid matrix of either 2% $HNO_3$/0.34% $H_2O_2$ or 2% $HNO_3$/1% HCl/0.34% $H_2O_2$. For the blank, a 15 mL conical tube was filled with 10 mL of ultra-trace metal grade water. For the 0 ng/mL solution, a 15 mL conical tube was filled with 10 mL of the acid matrix (2% $HNO_3$/0.34% $H_2O_2$ or 2% $HNO_3$/1% HCl/0.34% $H_2O_2$) only. Serial dilution was used to produce calibration solutions with mercury concentrations of 0.5, 1, 2, and 3 ng/mL. A batch of fresh calibration solutions was prepared for each run.

**ICP-MS Instrument & Batch Design.** This experiment used the Agilent 8800-QQQ ICP-MS instrument in the Molecular Instrumentation Center at UCLA. The instrument was operated using Agilent's MassHunter 4.1 software. Experiment batch design was reviewed with students prior to the start of the run. Method blank, SRM, and tuna samples were analyzed in triplicates. Rinse steps were inserted between the samples to prevent cross contamination. All reported concentrations were average values from five technical replicates.

**Instrument Parameters & Auto Tuning.** Key plasma parameters that were maintained for all experiments are shown here. KED voltage was set at either 2V or 4V for no gas or helium mode for each matrix. After the plasma warm-up, auto tuning was performed before each run to calibrate the quadrupole and ion lens voltages to reach optimal signal intensity for target elements in the tune mix solution.[31]

**Table 2. Instrument parameters.**

| | | | |
|---|---|---|---|
| *RF Power* | 1550 W | *Nebulizer Pump* | 0.10 revolutions per second |
| *RF Matching* | 1.80 V | *S/C Temp* | 2 °C |
| *Sampling Depth* | 8.0 mm | *Gas Switch* | Dilution Gas |
| *Carrier Gas* | 1.00 L/min | *Makeup/Dilution Gas* | 0.20 L/min |
| *Option Gas* | 0.0% | | |

**Hazards.** Students should always wear protective goggles, gloves, and lab coats during the experiment. 70% HNO3 and 37% HCl used for mercury extraction are extremely caustic and should be handled with care. The calibration solutions that contain HNO3, HCl, and mercury should be generated with care in a well-ventilated fume hood and should not be kept for more than a few days. Finally, it is recommended that the autosampler chamber be ventilated during and after the experiment so that any vapor build-up containing mercury is minimized.

## Results and Discussion

All data was generated by student volunteers who took Chem 184 (analytical chemistry class for upper-division undergraduate students) at UCLA during the Winter 2021 quarter. Calibration curves were excellent overall, possessing correlation coefficients between 0.9909 and 1.000 across 16 calibration curves. Two representative calibration curves are shown in **Figure 2**. Detection limits calculated by MassHunter software were all below 0.01 ng/mL, two orders of magnitude below the target mercury concentration range of both tuna and SRM samples. All

remaining calibration curves, as well as internal standard recovery curves, are provided in the

Supporting Information.
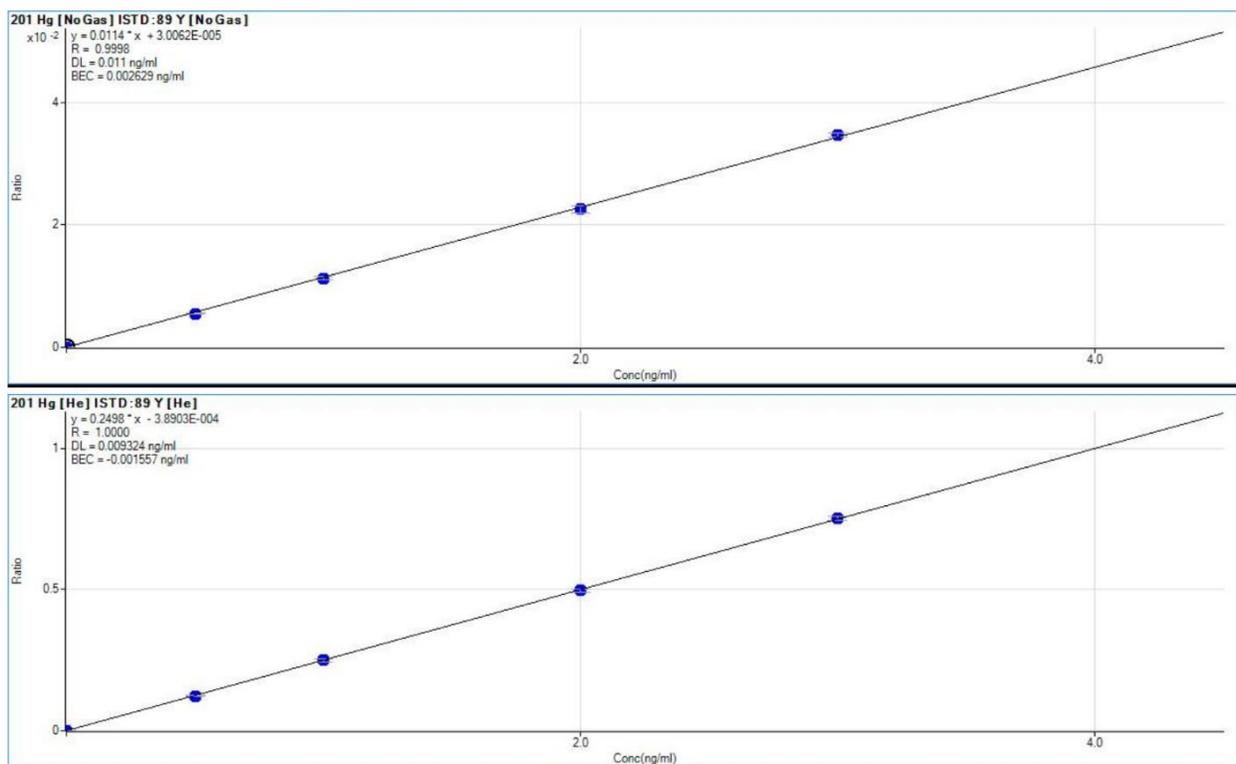


**Figure 2. Student-generated calibration curves.** Two representative calibration curves for 2% $HNO_3$/0.34% $H_2O_2$, 2V KED, no gas mode trial (top) and 2% $HNO_3$/1% HCl/0.34% $H_2O_2$, 4V KED, He mode trial (bottom) are shown here.

Mean tuna and SRM mercury concentrations calculated from triplicate measurements (noted as

[tuna] and [SRM] from here) on Day 1 and Day 8 are summarized in Figure 3. When samples

were analyzed with Matrix A (2% $HNO_3$/0.34% $H_2O_2$) on Day 1, trials with KED = 2V met the

NIST standard of ±10% of expected Hg concentration while trials with KED = 4V failed. Similar

trend was observed for both no gas mode and helium mode, in which heightened potential

energy barrier lead to loss of the target ions. This result indicates that the kinetic energy of the

target ions, determined by the instrument parameters set by the auto-tune run with Matrix A is

not strong enough to overcome the KED voltage at 4V and that there is minimal isobaric

polyatomic interference.
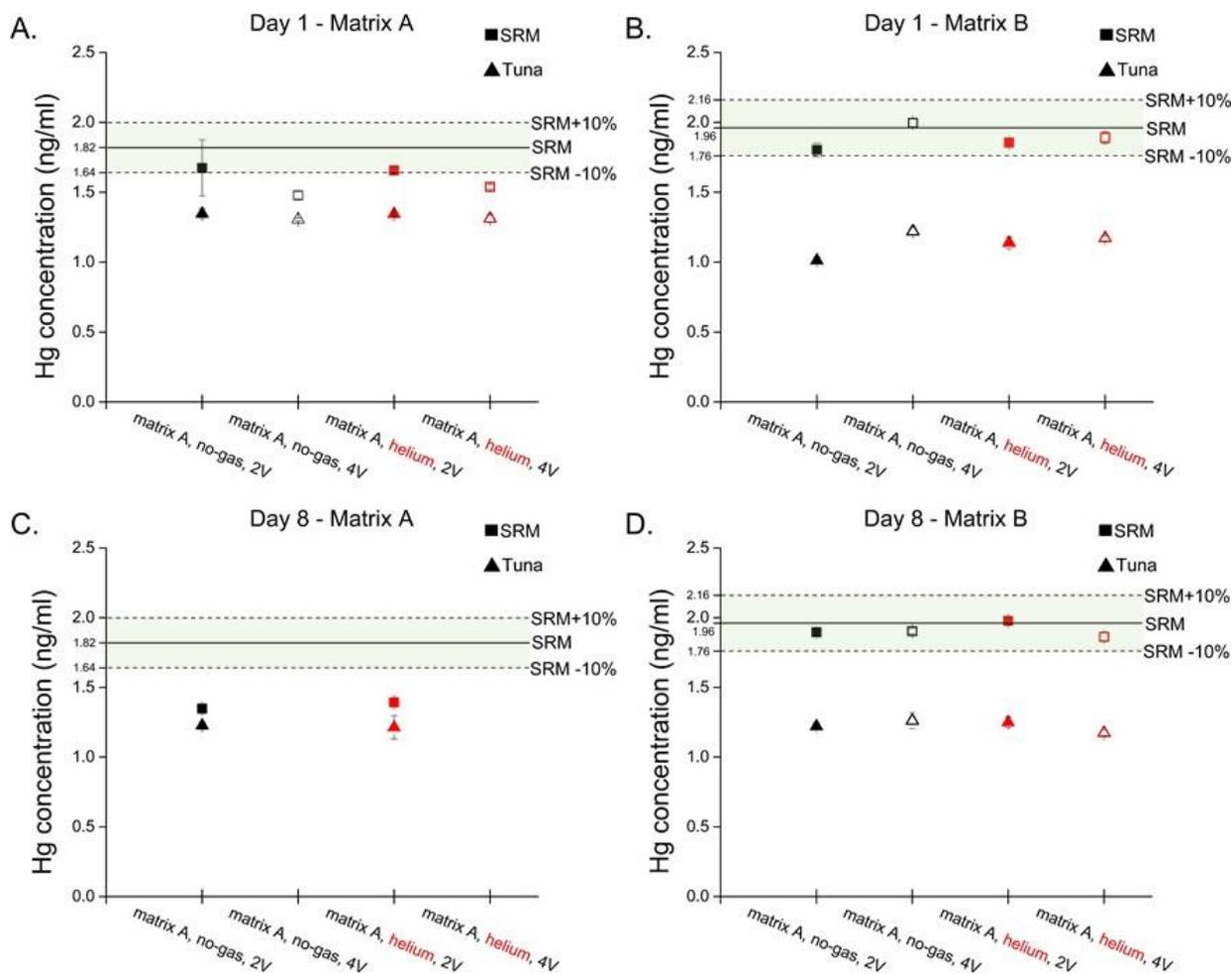


**Figure 3. Day 1 and Day 8 student analysis results.** A) Day 1 - Matrix A (2% $HNO_3$/0.34% $H_2O_2$) results. Mean tuna and SRM Hg concentrations were calculated from triplicate measurements. Expected SRM Hg concentration for Matrix A was 1.82 ng/mL and ±10% margin of error is highlighted in green. B) Day 1 - Matrix B (2% $HNO_3$/1% HCl/0.34% $H_2O_2$) results. Expected SRM Hg concentration for Matrix B was 1.96 ng/mL and ±10% margin of error is highlighted in green. C) Day 8 - Matrix A results. 4V KED trial for Matrix A was excluded from analysis since the condition failed to meet the NIST standard of ±10% margin of error on Day 1. D) Day 8 - Matrix B results.

Analysis of Day 8 samples followed identical methods; however, because the [SRM] in Matrix A

failed to meet validation criteria at KED = 4V on Day 1, this condition was excluded from Day 8

analysis. A decrease in [SRM] was observed for both no-gas mode (-19.4%) and He-mode (-15.9%) at KED = 2V. These results agree with previous studies demonstrating mercury loss of up to 50% after 10 days, depending on sample container.[28,29] This is especially relevant for undergraduate classes, since sample digestion availability may be limited due to cost or time constraints. Thus, unless the goal of the class is to investigate matrix effects on mercury stability over time, HNO3 or HNO3/ H2O2 matrix composition should be avoided for long-term sample storage.

In contrast, [SRM] for Matrix B (2% $HNO_3$/1% HCl/0.34% $H_2O_2$) met the ±10% margin of error for the expected value regardless of gas mode or KED voltage on Day 1. Curiously, higher potential energy barrier resulted in better accuracy for no gas mode, indicating that that interference attenuation by KED in no gas mode is possible. Similar result has been observed in a recent study but authors of the study states that there is no obvious explanation for this observation, and it may not be reproducible on all ICP-MS instruments, since it may be related to the plasma or interface system of the ICP-MS.[35] Thus, detailed discussion of this result was deemed outside the scope of this class and was not pursued further. When KED voltage was at 2V, accuracy improved slightly when helium mode used compared to the no gas mode, indicating that KED with helium mode leads to sufficient kinetic energy discrepancy between interfering polyatomic species and the target ions so that 2V potential barrier was effective. 4V potential barrier was not too high to harm target ion detection significantly but resulted in slightly less accurate result for Day 8.

For Matrix B, Day 8 results did not differ significantly from Day 1 results and all SRM analysis fell within ±5% margin of error from expected Hg concentration. This outcome agrees with previous studies that showed including HCl in the sample matrix enhances mercury stability over time.[28,29] These results demonstrate the importance of choosing the correct matrix for the

112

target element. Thus, for long-term storage of mercury samples for a class, $HNO_3/HCl$ or $HNO_3/HCl/H_2O_2$ matrix compositions are recommended.

Based on these results, students concluded that Matrix B is the superior matrix for trace Hg analysis in regard to both the accuracy and the sample stability over time. Students reported that performance of He mode was slightly better than no gas mode but may be dependent on KED voltage. Students also deduced that the KED voltage had a more pronounced effect on samples analyzed with Matrix A. This indicates that KED = 4V is too high for Matrix A and results in significant loss of sensitivity. Students concluded that both 2V and 4V KED voltages worked well with Matrix B. These results indicate that sensitivity loss from higher KED voltages had less impact when the matrix included HCl. From SRM analysis, students inferred that the tuna concentration from Matrix B, no gas, KED = 4V or Matrix B, helium, KED = 2V trials are the most accurate results. From these two trials, students calculated the expected tuna's original mercury concentrations as 0.183 ppm and 0.187 ppm, respectively. This was in close agreement with the FDA's value of 0.13 ppm.[7,33]


**Conclusion**

Students were asked to complete surveys before and after the experiment in order to assess learning objective progression. These survey results are provided in the Supplementary Information. In short, prior to the experiment, students expressed neutral confidence in ICP-MS experimental design and execution. Concepts concerning polyatomic interference, gas modes, and method validation, and the rationale underlying the choice of acid matrix, were areas of particularly low understanding among the volunteers. After the experiment, students reported high understanding in these four areas. Students also reported a high degree of confidence that they could design and carry out their own analysis using ICP-MS. Many undergraduate

institutions already have ICP-MS instrumentation for educational purposes. This protocol

provides a straightforward way to explore major aspects of ICP-MS analysis while piquing

students' interest by investigating a real-world problem.

**References**

1. Bernhoft, R. A. Mercury Toxicity and Treatment: A Review of the Literature. *J. Environ. Public Health.* **2021**, 460508 (2012).

2.Jernelöv, A., Landner, L. Larsson, T. Swedish Perspectives on Mercury Pollution. *J. Water Pollut. Control. Fed.* **47(4)**, 810–822 (1975).

3. Fujiki, M. & Tajima, S. The Pollution of Minamata Bay by Mercury. *Water Sci. Technol.* **25(11)**, 133–140 (1992).

4. Jiang, G., Shi, J., Feng, X. Mercury Pollution in China. *Environ. Sci. Technol.* **40(12)**, 3672-3678 (2006).

5. Mason, R. P. & Sheu, G.-R. Role of the ocean in the global mercury cycle. *Global Biogeochem. Cycles.* **16(4)**, 40-1 – 40-14 (2002).

6. Risher, J. F. & Amler, S. N. Mercury Exposure: Evaluation and Intervention: The Inappropriate Use of Chelating Agents in the Diagnosis and Treatment of Putative Mercury Poisoning. *Neurotoxicology* **26**, 691–699 (2005).

7. FDA, Mercury Concentrations in Fish from the FDA Monitoring Program (1990-2010) https://www.fda.gov/media/101237/download (2010)

8. FDA, Mercury Levels in Commercial Fish and Shellfish. https://www.fda.gov/food/metals-and-your-food/mercury-levels-commercial-fish-and-shellfish-1990-2012 (2012)

9. FDA. Advice about Eating Fish. https://www.fda.gov/food/consumers/advice-about-eating-fish (2021)

10. ATSDR. Public Health Statement for Mercury. https://wwwn.cdc.gov/TSP/PHS/PHS.aspx?phsid=112&toxid=24 (2015)

11. FDA. Guidance for Industry: Action Levels for Poisonous or Deleterious Substances in Human Food and Animal Feed. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/guidance-industry-action-levels-poisonous-or-deleterious-substances-human-food-and-animal-feed (2000)

12. FDA. Inductively Coupled Plasma-Mass Spectrometric Determination of Arsenic, Cadmium, Chromium, Lead, Mercury, and Other Elements in Food Using Microwave Assisted Digestion; FDA Elemental Analysis Manual. http://www.fda.gov/EAM (2020)

13. Pröfrock, D., Leonhard, P., Wilbur, S. & Prange, A. Sensitive, simultaneous determination of P, S, Cl, Br and I containing pesticides in environmental samples by GC hyphenated with collision-cell ICP-MS. *J. Anal. At. Spectrom.* **19**, 623–631 (2004).

14. Allibone, J., Fatemian, E. & Walker, P. J. Determination of mercury in potable water by ICP-MS using gold as a stabilising agent. *J. Anal. At. Spectrom.* **14**, 235–239 (1999).

15. Nardi, E. P. et al. The use of inductively coupled plasma mass spectrometry (ICP-MS) for the determination of toxic and essential elements in different types of food samples. *Food Chem.* **112**, 727–732 (2009).

16. Maione, C. et al. Comparative study of data mining techniques for the authentication of organic grape juice based on ICP-MS analysis. *Expert Syst. Appl.* **49**, 60–73 (2016).

17. Houk, R.S. Inductively coupled plasma-mass spectrometry and the European discovery of America. *J. Chem. Educ.* **77(5)**, 598-607 (2000).

18. Sanders, J. K. Inductively Coupled Plasma-Mass Spectrometry: Practices and Techniques (Taylor, Howard E.). *J. Chem. Educ.* **78**, 1465 (2001).

19. Donnell, A. M., Nahan, K., Holloway, D. & Vonderheide, A. P. Determination of Arsenic in Sinus Wash and Tap Water by Inductively Coupled Plasma–Mass Spectrometry. *J. Chem. Educ.* **93**, 738–741 (2016).

20. He, P., Colón, L. A. & Aga, D. S. Determination of Total Arsenic and Speciation in Apple Juice by Liquid Chromatography–Inductively Coupled Plasma Mass Spectrometry: An Experiment for the Analytical Chemistry Laboratory. *J. Chem. Educ.* **93**, 1939–1944 (2016).

21. Schwarz, G. et al. Protein Quantification by Elemental Mass Spectrometry: An Experiment for Graduate Students. *J. Chem. Educ.* **91**, 2167–2170 (2014).

22. Managh, A.J., Reid, P., & Knox, M.A. Development and Use of "ICP-MS TuneSim": A Software App that Allows Students to Simulate Tuning an Inductively Coupled Plasma Mass Spectrometer. *J. Chem. Ed.* **95**, 2059-2063 (2018).

23. Kristian, K.E. et al. A simplified digestion protocol for the analysis of Hg in fish by cold vapor atomic absorption spectroscopy. *J. Chem. Ed.* **92**, 698-702 (2015).

24. Bulska, E. et al. On the use of certified reference materials for assuring the quality of results for the determination of mercury in environmental samples. *Environ. Sci. Pollut. Res. Int.* **24**, 7889-7897 (2017).

25. NIST. SRM 1947 - Lake Michigan Fish Tissue. https://www-s.nist.gov/srmors/view_detail.cfm?srm=1947 (2020)

26. Agilent Technologies, Inc. Successful Low Level Mercury Analysis using the Agilent 7700 Series ICP-MS. 5990-7173EN (2011)

27. Wang, X.-M., Wang, K., Zeng, X.-C., Hsu, D. & Song, J.-E. Accurate Analysis of Trace Mercury in Cosmetics using the Agilent 8900 ICP-QQQ. 5994-0461EN (2019)

28. Krivan, V. & Haas, H. F. Prevention of loss of mercury(II) during storage of dilute solutions in various containers. *Fresenius' Zeitschrift für Anal. Chemie* **332**, 1–6 (1988).

29. Louie, H., Wong, C., Huang, Y. J. & Fredrickson, S. A study of techniques for the preservation of mercury and other trace elements in water for analysis by inductively coupled plasma mass spectrometry (ICP-MS). *Anal. Methods* **4**, 522-529 (2011)

30. Zhu, Y., Ahmady, B., Numata, M. & Chiba, K. Solid phase extraction using a sulfoxide adsorbent for preconcentration and separation of Hg(II) in natural water followed by ICP- MS measurements. *Anal. Sci.* **28**, 417–422 (2012).

31. Agilent. Handbook of ICP-QQQ Applications using the Agilent 8800 and 8900 4th edition. https://www.agilent.com/cs/library/applications/appcompendium_icp-qqq-5991-2802en-us-agilent.pdf (2020)

32. Iglesias, M., Gilon, N., Poussel, E., Mermet, J. Evaluation of an ICP-collision/reaction cell-MS system for the sensitive determination of spectrally interfered and non-interfered elements using the same gas conditions. *J. Anal. At. Spectrom.* **17**, 1240–1247 (2002).

33. EPA. EPA-FDA Fish Advice: Technical Information. https://www.epa.gov/fish-tech/epa-fda-fish-advice-technical-information (2012)

34. Yamada, N. Kinetic energy discrimination in collision/reaction cell ICP-MS: Theoretical review of principles and limitations. *Spectrochim. Acta Part B At. Spectrosc.* **110**, 31–44 (2015).

35. Sugiyama, N. Attenuation of doubly charged ion interferences on arsenic and selenium by ICP-MS under low kinetic energy collision cell conditions with hydrogen cell gas. *J. Anal. At. Spectrom.* **36**, 294–302 (2021).

**Supplementary Information**
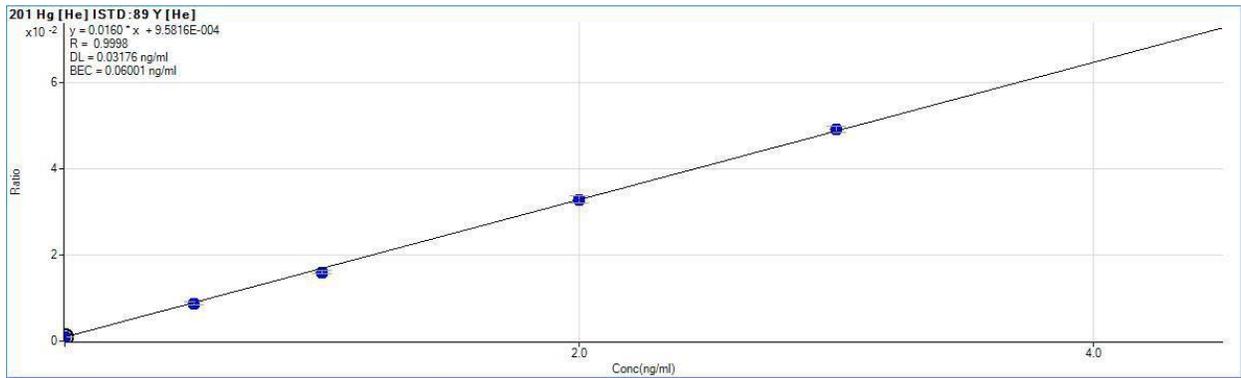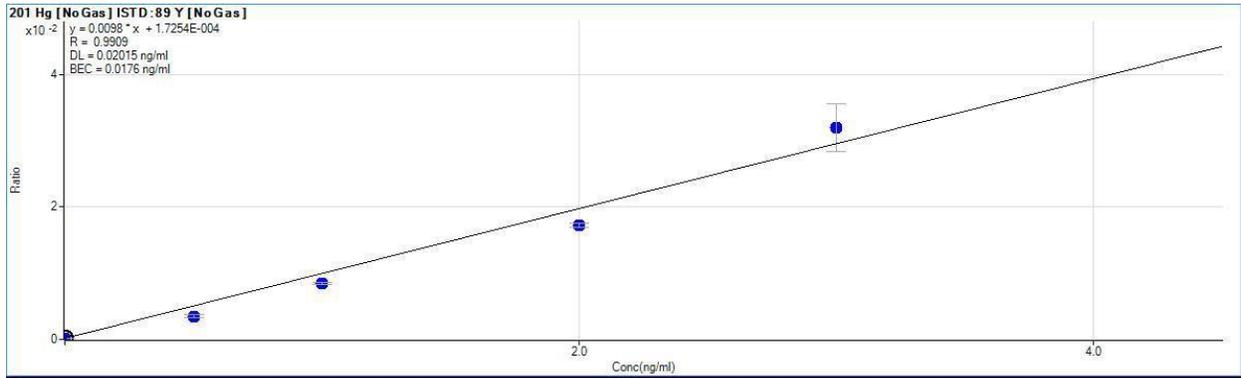




**Figure S1. Student-generated calibration curves for day 1, 2% HNO₃/1% HCl/0.34% H₂O₂, KED 2V.** Two calibration curves, no gas mode trial (top), He mode trial (bottom).
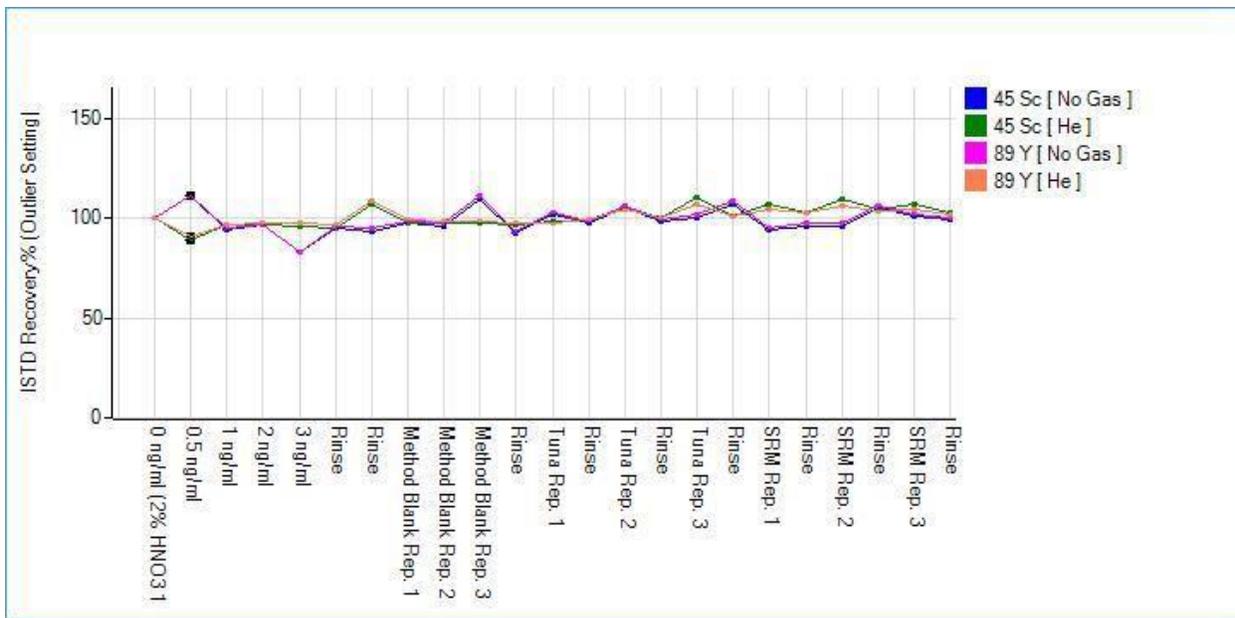
**Figure S2. Internal Standard recovery plot for day 1, 2% HNO₃/1% HCl/0.34% H₂O₂, KED 2V.** Internal standard percent recovery during no gas mode and helium mode for the above data set.
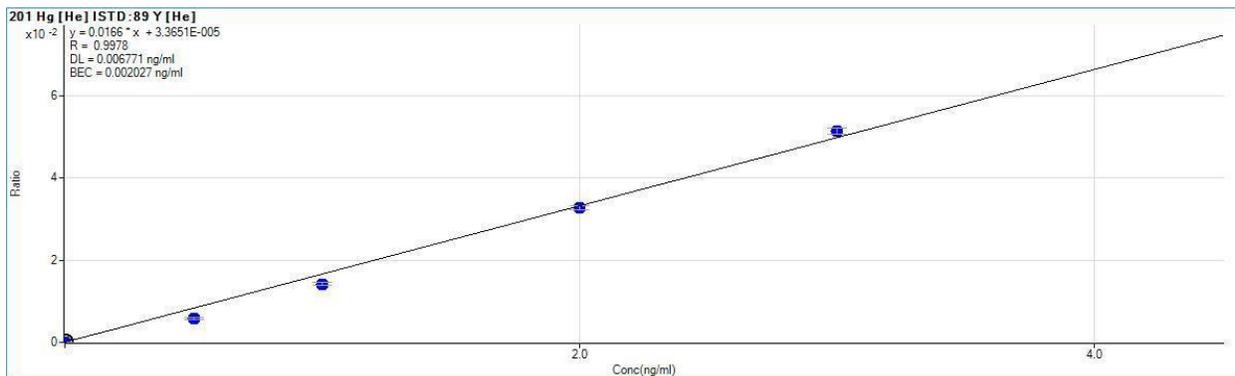


**Figure S3. Student-generated calibration curve for day 1, 2% HNO₃/0.34% H₂O₂, KED 2V. A calibration curve for the He mode trial.** The corresponding calibration curve for the no gas mode can be found in **Figure 3** within the main text.
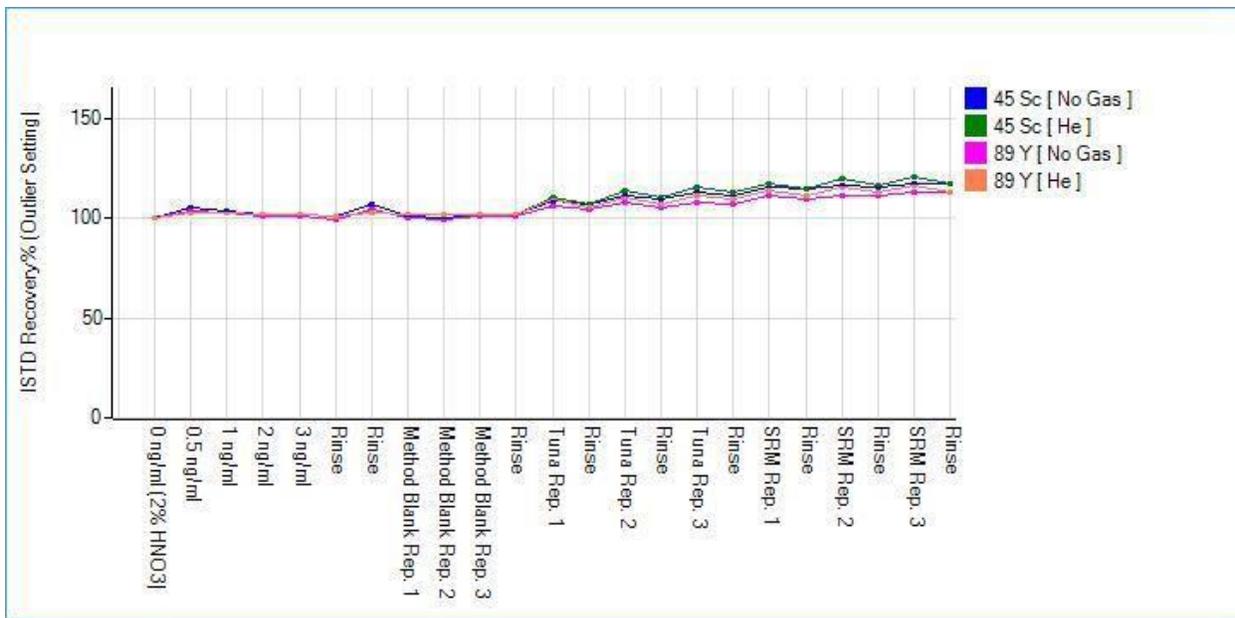
**Figure S4. Internal Standard recovery plot for day 1, 2% HNO$_3$/0.34% H$_2$O$_2$, KED 2V.**
Internal standard percent recovery during no gas mode and helium mode for the above data set.
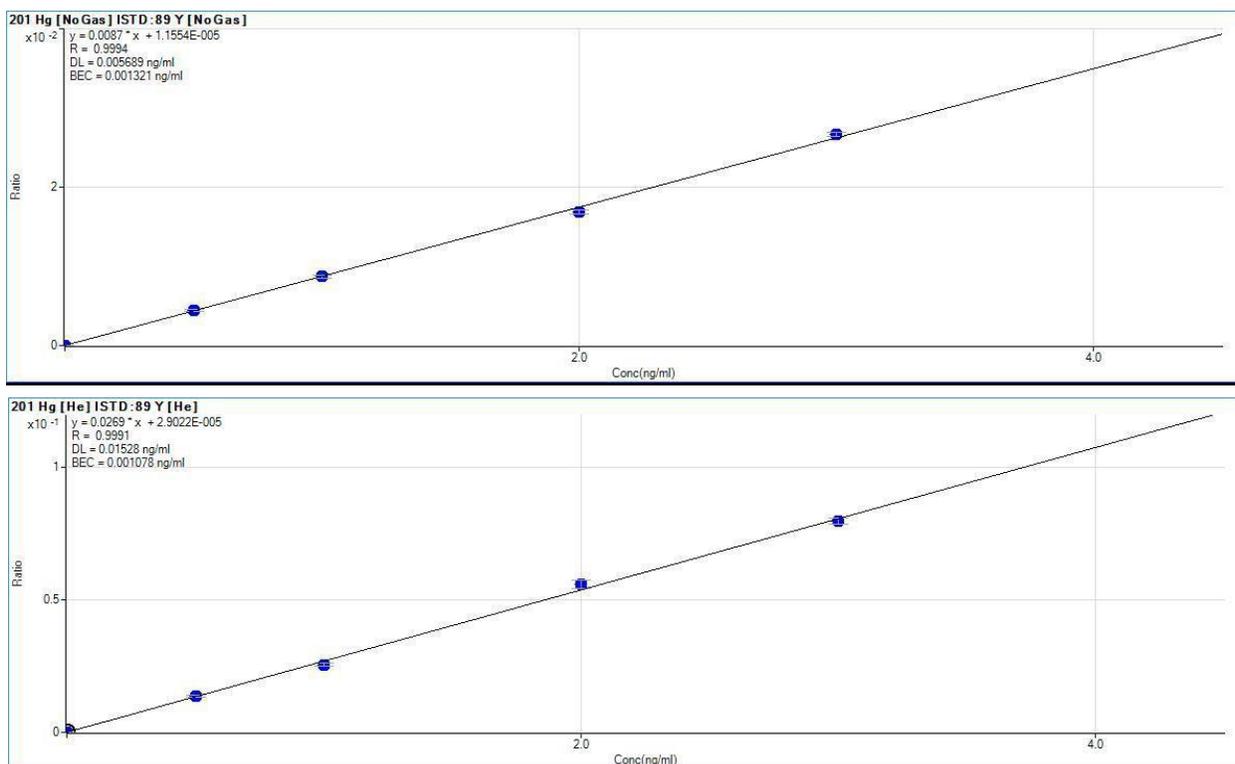


**Figure S5. Student-generated calibration curves for day 1, 2% HNO$_3$/1% HCl/0.34% H$_2$O$_2$, KED 4V.** Two calibration curves, no gas mode trial (top), He mode trial (bottom).
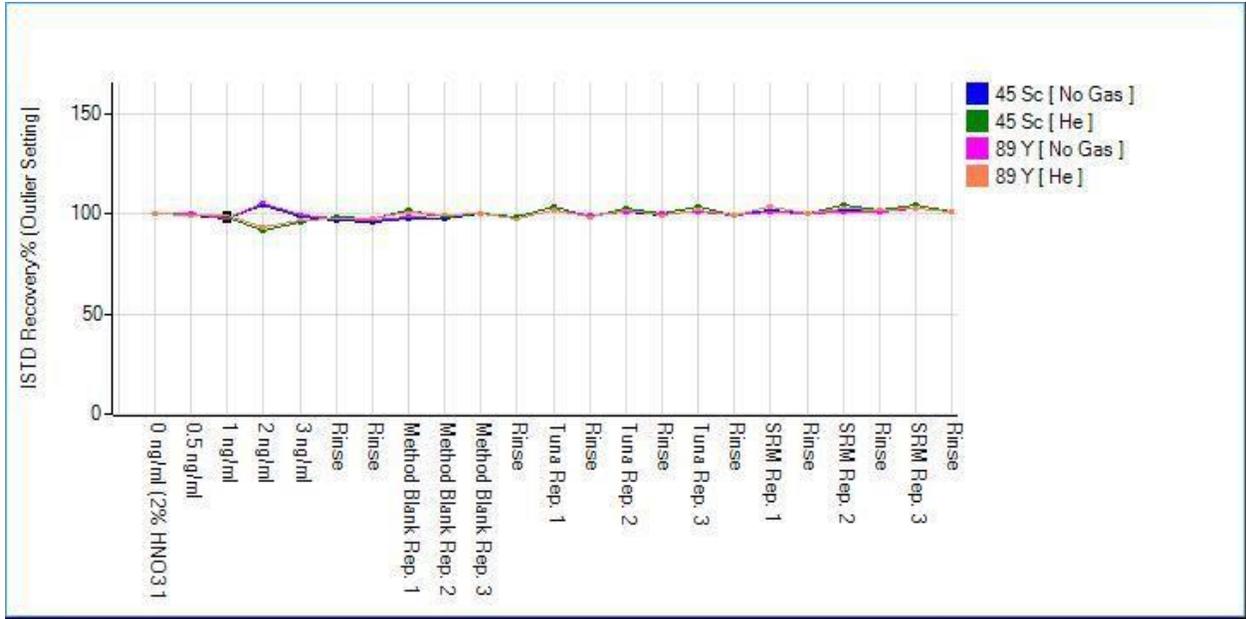
**Figure S6. Internal Standard recovery plot for day 1, 2% HNO₃/1% HCl/0.34% H₂O₂, KED 4V.** Internal standard percent recovery during no gas mode and helium mode for the above data set.
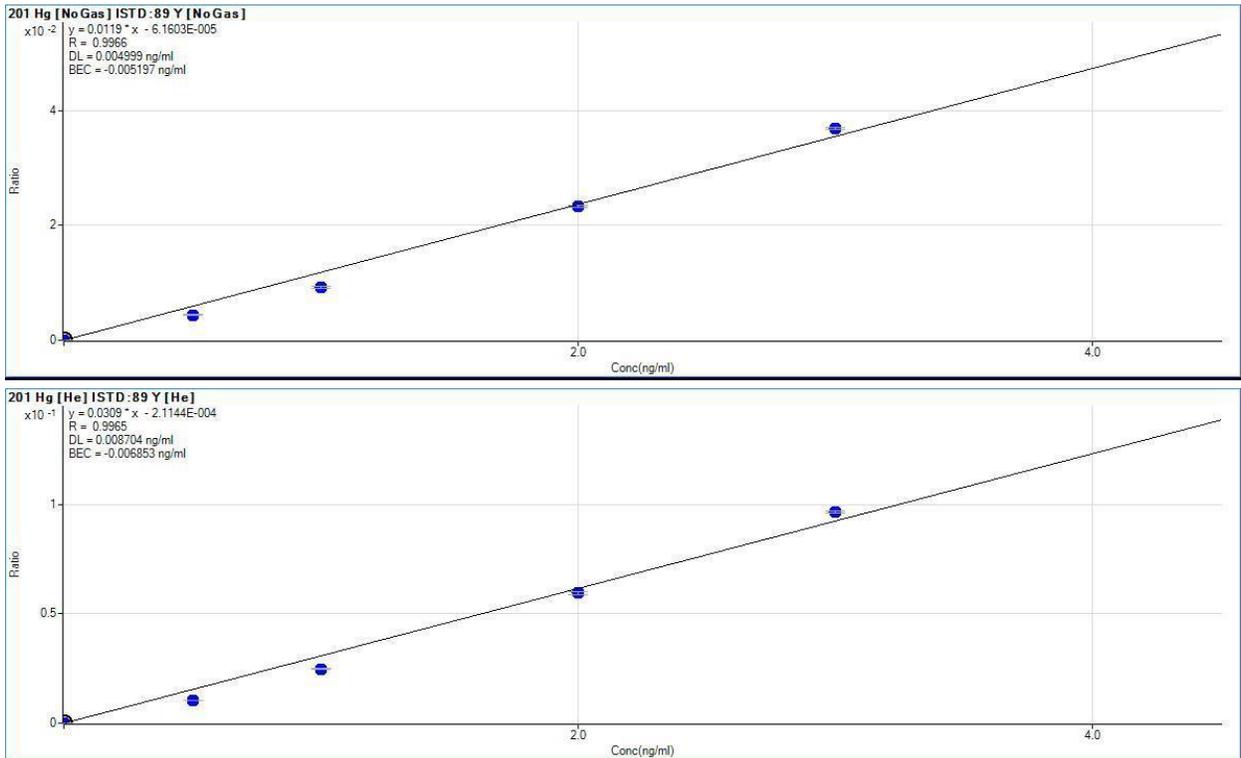


**Figure S7. Student-generated calibration curves for day 1, 2% HNO₃/0.34% H₂O₂, KED 4V.** Two calibration curves, no gas mode trial (top), He mode trial (bottom).
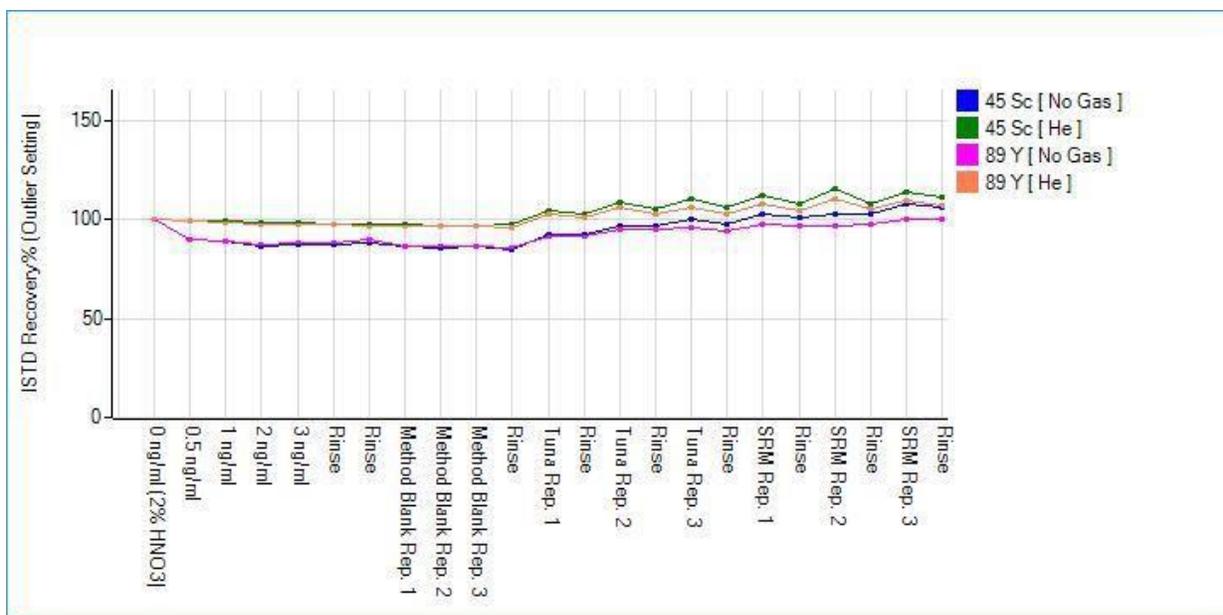
**Figure S8. Internal Standard recovery plot for day 1, 2% HNO₃/0.34% H₂O₂, KED 4V.**
Internal standard percent recovery during no gas mode and helium mode for the above data set.
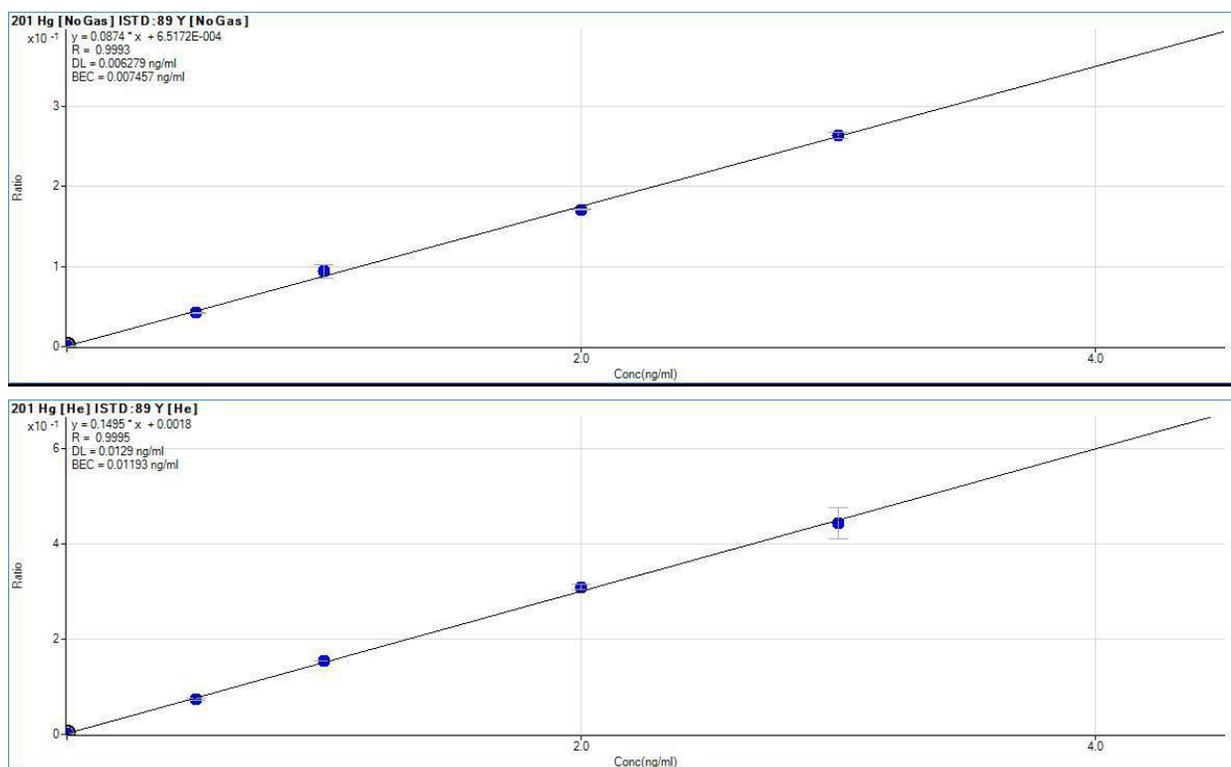


**Figure S9. Student-generated calibration curves for day 8, 2% HNO₃/1% HCl/0.34% H₂O₂, KED 2V.** Two calibration curves, no gas mode trial (top), He mode trial (bottom).

122
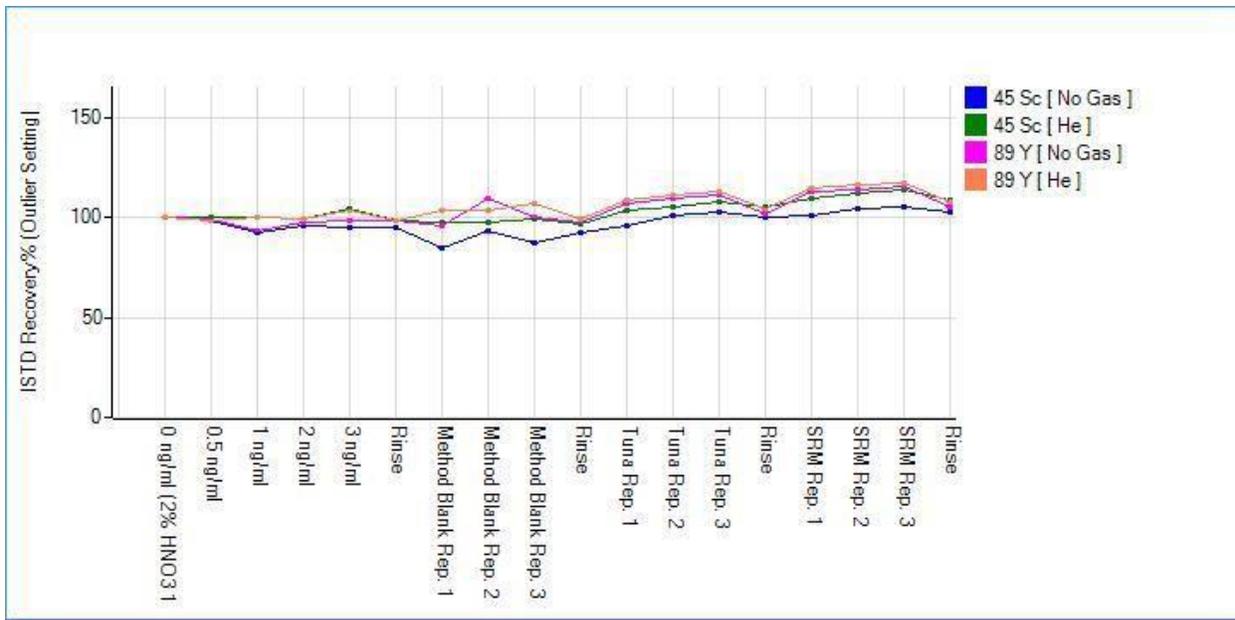
**Figure S10. Internal Standard recovery plot for day 8, 2% HNO₃/1% HCl/0.34% H₂O₂, KED 2V.** Internal standard percent recovery during no gas mode and helium mode for the above data set.
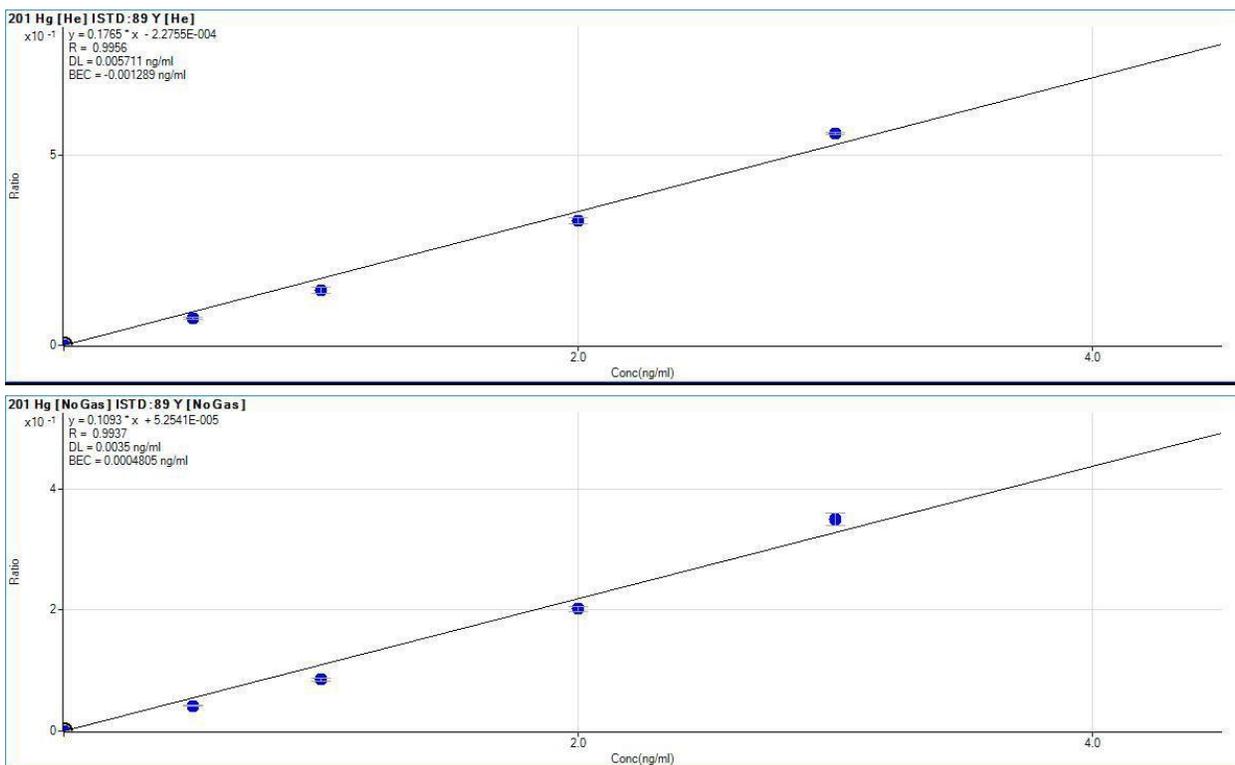


**Figure S11. Student-generated calibration curves for day 8, 2% HNO₃/0.34% H₂O₂, KED 2V.** Two calibration curves, no gas mode trial (top), He mode trial (bottom).
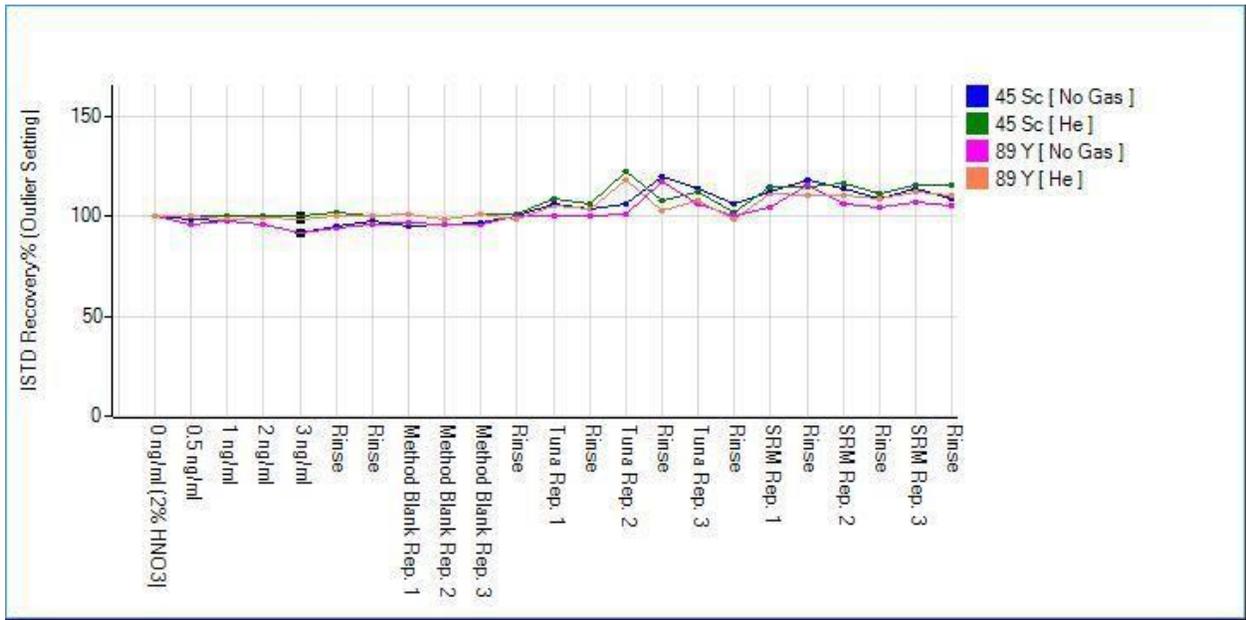
**Figure S12. Internal Standard recovery plot for day 8, 2% HNO$_3$/0.34% H$_2$O$_2$, KED 2V.**
Internal standard percent recovery during no gas mode and helium mode for the above data set.
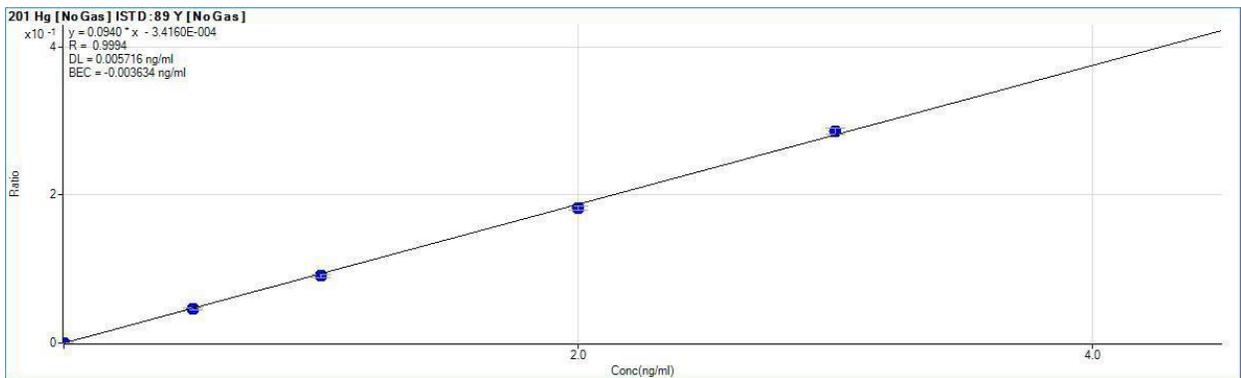


**Figure S13. Student-generated calibration curve for day 8, 2% HNO$_3$/1% HCl/0.34% H$_2$O$_2$, KED 4V.** Calibration curve for no gas mode trial. The corresponding calibration curve for the He mode can be found in **Figure 3** within the main text.
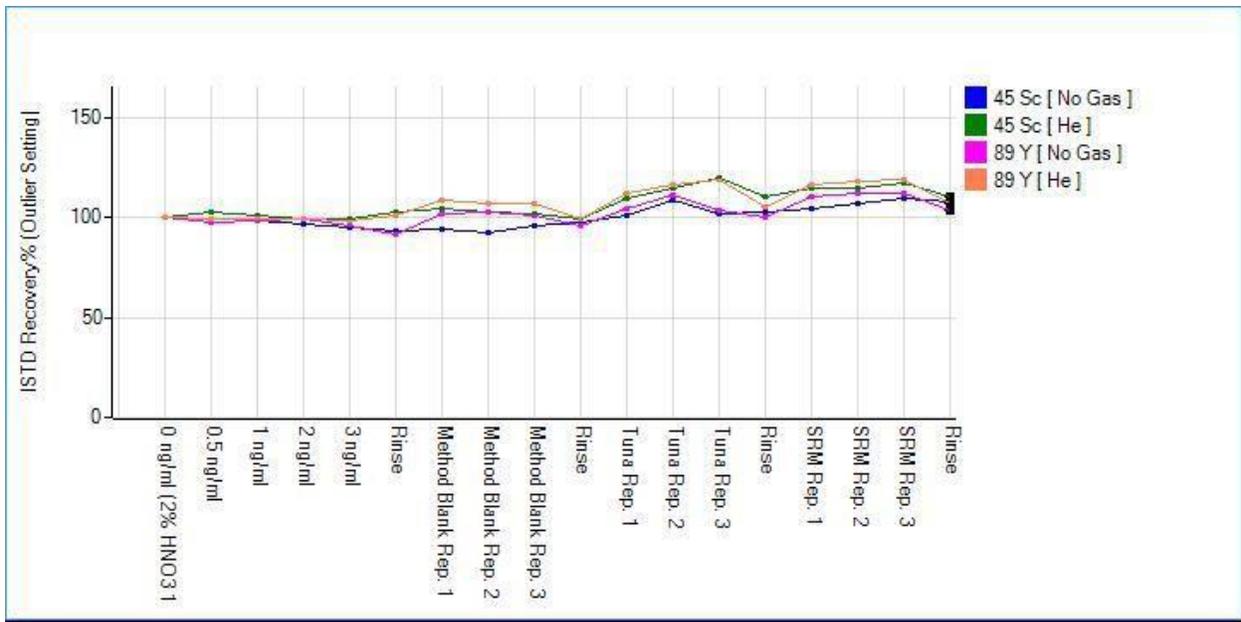
**Figure S14. Internal Standard recovery plot for day 8, 2% HNO₃/1% HCl/0.34% H₂O₂, KED 4V.** Internal standard percent recovery during no gas mode and helium mode for the above data set.
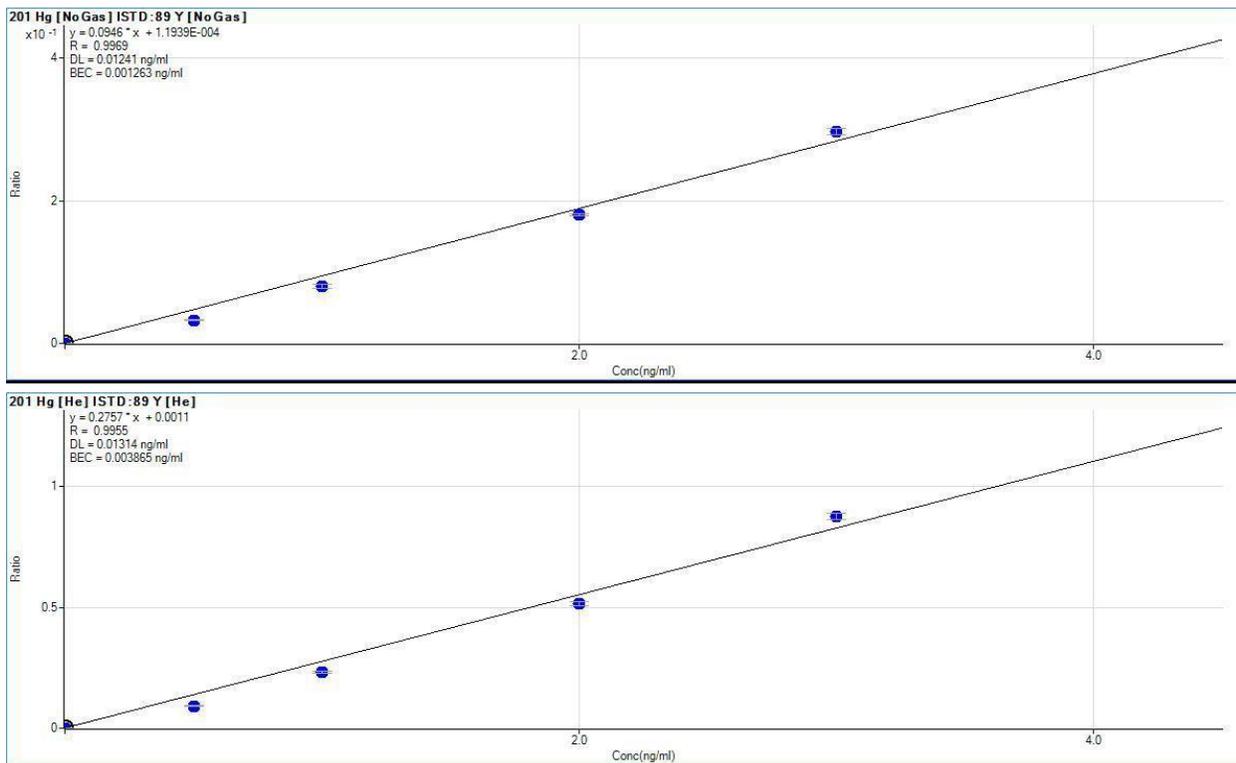


**Figure S15. Student-generated calibration curves for day 8, 2% HNO₃/0.34% H₂O₂, KED 4V.** Two calibration curves, no gas mode trial (top), He mode trial (bottom).
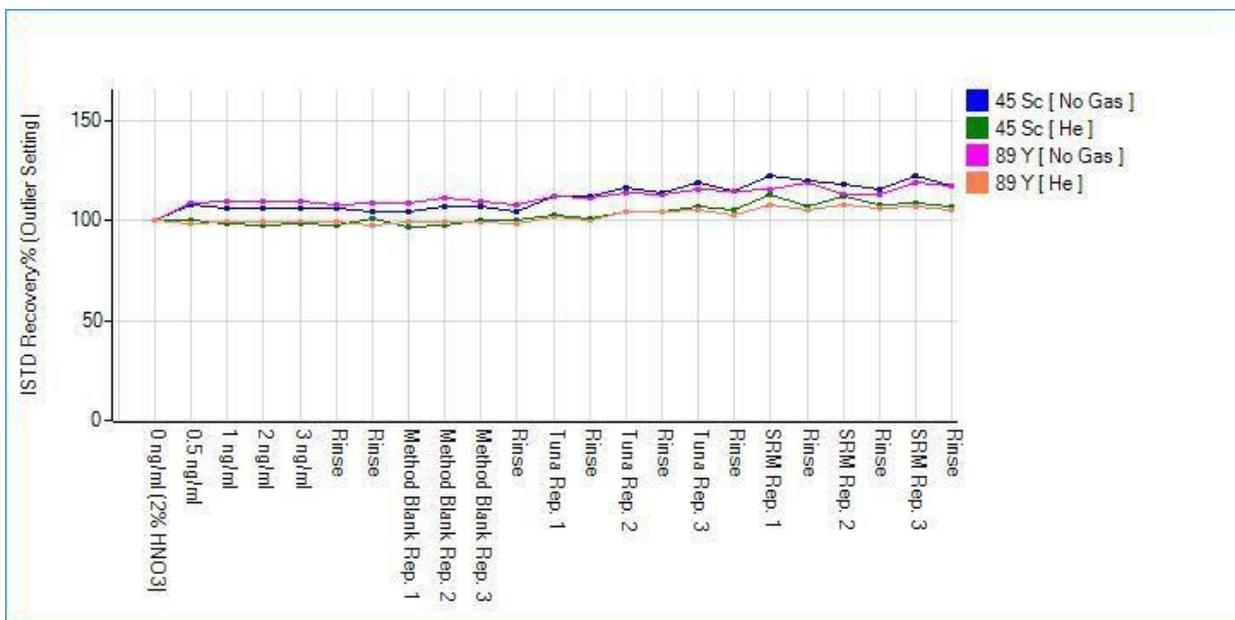
**Figure S16. Internal Standard recovery plot for day 8, 2% HNO$_3$/0.34% H$_2$O$_2$, KED 4V.**
Internal standard percent recovery during no gas mode and helium mode for the above data set.

**Survey results from student volunteers (N=5)**

**Q1: How would you rate your current understanding of ICP-MS experimental design?**
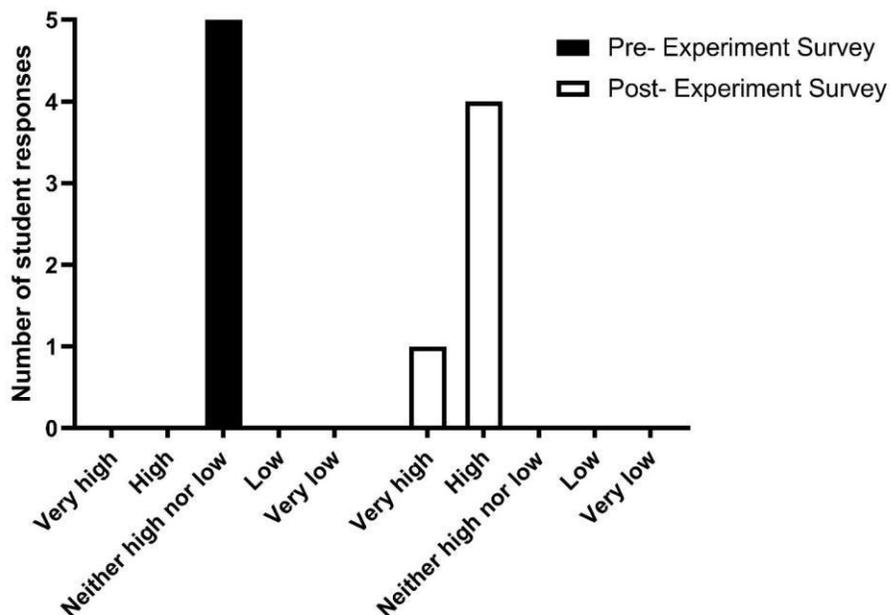


**Q2: How would you rate your current understanding of polyatomic interference in ICP-MS?**
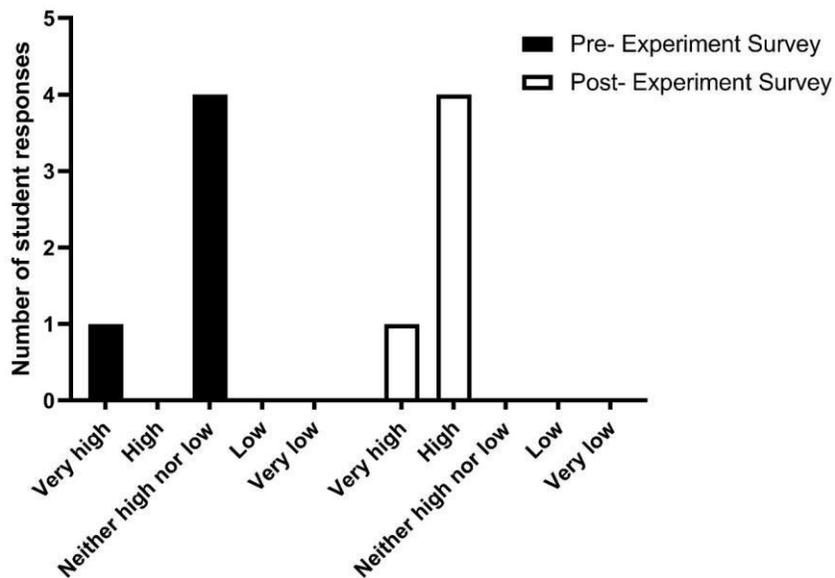
**Q3: How would you rate your current understanding of the concept behind, purpose of and use for standard reference materials (SRMs/CRMs)?**
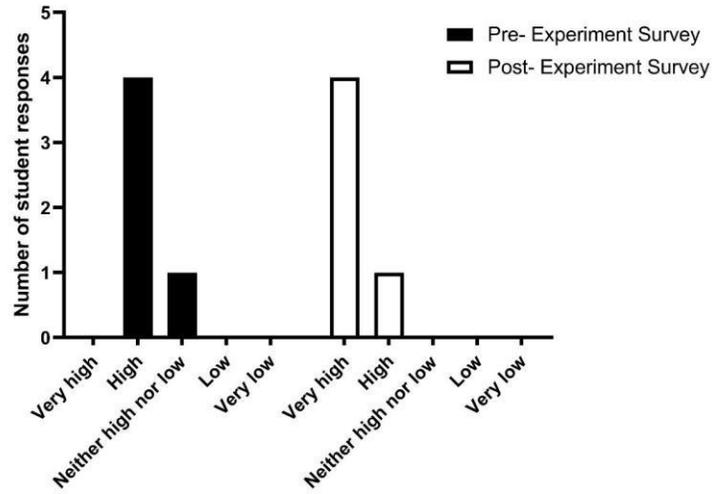


**Q4: How would you rate your current understanding of the purpose of using different gas modes, including no-gas mode, in an ICP-MS experiment?**
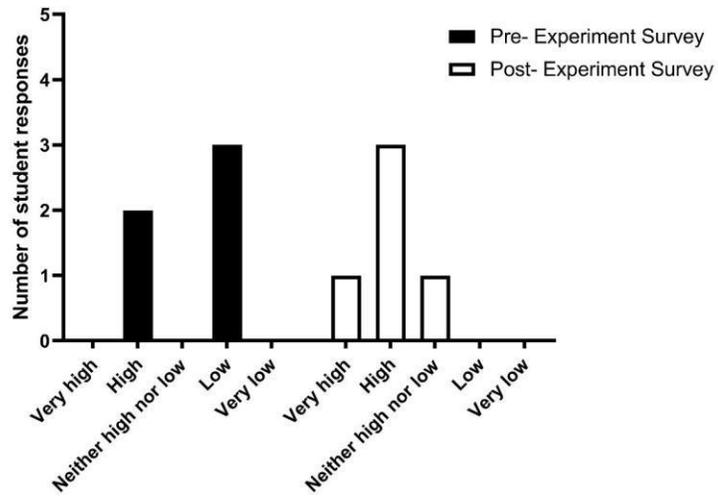
**Q5: How would you rate your understanding of the concept and purpose of ICP-MS method validation?**
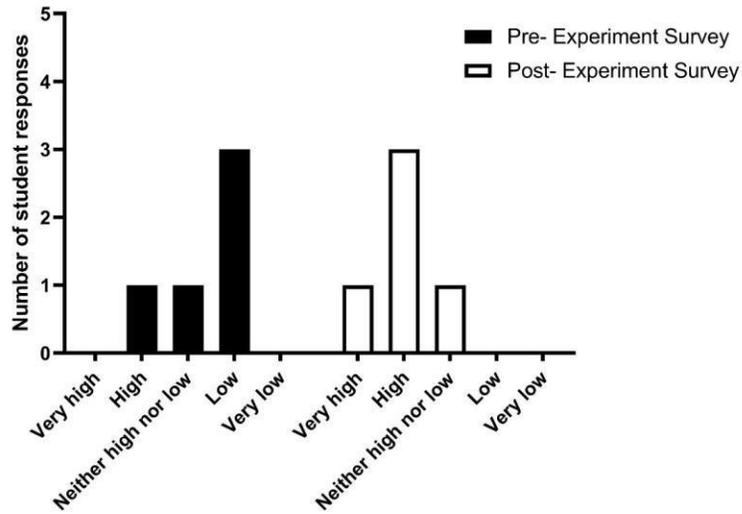


**Q6: How would you rate your current understanding of the concepts "Limit of Detection (LOD)" and "Limit of Quantitation (LOQ)" and how these relate to ICP-MS?**
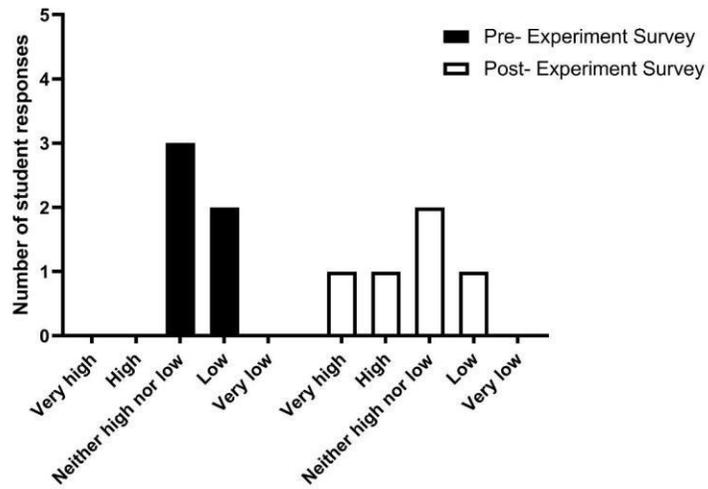
**Q7: How would you rate your current understanding of how the choice of acid utilized in sample digestion affects sample storage?**
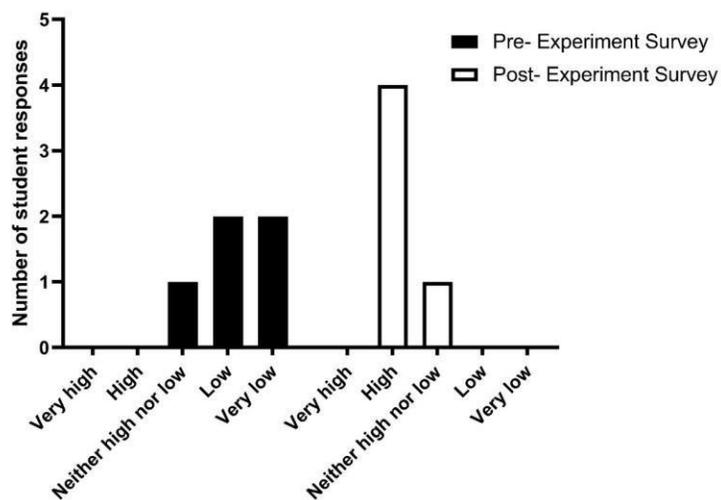


**Q8: Would you feel comfortable explaining the concept and purpose of kinetic energy discrimination (KED) in ICP-MS to another student?**
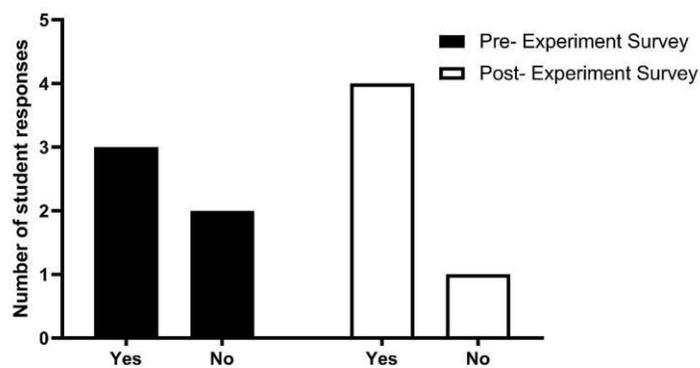
# Best practices and benchmarks for intact protein analysis for top-down mass spectrometry

Daniel P. Donnelly [1,16], Catherine M. Rawlins [1,16], Caroline J. DeHart[2], Luca Fornelli[2], Luis F. Schachner [2], Ziqing Lin[3], Jennifer L. Lippens [4], Krishna C. Aluri [1,5], Richa Sarin [1,6], Bifan Chen [3], Carter Lantz[7], Wonhyeuk Jung[7], Kendall R. Johnson[1], Antonius Koller[1], Jeremy J. Wolff[8], Iain D. G. Campuzano [4], Jared R. Auclair [9], Alexander R. Ivanov[1], Julian P. Whitelegge [10], Ljiljana Paša-Tolić [11], Julia Chamot-Rooke [12], Paul O. Danis[13], Lloyd M. Smith[14], Yury O. Tsybin [15], Joseph A. Loo [7], Ying Ge [3], Neil L. Kelleher [2] and Jeffrey N. Agar [1]*

**One gene can give rise to many functionally distinct proteoforms, each of which has a characteristic molecular mass. Top-down mass spectrometry enables the analysis of intact proteins and proteoforms. Here members of the Consortium for Top-Down Proteomics provide a decision tree that guides researchers to robust protocols for mass analysis of intact proteins (antibodies, membrane proteins and others) from mixtures of varying complexity. We also present cross-platform analytical benchmarks using a protein standard sample, to allow users to gauge their proficiency.**

Mutations, polymorphisms, RNA processing and post-translational modifications (PTMs) such as acetylation, methylation and phosphorylation can lead to a single gene producing many functionally distinct 'proteoforms'[1]. These proteoforms can have different effects on important biological processes, including gene regulation, cell signaling and protein activity; consequently, the ability to characterize these species is essential for an understanding of the biological response to disease. The identity of a proteoform can often be inferred[2] from an accurate experimentally determined intact mass[3]. One can increase the sensitivity of intact-mass-based proteoform identification by determining the relative abundance of a particular amino acid by using isotopic labeling, by using mass similarities to cluster proteoforms into gene families and by reducing the search space using sample-specific search databases[2]. Localizing PTMs, and in some cases the definitive proteoform identifier, requires tandem mass spectrometry (MS[n]) analysis. The measurement of intact protein mass followed by MS[n] has been coined 'top-down' mass spectrometry[4–8], with its origins in Fenn and colleagues' discovery that large biomolecules could be ionized[9] and fragmented[10–12] using electrospray ionization (ESI)-MS. Top-down MS protocols, unlike widely used bottom-up protocols[13,14], do not require endoproteinase digestion before analysis, do not conflate proteoforms and tend to complement native MS analysis.

One advantage of ESI over the alternative 'soft' ionization method, matrix-assisted laser desorption-ionization (MALDI), is that ESI imparts more charge per protein. This enables the mass determination of large biomolecules using mass analyzers with moderate mass-to-charge ratio upper limits (for example, $m/z \leq$ 4,000), which happen to offer the highest resolving power. Higher charge per molecular mass also facilitates gas-phase fragmentation and, therefore, the characterization of primary sequence and PTMs by MS[n] (refs. [15,16]). Because of this superior fragmentation and the ability to interface with liquid chromatography (LC) systems, ESI is used for most top-down MS experiments. Projects requiring rapid MS analysis[17], the ability to analyze hundreds of proteins in a single spectrum, protein imaging capabilities, or less signal suppression by common protein buffer components[18] may be better suited for MALDI-MS.

Compared to bottom-up workflows, top-down approaches provide additional layers of information, including detecting modifications that are removed or scrambled[19] during peptide sample preparation (for example, *S*-thiolation), elucidating functional relationships (for example, cross-talk) between PTMs on the same protein molecule, characterizing drug–target interactions, observing important modifications on biopharmaceuticals, and identifying and quantifying distinct proteoforms that would have been convoluted by endoproteinase digestion[20–24]. In addition, sample preparation for intact protein MS comprises fewer steps than bottom-up approaches and does not require chemical modification (for example, reduction and alkylation), thereby reducing the number of

[1]Barnett Institute of Chemical and Biological Analysis and Departments of Chemistry & Chemical Biology and Pharmaceutical Sciences, Northeastern University, Boston, MA, USA. [2]Departments of Chemistry and Molecular Biosciences and the Proteomics Center of Excellence, Northwestern University, Evanston, IL, USA. [3]Department of Cell and Regenerative Biology, Department of Chemistry, Human Proteomics Program, University of Wisconsin-Madison, Madison, WI, USA. [4]Amgen Research, Discovery Attribute Sciences, Amgen, Thousand Oaks, CA, USA. [5]Alnylam Pharmaceuticals, Cambridge, MA, USA. [6]Biogen, Cambridge, MA, USA. [7]Department of Chemistry and Biochemistry, Department of Biological Chemistry, and UCLA/DOE Institute of Genomics and Proteomics, University of California, Los Angeles, Los Angeles, CA, USA. [8]Bruker Daltonics, Billerica, MA, USA. [9]Biopharmaceutical Analysis Training Laboratory, Northeastern University, Burlington, MA, USA. [10]The Pasarow Mass Spectrometry Laboratory, The Jane and Terry Semel Institute for Neuroscience and Human Behavior, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA, USA. [11]Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland, WA, USA. [12]Mass Spectrometry for Biology Unit, Institut Pasteur, USR 2000, CNRS, Paris, France. [13]Eastwoods Consulting, Boylston, MA, USA. [14]Department of Chemistry, Genome Center of Wisconsin, University of Wisconsin-Madison, Madison, WI, USA. [15]Spectroswiss, Lausanne, Switzerland. [16]These authors contributed equally: Daniel P. Donnelly, Catherine M. Rawlins. *e-mail: j.agar@northeastern.edu

experimental artifacts[25]. Current top-down sample cleanup methods (for example, protein precipitation[26] and molecular weight cutoff (MWCO) ultrafiltration) are not applicable to all sample types or downstream MS analyses. The demand for robust, generally applicable methods for intact protein MS is the most common request made to members of the Consortium for Top-Down Proteomics[27,28] (http://topdownproteomics.org/).

Our goal here is to address this unmet need, by providing a guide to enable users with all levels of expertise to acquire high-quality intact protein mass spectra by ESI-MS. First we discuss signal suppression associated with common buffer components and biotherapeutic excipients. This provides the rationale for most failed intact MS measurements and, in addition, a path to designing MS-compatible buffers. Then, we present a decision tree based on sample composition and experimental goals, which guides users to a best-practices protocol and corresponding benchmark data.

## Origins of signal suppression and signal spreading

Biological, biochemical and biotherapeutic sample preparations usually contain numerous interfering substances (for example, salts, detergents, chaotropes and buffers) that lead to signal suppression during ESI-MS analysis. To provide a theoretical context, we describe the two major drivers of the quality of intact protein (positive ion) ESI-MS and how these are affected by interfering substances. The first driver of quality is the formation of desolvated protein ions, which can be understood in terms of a few critical steps during the ESI process[16,29,30]. Interfering substances generally affect the ESI process after the formation of nanodroplets at the Rayleigh charge limit. Two salient, often opposing, processes that occur within these nanodroplets are the partitioning of net charge toward the droplet surface and the minimization of solvation energy. Polar species such as salts and native proteins partition toward the droplet interior to optimize solvation energy; their ionization, therefore, requires evaporation of solvent molecules[16]. Hydrophobic species such as detergent monomers and unfolded proteins migrate to the droplet surface to optimize solvation energy and, in a faster process that requires less energy, evaporate or are ejected. Many of the techniques presented here for reducing signal suppression can be rationalized within the framework above. For example, organic solvents that decrease surface tension should promote the ionization of both polar and nonpolar analytes; detergents partition to the surface, where they can outcompete analytes for a limited number of protons; organic solvents and acids that unfold proteins should promote ejection-based ionization; native MS (nMS) requires greater desolvation energy and is more sensitive to polar contaminants.

The second driver of the quality of intact protein MS is signal spreading (that is, the distribution of the signal from a single proteoform across multiple channels), which increases with protein mass. Each channel has its own respective noise; consequently, the cumulative noise increases proportionally to the number of channels. The ESI process promotes signal spreading, via adduct formation, by increasing the concentrations of interfering substances and proteins. Heavy isotopes and charge states further distribute signal intensity across multiple channels; the former can be mitigated by isotope depletion[31]. Here we describe experimental techniques that minimize signal spreading (increase signal-to-noise ratio, or S/N), including using nMS to reduce the number of charge states, and the use of volatile salts (for example, ammonium acetate) or purification to minimize the effects of alkaline salts.

## Signal suppression by common buffer components

Using the intact protein standard mixture (ubiquitin, myoglobin, trypsinogen and carbonic anhydrase) established by the National Resource for Translational and Developmental Proteomics (NRTDP) (http://nrtdp.northwestern.edu/protocols/), we evaluated common buffer components (Fig. 1) to quantify the concentration
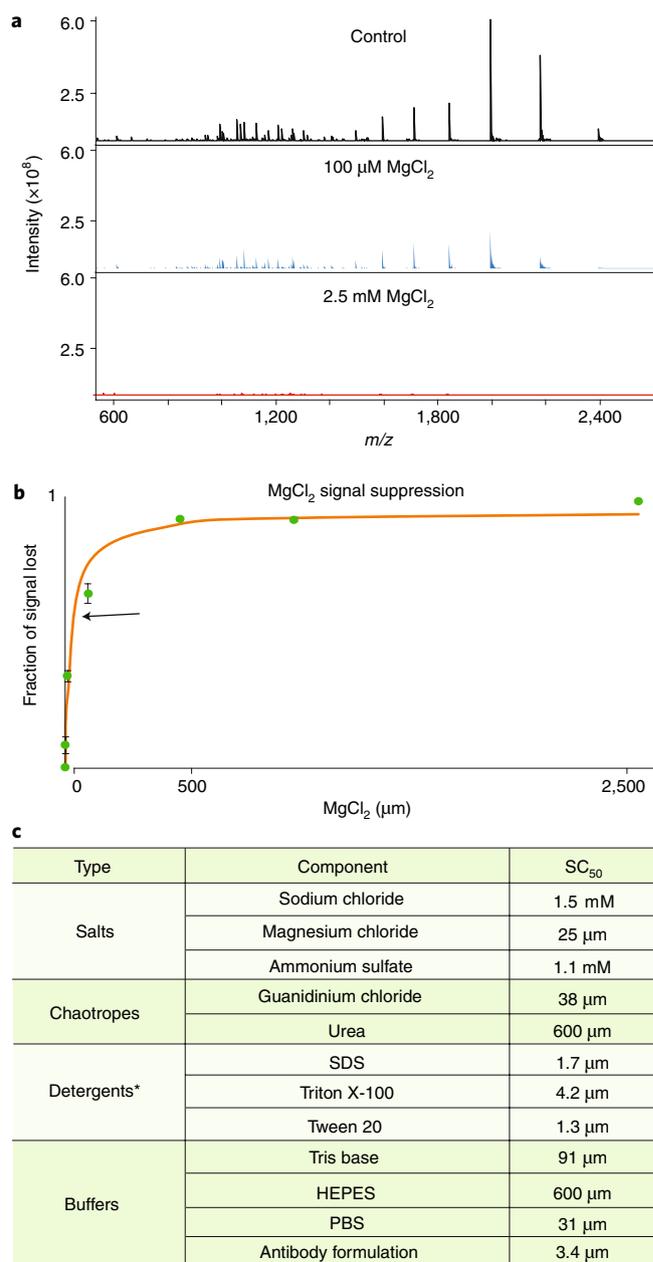


**Fig. 1 | Common buffer components suppress MS signal. a**, $MgCl_2$ reduces signal (and S/N) in a concentration-dependent manner. **b**, Fit of experimental data to determine the concentration of $MgCl_2$ required for 50% signal suppression ($SC_{50}$; black arrow), **c**, Table of common buffer components and the concentration threshold for 50% $SC_{50}$ (experimental data curves and their fits are shown in Supplementary Fig. 1) and calculations in Supplementary Note 2. Detergents compatible with mass spectrometry are discussed in Protocol 2b. *Signal suppression by detergents is less pronounced above their critical micellar concentration (CMC) (described in Protocol 2b).

required for 50% signal suppression during direct infusion ESI. By analogy to half-maximum inhibitory concentration ($IC_{50}$) nomenclature, we termed this metric the half-maximum suppression concentration ($SC_{50}$) (Fig. 1, Supplementary Fig. 1). At their typical concentrations, all common buffer additives suppressed ESI signal considerably. Consistent with the mechanisms of ESI ionization described above, detergents produced the most signal suppression,
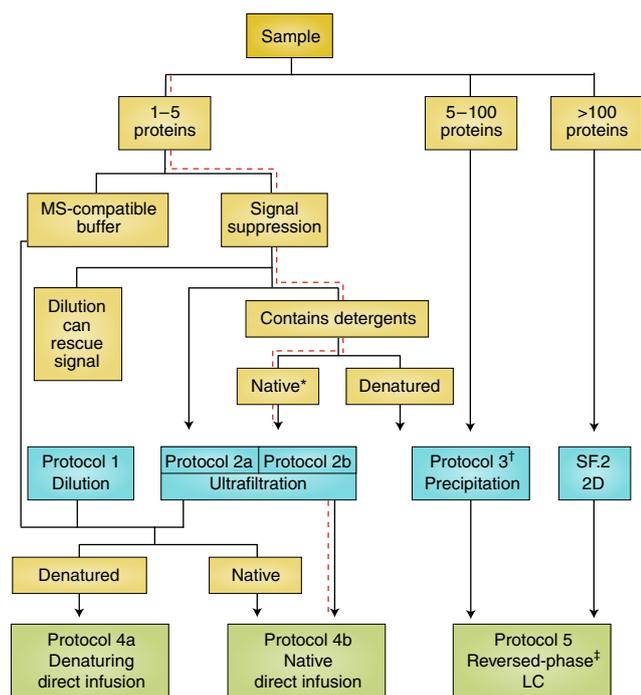
**Fig. 2 | Decision tree for intact protein sample clean-up, preparation and analysis.** The red dashed line, for example, denotes the decision path for the native MS analysis of a membrane protein. *LC can also be applied at this stage in the decision tree. †Minimally complex protein samples prepared following Protocol 3 can be analyzed via denaturing direct infusion (Protocol 4a) if desired. Supplementary Fig. 2 (SF.2) shows a recommended example of 2D separation following the GELFrEE protocol. ‡Other viable alternative separation techniques include capillary zone electrophoresis, ion exchange and size exclusion chromatography.

less volatile (for example, metallic) salts produced intermediate suppression and volatile components lowest suppression. Additional details of the experimental parameters used here are provided in Supplementary Notes 1 and 2.

The $SC_{50}$ values given in Fig. 1 allow users to design MS-compatible buffers. In addition, the $SC_{50}$ and buffer composition serve as the entry point into the decision tree outlined below, leading users to the appropriate protocol. Although the trends in $SC_{50}$ values reported here should generally be consistent across MS platforms, parameter-dependent variations in the reported values are likely (in particular, flow rate, voltages, temperatures, and pressures that affect ionization and desolvation). Here, for example, we calculated $SC_{50}$ obtained by direct infusion using a standard microflow ESI source (about a microliter per minute), but nano-ESI (less than a microliter per minute) is less affected by salts because of the order-of-magnitude decrease in initial droplet size[32,33].

## Intact protein MS (IPMS) decision tree

The IPMS decision tree (Fig. 2) directs practitioners to a protocol or a combination of protocols based on buffer composition, the number of proteins in the sample, and whether native or denaturing conditions are to be used. Consider, for example, a purified protein in phosphate-buffered saline (PBS). Based on the 1.5 mM $SC_{50}$ exhibited by NaCl (Fig. 1c) and the 137 mM NaCl present in PBS, a protein sample in PBS requires a 91-fold dilution to achieve 50% of the potential MS signal. Therefore, if the protein concentration is greater than 90 μM and salt adducts will not impede data analysis, the sample can be diluted following Protocol 1. Otherwise, sample cleanup by ultrafiltration using spin cartridges with a MWCO-membrane is recommended following Protocol 2.

Interest in certain PTMs (for example, metallation) or protein complex quaternary structure would dictate the use of native MS methods following Protocol 4b; otherwise the denaturing MS Protocol 4a is recommended. Depending on the complexity of the sample, additional separation techniques such as GELFrEE may be required (Supplementary Fig. 2). The objective of this decision tree is to provide a proven workflow for any sample, not to rule out alternative methods. For example, depending on sample stability, user expertise and available resources, precipitation (Protocol 3), size exclusion 'spin cartridges', or LC (Protocol 5) could be suitable alternatives to MWCO ultrafiltration. All protocols and benchmarks referenced by the decision tree and alternative methods are summarized below and further detailed in Supplementary Notes 1–5 and Supplementary Protocols 1–5.

## Protein standards and benchmarks

To promote standardization and allow users to benchmark their own data using readily available proteins, we provided representative results for each protocol using the following commercially available standards: (i) the NRTDP intact protein standard mixture (see Supplementary Note 1 for preparation instructions), (ii) NIST monoclonal antibody reference material 8671 (NISTmAb), containing humanized IgG1κ in 12.5 mM L-histidine, 12.5 mM L-histidine HCl (pH 6.0), and (iii) Sigma bacteriorhodopsin from *Halobacterium salinarum* (B0184). Benchmarks for mass accuracy depend upon the instrumentation platform and have been reviewed[3,34–39]. Rules of thumb include requiring 10 p.p.m. accuracy for modern Fourier transform MS and 20 p.p.m. accuracy for modern quadrupole time-of-flight (QTOF) MS. We suggest the use of ProForma notation[40] for standardized proteoform nomenclature, and note that the PeptideMass tool (https://web.expasy.org/peptide_mass/) can be used to calculate the mass of a given sequence or of proteoforms contained in the UniProt database.

## Protocol 1: sample preparation by dilution of interfering substances

Consistent with the mechanisms of ESI and signal spreading detailed above, common buffer components render proteins undetectable by MS (Fig. 3). Minimally complex, concentrated protein solutions can often be analyzed by direct infusion, following dilution to ~1 μM final protein concentration in the appropriate sample buffer. Users should consider using this protocol if dilution can decrease the concentration of a given interfering substance below its $SC_{50}$ value (Fig. 1, Supplementary Protocol 1). Assuming a practical upper limit of ~10 mM protein concentration, this protocol is potentially applicable to any of the components listed in Fig. 1. As detailed above, however, nMS utilizes an ESI process that is more sensitive to many interferents, including salts. Consequently, dilution is less likely to adequately improve nMS. Protocol 4 describes methods to dilute native proteins into whichever solution will be used to introduce samples to the MS. However, mass spectra obtained by this method have the lowest S/N of any of the protocols described here and may contain adducts.

## Protocol 2: sample preparation using MWCO ultrafiltration

We recommend remediating nonvolatile salt adducts by buffer exchange into a solution of volatile salts. The MWCO of the ultrafiltration device should not exceed half the molecular mass of any given protein in a sample to prevent possible sample loss. No particular pH is optimal for all proteins, but pH extremes should be avoided, as should pH that is equivalent to a protein's pI, where protein solubility is at a local minimum[41]. We recommend using ammonium acetate throughout these protocols owing to its volatility and ability to act as a stabilizing background electrolyte during ESI[42]. Ammonium acetate provides maximal buffering around pH 4.75 (acetate) and 9.25 (ammonium), and results in a neutral pH
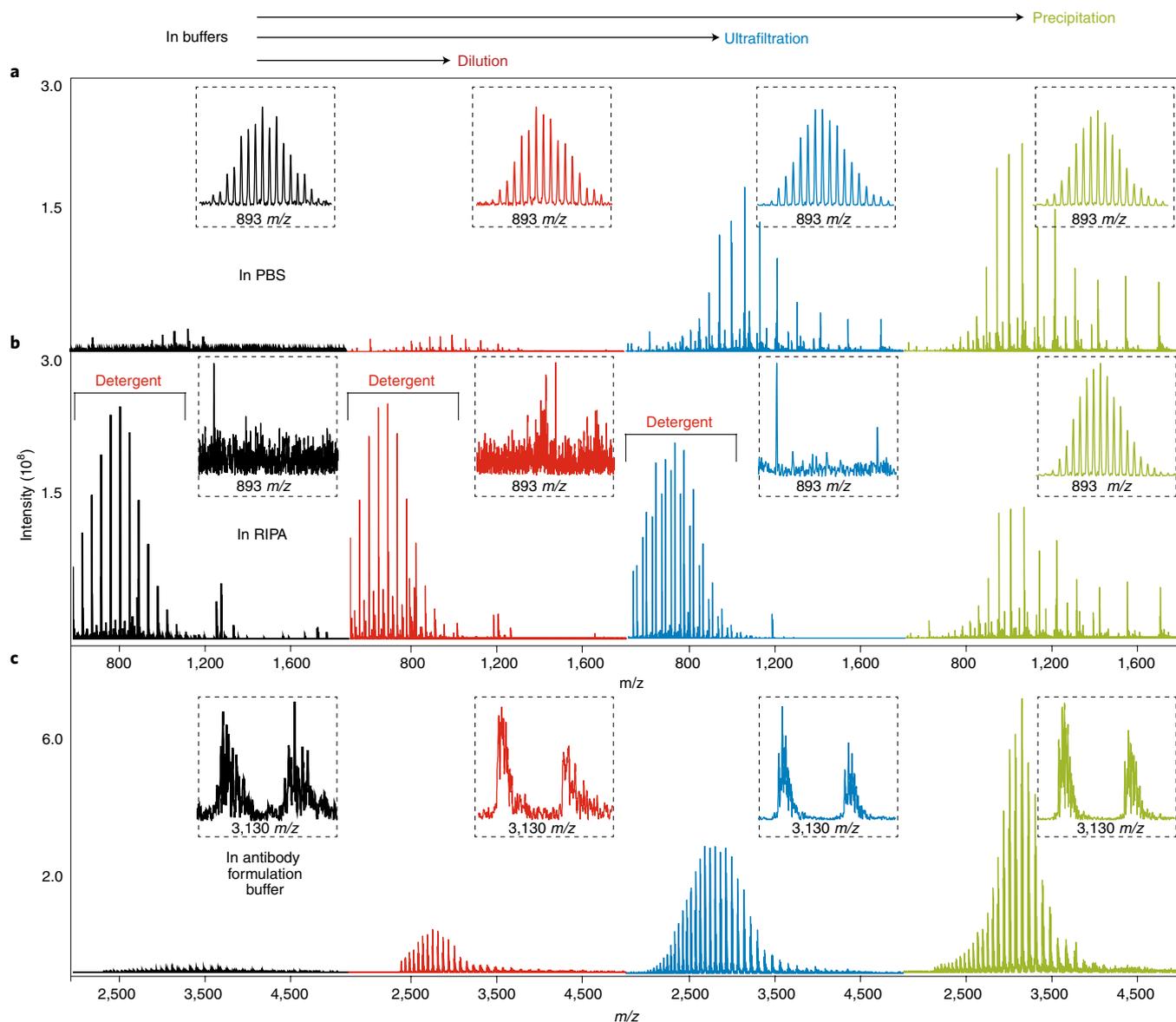
**Fig. 3 | Dilution (Protocol 1), MWCO ultrafiltration (Protocol 2a) and precipitation (Protocol 3) sample preparation protocols applied to common buffers. a,b**, Protein standard mixture in PBS (**a**) and detergent-containing RIPA buffer (**b**). In the buffer containing harsh detergents, protein signal is attained only with precipitation. **c**, NISTmAb in 12.5 mM L-histidine, 12.5 mM L-histidine HCl (pH 6.0). Mass spectra were obtained using a Fourier transform ion cyclotron resonance MS (FT-ICR) (Bruker Daltronics SolariX 9.4T MS) using denaturing direct infusion (Protocol 4a). See Supplementary Fig. 3 for additional results with 'gentle elution' immunoaffinity elution buffer and a second antibody buffer.

upon dissolving in water (approximate pH 6.5–7). Before adding protein sample, MWCO-ultrafiltration devices should be rinsed with the appropriate buffer. Additional details for this method can be found in Supplementary Protocol 2.

**Protocol 2a, soluble proteins.** On the basis of the protein masses in the NRTDP intact protein standard, we recommend using a MWCO of 3 kDa according to the manufacturer's instructions. The protein preparation should be subjected to three (1:20 dilution) buffer exchanges into 10 mM ammonium acetate (pH 6.5) using a MWCO-ultrafiltration device, followed by an additional three exchanges into 2.5 mM ammonium acetate (pH 6.5) (exemplary data in Fig. 3, Supplementary Figs. 3 and 4, see also Supplementary Protocol 2a and Supplementary Note 3). Denaturing and nondenaturing samples can then be diluted and introduced to the MS as described below in Protocol 4a.

**Protocol 2b, native membrane proteins.** Membrane proteins are estimated to account for 23% of the total human proteome and represent ~60% of targets for currently approved drugs[43,44]. The mass analysis of native, intact membrane proteins can further provide key information regarding stoichiometry, ligand binding and lipid association. A typical analysis of a membrane protein complex requires either size-exclusion chromatography (SEC) or MWCO ultrafiltration to remove alkali salt adducts while maintaining the detergent used to solubilize the protein (Supplementary Protocol 2b)[45]. This differs fundamentally from the MWCO ultrafiltration used during filter-aided sample preparation (FASP) to improve the bottom-up proteomics analysis of membrane proteins, which removes detergents[46,47]. For users interested in native membrane proteins, we recommend the protocols of Robinson and coworkers[45]. Their protocols are based on comprehensive optimization and include a complete list of non-ionic detergents compatible with MS and

detailed sample preparation considerations. We demonstrate an example application of Robinson and coworkers' protocols for the native tetramer of Aquaporin Z (AqpZ) from *Escherichia coli* (Supplementary Fig. 5).

## Protocol 3: sample preparation using protein precipitation

Common precipitation protocols use organic solvents to agglomerate proteins while leaving small molecules, including salts and detergents, solubilized. Whereas MWCO ultrafiltration using Protocol 2a does not rescue protein signal from a preparation containing harsh surfactants (for example, SDS and Triton), precipitation of proteins following Protocol 3 does (Fig. 3, Supplementary Protocol 3). A volume ratio of 1:1:4:3 of aqueous protein sample:chloroform:methanol:water is recommended to precipitate proteins[26]. The supernatant is removed by aspiration, and the precipitated pellet can be further washed with one more addition and removal of methanol. Pellets are resolubilized for 15 minutes at −20 °C using a small volume of 80% (v/v) formic acid (~25% of the starting volume) and are then diluted to the starting sample volume with HPLC-grade water or a solution of volatile salts (for example, ammonium acetate)[48]. As an alternative method, acetone precipitation has the distinct advantage of leaving many proteins folded. This method, however, has been shown to modify proteins with +98 Da adducts[49], requires longer incubation at −20 °C (at least 1 h), requires that all steps be performed at or below 0 °C to maximize resolubilization, and can be compromised by detergents.

## Protocol 4a: denaturing direct-infusion MS

Denaturing direct-infusion ESI mass spectra can usually be obtained by introducing samples to the MS in a mixture of 49.95% HPLC-grade acetonitrile, 49.95% HPLC-grade water, and 0.1% formic acid (v/v). A 60:35:5 ratio of HPLC-grade methanol:water:acetic acid may be used as an alternative and, in some cases, can improve S/N[9,50]. As described above, the use of these organic solvents and acids results in efficient ionization from a droplet's surface, often allowing MS analysis to be performed using instrumentation parameters typically used for peptides. A more detailed description of instrument parameters for the Bruker SolariX FT-ICR MS used during denaturing direct infusion studies is found in Supplementary Protocol 4a.

## Protocol 4b: native direct-infusion MS

Although native MS protocols may not necessarily produce folded ions that match exactly to their in-solution structures, they can be used to achieve accurate mass measurements of native structures and complexes[51]. Consequently, native direct-infusion MS can provide unique structural information, including the characterization of labile PTMs, metal-binding sites, noncovalent interactions with small molecules, and protein tertiary and quaternary structure. Detergent-free samples can be infused directly in aqueous 2.5 mM ammonium acetate[52], the same solution used in the final stage of Protocol 2a (concentrations of ammonium acetate up to 500 mM can even be used).

Figure 4 compares mass spectra of carbonic anhydrase in denatured and native states, with the intensity of the base peak in the native sample being about twofold higher than that of the denatured sample. This comparison was repeated in four additional labs on six different instruments to illustrate the possible range of relative intensities (Supplementary Fig. 6, Supplementary Protocol 4b). Membrane protein complexes with MS-compatible detergents can be infused directly from the final solution described in Protocol 2b[45]. To observe native membrane proteins, detergent ions must be removed from the protein–micelle complex by increased collisional activation. This may be achieved through an increase in collision voltage applied to the source or the collision cell (typically 50−200 V), but it could require additional critical parameters that
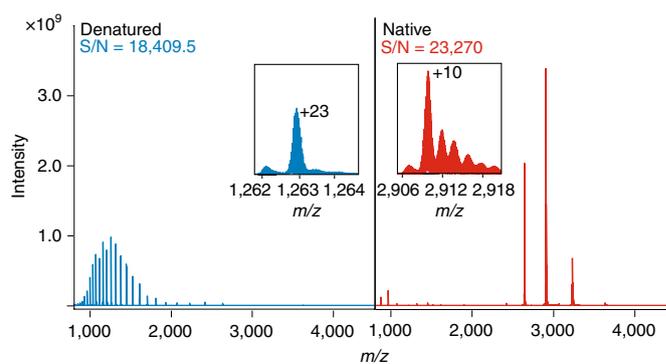


**Fig. 4 | Denatured versus native ESI-MS of carbonic anhydrase.** Intensity is scaled to demonstrate the difference between denaturing MS (left) and native MS (right). These spectra were collected on the same instrument using the same concentration (10 μM). Native MS results in lower and fewer charge states, and thus the signals have higher intensity and appear at a higher *m/z*. The inset includes the most abundant charge state and the S/N.

are described in detail by Robinson and coworkers, and in part in Supplementary Protocol 4b[45,53,54].

## Protocol 5: intact protein analysis using LC-MS

Ionization suppression by excipients and by other proteins generally makes the analysis of multiple proteins and proteoforms by direct infusion intractable. For example, many 'high-purity' proteins (as judged by SDS-PAGE) contain numerous proteoforms that cannot be reliably detected and quantitatively assessed without up-front separation[55,56]. Liquid phase separation approaches, including LC (for example, reversed-phase (RP), size-exclusion, ion exchange, chromatofocusing) and capillary electrophoresis techniques (for example, capillary zone electrophoresis, capillary isoelectric focusing) can remove excipients and provide the resolving power for deep characterization of proteins. As directed in the decision tree (Fig. 2), separation of particularly complex samples (>100 proteins) requires an additional dimension of separation before LC-MS. Supplementary Fig. 2 shows the use of GELFrEE separation prior to LC, which fractionates samples on the basis of protein molecular weights and has resulted in the largest number of characterized proteoforms to date[57].

**Protocol 5a: LC-MS of soluble proteins.** RP-LC is recommended for all samples containing more than five unique proteins but is also a viable option for samples with fewer proteins, provided they do not contain high salt concentrations (>1 M) or harsh detergents. The recommended reversed-phase LC protocol is described in Supplementary Protocol 5a and at http://nrtdp.northwestern.edu/protocols/.

Figure 5 demonstrates that sufficient intact MS signal was attained, and four unique chromatographic peaks were observed, using Protocol 5a with a PLRP-S stationary phase (1,000-Å pore size, 5-μm particle size) on a Dionex UPLC coupled to a Thermo Orbitrap Elite. We provide benchmarks for this standard operating procedure (SOP), as well as for additional data acquired using Monolithic and C4 stationary phases, for six widely used platforms (Waters Xevo G2-S QTOF, Supplementary Fig. 7; Bruker Impact II QTOF and Bruker SolariX FT-ICR, Supplementary Fig. 8; Thermo Orbitrap Elite, Thermo Orbitrap Fusion Lumos, and Thermo Orbitrap QE-HF, Supplementary Figs. 9 and 10). To allow users to compare their performance with that of experienced operators using instruments that are operating within specifications, we report S/N for the platforms used here (Fig. 5). However, instrument vendors use proprietary, non-standardized techniques to preprocess data, display data and determine S/N, and, as a result, our data cannot
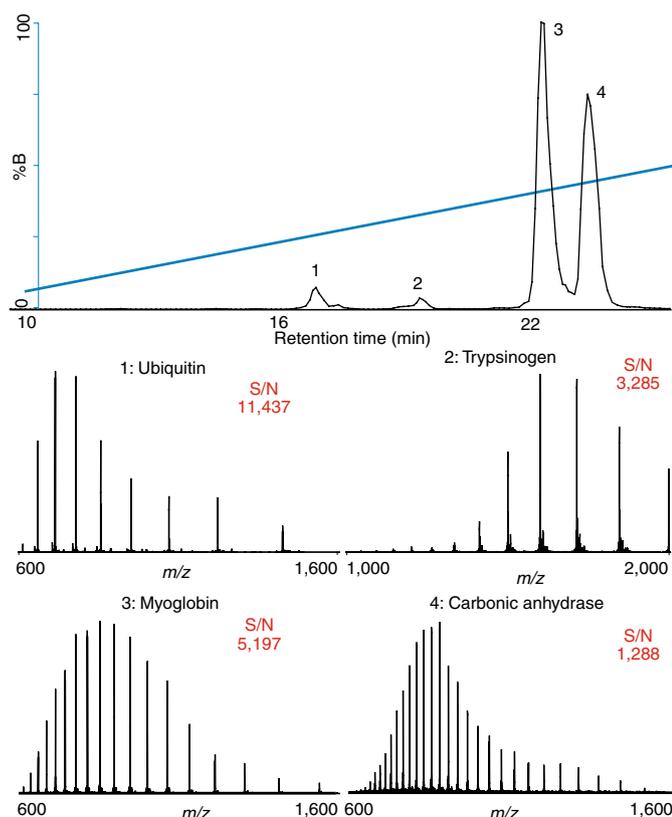
**Fig. 5 | LC-MS of protein standard mixture prepared following Protocol 5a and separated on a Dionex UPLC with a Thermo Orbitrap Elite system using PLRP-S stationary phase.** Final concentrations of each protein loaded onto the column were 0.14 pmol ubiquitin, 0.49 pmol trypsinogen, 1.09 pmol myoglobin and 0.64 pmol carbonic anhydrase (top). Summary of S/N values calculated for each protein on all instrumentation platforms using the given SOP (bottom) including Dionex Ultimate 3000–Thermo Orbitrap Elite, Waters Acquity–Xevo G2-S QTOF, Waters nanoAcquity–Bruker Impact II QTOF, Waters nanoAcquity–Bruker SolariX FT-ICR, Dionex Ultimate 3000–Thermo Fusion Lumos, and Dionex Ultimate 3000–Thermo QE-HF. As described, S/N calculations differ per manufacturer and do not reflect absolute performance.

be used for a cross-platform comparison. As an example of a viable alternative method that is notably better suited for proteoforms with similar mass and RP-LC retention (for example, deamidation), we provide a separation of the same protein mix using capillary zone electrophoresis (Supplementary Fig. 11).

**Protocol 5b: intact membrane protein LC-MS.** Denaturing LC-MS of intact membrane proteins is not straightforward because of their inherent hydrophobicity[58,59]. Whitelegge et al. provided the earliest example of denaturing LC-MS of membrane proteins using high concentrations of mobile phase additives and demonstrated that ESI of membrane proteins could achieve the 0.01% mass accuracy benchmark established for ESI of soluble proteins[58]. For
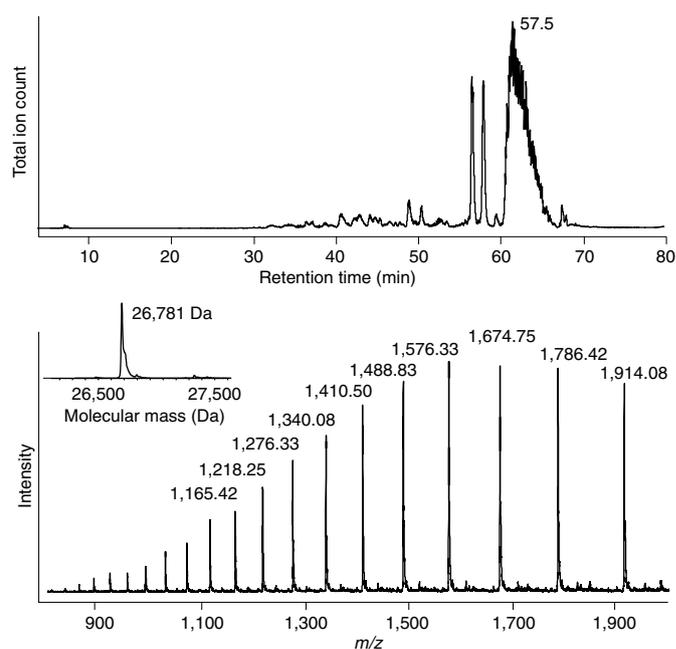


**Fig. 6 | LC-MS of bacteriorhodopsin-containing purple membrane of *H. salinarum* prepared following Protocol 5b and analyzed on an Agilent HPLC system coupled to a Thermo linear ion trap (LTQ) mass spectrometer.** Proteins were separated using an Agilent PLRP-S 300 Å, 2.1 × 150 mm, 3 µm. Supplementary Fig. 12 demonstrates this analysis on four additional instrumentation platforms.

thorough reviews of the current state of membrane protein analysis via LC-MS[60,61] and the corresponding protocols, we direct readers to refs. [60–62].

Current denaturing LC-MS methods for membrane proteins use either size-exclusion[63,64] or reversed-phase separation. Owing to the ease of implementation across a variety of MS platforms, we suggest analysis via reversed-phase LC-MS using a polystyrene-divinyl benzene co-polymer stationary phase (PLRP-S, 300 Å, Agilent). We do not recommend the use of long chain bonded stationary phases such as C8 and C18, as membrane proteins are likely to be retained on the column. As an example, we solubilized enriched bacteriorhodopsin from *H. salinarum* (Sigma B0184) in 88% formic acid to separate the protein from lipid contaminants. To avoid the risk of formic acid adduction (+28 Da), samples are immediately injected onto the column and solvent exchanged to much lower acid concentrations (0.1%). In the case of membrane protein preparations containing high enough concentrations of lipid contaminants to confound analysis or damage the column, we recommend precipitation following Protocol 3 before analysis. Proteins are eluted using an increasing gradient of 49.95% acetonitrile, 49.95% isopropanol, 0.1% formic acid. Figure 6 shows the analysis of denatured bacteriorhodopsin of *H. salinarum* following this protocol. Although elution efficiency for some integral proteins may fall well below 100%, PLRP-S columns can be regenerated with 90% formic acid injections. This protocol was performed in five labs on five different instrument platforms (Supplementary Fig. 12, Supplementary Protocol 5b). An example of an alternative LC-MS method using a more common stationary phase (ZORBAX RRHD 300SB-C3) is provided for aquaporin Z in Supplementary Fig. 5b.

## Special methodological considerations for intact antibody mass spectrometry

With the increasing development of biotherapeutics and biosimilars in the pharmaceutical industry, and an increasingly stringent route

to regulatory approvals, there is a growing need for intact antibody MS. Every protocol presented here can be applied to the analysis of intact antibodies (Fig. 3, Supplementary Fig. 13, Supplementary Note 4). However, as antibodies are relatively large and signal spreading increases in proportion to protein size, we recommend against the use of Protocol 1 (dilution) for any regulatory filing.

## Discussion

The IPMS decision tree (Fig. 2) guides practitioners of all levels toward broadly applicable methods to obtain high-quality intact mass spectra from any protein sample. The protocols described here have been scrutinized and optimized in over ten expert intact protein MS labs, and successfully applied in laboratories without experience in intact protein MS. We hope that these protocols will enable any research group to adopt intact protein mass analysis.

The accurate mass measurement of an intact protein is the *sine qua non* of top-down mass spectrometry, which can characterize how proteoforms interact and identify PTMs that are lost in other analyses. High-throughput top-down analysis of whole proteomes has proven successful in the unambiguous identification of hundreds of proteins and proteoforms from a single biological sample[65] and revealed prevalent yet previously uncharacterized biologically relevant modifications[66]. Quantitative top-down proteomics has been used to identify disease-relevant differences in protein levels, an encouraging step forward in the field of proteomics-based personalized medicine[67]. Additionally, by using native mass spectrometry following the top-down workflow, one can observe previously unknown protein–protein interactions, protein–ligand binding, protein–cofactor association and protein-complex stoichiometry, and assess their relationships to important biological pathways[68]. We believe that by starting with intact mass analysis, using these intact protein MS protocols coupled to top-down MS analysis, and by identifying proteoforms rather than proteins, scientists can gain new insights into the human proteome. We also hope that these protocols serve as a starting point for users to push, even further, the current limits of high-molecular-weight mass spectrometry.

All general protocols are available as Supplementary Protocols.

## References

1. Smith, L. M. & Kelleher, N. L. Proteoform: a single term describing protein complexity. *Nat. Methods* **10**, 186–187 (2013).
2. Shortreed, M. R. et al. Elucidating proteoform families from proteoform intact-mass and lysine-count measurements. *J. Proteome Res.* **15**, 1213–1221 (2016).
3. Mann, M. & Kelleher, N. L. Precision proteomics: The case for high resolution and high mass accuracy. *Proc. Natl Acad. Sci. USA* **105**, 18132 (2008).
4. Shaw, J. B. et al. Complete protein characterization using top-down mass spectrometry and ultraviolet photodissociation. *J. Am. Chem. Soc.* **135**, 12646–12651 (2013).
5. Ge, Y. et al. Top down characterization of larger proteins (45 kDa) by electron capture dissociation mass spectrometry. *J. Am. Chem. Soc.* **124**, 672–678 (2002).
6. Siuti, N. & Kelleher, N. L. Decoding protein modifications using top-down mass spectrometry. *Nat. Methods* **4**, 817–821 (2007).
7. Ge, Y., Rybakova, I. N., Xu, Q. G. & Moss, R. L. Top-down high-resolution mass spectrometry of cardiac myosin binding protein C revealed that truncation alters protein phosphorylation state. *Proc. Natl Acad. Sci. USA* **106**, 12658–12663 (2009).
8. 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).
9. 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).
10. Loo, J. A., Edmonds, C. G. & Smith, R. D. Tandem mass spectrometry of very large molecules: serum albumin sequence information from multiply charged ions formed by electrospray ionization. *Anal. Chem.* **63**, 2488–2499 (1991).
11. Reid, G. E. & McLuckey, S. A. 'Top down' protein characterization via tandem mass spectrometry. *J. Am. Soc. Mass Spectrom.* **37**, 663–675 (2002).
12. Alexandrov, M. L. et al. Extraction of ions from solutions under atmospheric pressure as a method for mass spectrometric analysis of bioorganic compounds. *Rapid Commun. Mass Spectrom.* **22**, 267–270 (2008).
13. McCormack, A. L. et al. Direct analysis and identification of proteins in mixtures by LC/MS/MS and database searching at the low-femtomole level. *Anal. Chem.* **69**, 767–776 (1997).
14. Zhang, Y., Fonslow, B. R., Shan, B., Baek, M.-C. & Yates, J. R. Protein analysis by shotgun/bottom-up proteomics. *Chem. Rev.* **113**, 2343–2394 (2013).
15. Kelleher, N. L. et al. Top down versus bottom up protein characterization by tandem high-resolution mass spectrometry. *J. Am. Chem. Soc.* **121**, 806–812 (1999).
16. Konermann, L., Ahadi, E., Rodriguez, A. D. & Vahidi, S. Unraveling the mechanism of electrospray ionization. *Anal. Chem.* **85**, 2–9 (2013).
17. Lohnes, K. et al. Combining high-throughput MALDI-TOF mass spectrometry and isoelectric focusing gel electrophoresis for virtual 2D gel-based proteomics. *Methods* **104**, 163–169 (2016).
18. Fenselau, C. & Demirev, P. A. Characterization of intact microorganisms by MALDI mass spectrometry. *Mass Spectrom. Rev.* **20**, 157–171 (2001).
19. Auclair, J. R. et al. Artifacts to avoid while taking advantage of top-down mass spectrometry based detection of protein S-thiolation. *Proteomics* **14**, 1152–1157 (2014).
20. Auclair, J. R. et al. Post-translational modification by cysteine protects Cu/Zn-superoxide dismutase from oxidative damage. *Biochemistry* **52**, 6137–6144 (2013).
21. Ayaz-Guner, S., Zhang, J., Li, L., Walker, J. W. & Ge, Y. In vivo phosphorylation site mapping in mouse cardiac troponin I by high resolution top-down electron capture dissociation mass spectrometry: Ser22/23 are the only sites basally phosphorylated. *Biochemistry* **48**, 8161–8170 (2009).
22. Martinez-Garcia, E. et al. The MMSET histone methyl transferase switches global histone methylation and alters gene expression in t(4;14) multiple myeloma cells. *Blood* **117**, 211–220 (2011).
23. Wu, D., Struwe, W. B., Harvey, D. J., Ferguson, M. A. J. & Robinson, C. V. N-glycan microheterogeneity regulates interactions of plasma proteins. *Proc. Natl Acad. Sci. USA* **115**, 8763 (2018).
24. Doll, S. & Burlingame, A. L. Mass spectrometry-based detection and assignment of protein posttranslational modifications. *ACS Chem. Biol.* **10**, 63–71 (2015).
25. Chait, B. T. Mass spectrometry: bottom-up or top-down? *Science* **314**, 65 (2006).
26. Wessel, D. & Flügge, U. I. A method for the quantitative recovery of protein in dilute solution in the presence of detergents and lipids. *Anal. Biochem.* **138**, 141–143 (1984).
27. Fornelli, L. et al. Top-down proteomics: where we are, where we are going? *J. Proteomics* **75**, 3–4 (2017).
28. Dang, X. et al. The first pilot project of the Consortium for Top-Down Proteomics: a status report. *Proteomics* **14**, 1130–1140 (2014).
29. Cech, N. B. & Enke, C. G. Practical implications of some recent studies in electrospray ionization fundamentals. *Mass Spectrom. Rev.* **20**, 362–387 (2001).
30. Metwally, H., McAllister, R. G. & Konermann, L. Exploring the mechanism of salt-induced signal suppression in protein electrospray mass spectrometry using experiments and molecular dynamics simulations. *Anal. Chem.* **87**, 2434–2442 (2015).
31. Compton, P. D., Zamdborg, L., Thomas, P. M. & Kelleher, N. L. On the scalability and requirements of whole protein mass spectrometry. *Anal. Chem.* **83**, 6868–6874 (2011).
32. Karas, M., Bahr, U. & Dülcks, T. Nano-electrospray ionization mass spectrometry: addressing analytical problems beyond routine. *Fresenius J. Anal. Chem.* **366**, 669–676 (2000).
33. Juraschek, R., Dulcks, T. & Karas, M. Nanoelectrospray—more than just a minimized-flow electrospray ionization source. *J. Am. Soc. Mass Spectrom.* **10**, 300–308 (1999).
34. Evershed, R. P., Robertson, D. H., Beynon, R. J. & Green, B. N. Application of electrospray ionization mass spectrometry with maximum-entropy analysis to allelic 'fingerprinting' of major urinary proteins. *Rapid Commun. Mass Spectrom.* **7**, 882–886 (1993).
35. Ferrige, A. G. et al. Disentangling electrospray spectra with maximum entropy. *Rapid Commun. Mass Spectrom.* **6**, 707–711 (1992).
36. Maquin, F., Schoot, B. M., Devaux, P. G. & Green, B. N. Molecular weight determination of recombinant interleukin 2 and interferon gamma by electrospray ionization mass spectroscopy. *Rapid Commun. Mass Spectrom.* **5**, 299–302 (1991).
37. Loo, J. A., Udseth, H. R. & Smith, R. D. Peptide and protein analysis by electrospray ionization-mass spectrometry and capillary electrophoresis-mass spectrometry. *Anal. Biochem.* **179**, 404–412 (1989).
38. Morris, H. R. et al. High sensitivity collisionally-activated decomposition tandem mass spectrometry on a novel quadrupole/orthogonal-acceleration time-of-flight mass spectrometer. *Rapid Commun. Mass Spectrom.* **10**, 889–896 (1996).

39. Kelleher, N. L., Senko, M. W., Little, D. P., O'Connor, P. B. & McLafferty, F. W. Complete large-molecule high-resolution mass spectra from 50-femtomole microvolume injection. *J. Am. Soc. Mass Spectrom.* **6**, 220–221 (1995).

40. LeDuc, R. D., Schwammle, V., Shortreed, M. R. & Cesnik, A. J. ProForma: a standard proteoform notation. *J. Proteome Res.* **17**, 1321–1325 (2018).

41. Shaw, K. L., Grimsley, G. R., Yakovlev, G. I., Makarov, A. A. & Pace, C. N. The effect of net charge on the solubility, activity, and stability of ribonuclease Sa. *Protein Sci.* **10**, 1206–1215 (2001).

42. Konermann, L. Addressing a common misconception: ammonium acetate as neutral pH "buffer" for native electrospray mass spectrometry. *J. Am. Soc. Mass Spectrom.* **28**, 1827–1835 (2017).

43. Uhlén, M. et al. Tissue-based map of the human proteome. *Science* **347**, 1260419 (2015).

44. Overington, J. P., Al-Lazikani, B. & Hopkins, A. L. How many drug targets are there? *Nat. Rev. Drug Discov.* **5**, 993 (2006).

45. Laganowsky, A., Reading, E., Hopper, J. T. S. & Robinson, C. V. Mass spectrometry of intact membrane protein complexes. *Nat. Protoc.* **8**, 639–651 (2013).

46. Wisniewski, J. R., Zougman, A., Nagaraj, N. & Mann, M. Universal sample preparation method for proteome analysis. *Nat. Methods* **6**, 359–362 (2009).

47. Erde, J., Loo, R. R. O. & Loo, J. A. Enhanced FASP (eFASP) to increase proteome coverage and sample recovery for quantitative proteomic experiments. *J. Proteome Res.* **13**, 1885–1895 (2014).

48. Doucette, A. A., Vieira, D. B., Orton, D. J. & Wall, M. J. Resolubilization of precipitated intact membrane proteins with cold formic acid for analysis by mass spectrometry. *J. Proteome Res.* **13**, 6001–6012 (2014).

49. Güray, M. Z., Zheng, S. & Doucette, A. A. Mass spectrometry of intact proteins reveals +98 u chemical artifacts following precipitation in acetone. *J. Proteome Res.* **16**, 889–897 (2017).

50. Smith, R. D., Loo, J. A., Edmonds, C. G., Barinaga, C. J. & Udseth, H. R. New developments in biochemical mass spectrometry: electrospray ionization. *Anal. Chem.* **62**, 882–899 (1990).

51. Leney, A. C. & Heck, A. J. R. Native mass spectrometry: what is in the name? *J. Am. Soc. Mass Spectrom.* **28**, 5–13 (2017).

52. Hernández, H. & Robinson, C. V. Determining the stoichiometry and interactions of macromolecular assemblies from mass spectrometry. *Nat. Protoc.* **2**, 715 (2007).

53. Lössl, P., Snijder, J. & Heck, A. J. R. Boundaries of mass resolution in native mass spectrometry. *J. Am. Soc. Mass Spectrom.* **25**, 906–917 (2014).

54. Gault, J. et al. High-resolution mass spectrometry of small molecules bound to membrane proteins. *Nat. Methods* **13**, 333–336 (2016).

55. Belov, A. M. et al. Analysis of proteins, protein complexes, and organellar proteomes using sheathless capillary zone electrophoresis - native mass spectrometry. *J. Am. Soc. Mass Spectrom.* **28**, 2614–2634 (2017).

56. Belov, A. M. et al. Complementary middle-down and intact monoclonal antibody proteoform characterization by capillary zone electrophoresis - mass spectrometry. *Electrophoresis* **39**, 2069–2082 (2018).

57. Tran, J. C. & Doucette, A. A. Multiplexed size separation of intact proteins in solution phase for mass spectrometry. *Anal. Chem.* **81**, 6201–6209 (2009).

58. Whitelegge, J. P., Gundersen, C. B. & Faull, K. F. Electrospray-ionization mass spectrometry of intact intrinsic membrane proteins. *Protein Sci.* **7**, 1423–1430 (1998).

59. le Coutre, J. et al. Proteomics on full-length membrane proteins using mass spectrometry. *Biochemistry* **39**, 4237–4242 (2000).

60. Whitelegge, J. P. HPLC and mass spectrometry of intrinsic membrane proteins. In *HPLC of Peptides and Proteins: Methods and Protocols* (ed. Aguilar, M.-I.) 323–339 (Springer, 2004).

61. Souda, P., Ryan, C. M., Cramer, W. A. & Whitelegge, J. Profiling of integral membrane proteins and their post translational modifications using high-resolution mass spectrometry. *Methods* **55**, 330–336 (2011).

62. Whitelegge, J. P. Integral membrane proteins and bilayer proteomics. *Anal. Chem.* **85**, 2558–2568 (2013).

63. Howery, A. E. et al. A designed inhibitor of a CLC antiporter blocks function through a unique binding mode. *Chem. Biol.* **19**, 1460–1470 (2012).

64. Whitelegge, J. P. et al. Toward the bilayer proteome, electrospray ionization-mass spectrometry of large, intact transmembrane proteins. *Proc. Natl Acad. Sci. USA* **96**, 10695–10698 (1999).

65. Cleland, T. P. et al. High-throughput analysis of intact human proteins using UVPD and HCD on an Orbitrap mass spectrometer. *J. Proteome Res.* **16**, 2072–2079 (2017).

66. Ansong, C. et al. Top-down proteomics reveals a unique protein S-thiolation switch in *Salmonella* Typhimurium in response to infection-like conditions. *Proc. Natl Acad. Sci. USA* **110**, 10153–10158 (2013).

67. Ntai, I. et al. Applying label-free quantitation to top down proteomics. *Anal. Chem.* **86**, 4961–4968 (2014).

68. Skinner, O. S. et al. Top-down characterization of endogenous protein complexes with native proteomics. *Nat. Chem. Biol.* **14**, 36 (2017).

## Author contributions

D.P.D. and C.M.R. contributed by making all the figures and writing the manuscript with J.N.A. Experiments were performed by D.P.D., C.M.R., C.J.D., L.F., L.F.S., Z.L., J.J.W., J.L.L., K.C.A, R.S., K.J., A.K., I.D.G.C., J.R.A., B.C., C. L., W.J., A.R.I. and J.P.W. who all contributed data that was incorporated into the final results. L.P., J.C., P.O.D., L.M.S., Y.O.T., J.A.L, Y.G., N.L.K. and J.N.A make up the Consortium for Top-Down Proteomics Board of Directors who designed and guided this project. All authors contributed to the editing and formatting of the manuscript.

## Competing interests

Authors with company affiliations include J.L.L. (Amgen), K.C.A. (Alnylam), R.S. (Biogen), J.J.W. (Bruker), I.D.G.C. (Amgen), P.O.D. (Eastwoods Consulting) and Y.O.T. (Spectroswiss).

## Additional information

**Supplementary information** is available for this paper at https://doi.org/10.1038/s41592-019-0457-0.

**Reprints and permissions information** is available at www.nature.com/reprints.

**Correspondence** should be addressed to J.N.A.

**Peer review information:** Allison Doerr was the primary editor on this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

# Internal Fragments Generated by Electron Ionization Dissociation Enhance Protein Top-Down Mass Spectrometry

Muhammad A. Zenaidee, Carter Lantz, Taylor Perkins, Wonhyuek Jung, Rachel R. Ogorzalek Loo, and Joseph A. Loo*

ACCESS | Metrics & More | Article Recommendations | Supporting Information

**ABSTRACT:** Top-down proteomics by mass spectrometry (MS) involves the mass measurement of an intact protein followed by subsequent activation of the protein to generate product ions. Electron-based fragmentation methods like electron capture dissociation and electron transfer dissociation are widely used for these types of analyses. Recently, electron ionization dissociation (EID), which utilizes higher energy electrons (>20 eV) has been suggested to be more efficient for top-down protein fragmentation compared to other electron-based dissociation methods. Here, we demonstrate that the use of EID enhances protein fragmentation and subsequent detection of protein fragments. Protein product ions can form by either single cleavage events, resulting in terminal fragments containing the C-terminus or N-terminus of the protein, or by multiple cleavage events to give rise to internal fragments that include neither the C-terminus nor the N-terminus of the protein. Conventionally, internal fragments have been disregarded, as reliable assignments of these fragments were limited. Here, we demonstrate that internal fragments generated by EID can account for ∼20−40% of the mass spectral signals detected by top-down EID-MS experiments. By including internal fragments, the extent of the protein sequence that can be explained from a single tandem mass spectrum increases from ∼50 to ∼99% for 29 kDa carbonic anhydrase II and 8.6 kDa ubiquitin. When searching for internal fragments during data analysis, previously unassigned peaks can be readily and accurately assigned to confirm a given protein sequence and to enhance the utility of top-down protein sequencing experiments.
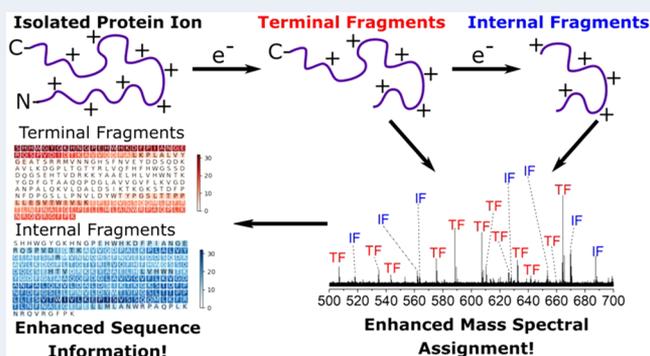
## INTRODUCTION

Top-down mass spectrometry (MS) has emerged as a technique to characterize proteins and to elucidate unique proteoforms.[1,2] Typically, intact protein ions are generated using electrospray ionization (ESI), followed by dissociation of the intact protein ion within the mass spectrometer to generate product ions that can be used to return information about protein identification and primary structure, i.e., sequence. Electron-based dissociation techniques such as electron capture dissociation (ECD)[3,4] and electron transfer dissociation (ETD)[5] employ low-energy electrons to generate protein fragment ions. ECD/ETD confers many advantages over other dissociation techniques (e.g., collision induced dissociation (CID),[6] surface induced dissociation (SID),[7] and ultraviolet photodissociation (UVPD)[8]), including but not limited to conserving post-translational modifications and nonselective fragmentation of the protein backbone.[9,10] Due to nonspecific cleavage during electron-based dissociation, fragmentation by electron-based methods has the potential to generate more protein fragments that allow for richer sequence information.[11,12] Despite the many advantages and prevalent use of ECD/ETD (ExD),[13] these fragmentation techniques can be limiting due to the reliance of generating protein ions in higher charge states.[14] In addition, proteins can have low electron capture efficiencies, thus potentially limiting ExD efficiency.[3,15]

Electron ionization dissociation (EID) is a recently discovered alternative ExD fragmentation technique for peptide and protein characterization.[16−18] EID utilizes high-energy electrons (>20 eV) to induce protein fragmentation along the backbone. In EID, the interaction of a multiply charged protein ion with a high-energy electron results in the formation of the oxidized species. Subsequent rearrangement of the oxidized species and/or capture of a second electron promotes backbone fragmentation. Using EID, Zubarev and co-workers demonstrated that fragmentation efficiency for some proteins and peptides can be close to 100%.[18] This data suggests that EID can result in more efficient fragmentation of

polypeptides compared to ECD, which would be especially beneficial for the analysis of large proteins. Recently, Loo and co-workers demonstrated that by using EID for native top-down MS, extensive fragmentation of apo-human superoxide dismutase 1 homodimer complex (32 kDa) could be achieved, whereas ECD resulted only in charge-reduced precursors and no protein fragmentation.[17] In addition, EID can be used to probe the metal-binding sites of proteins and protein complexes, which suggests that EID could be beneficial for investigating the binding of noncovalent ligands and labile PTMs.[17] Despite the potential advantages conferred by EID compared to ECD/ETD, the use of EID for protein top-down MS has not been extensively explored.

Protein product ions can either be (i) a terminal fragment ion, where only a single cleavage event occurs to generate N-terminal-containing $a$, $b$, and $c$ fragments or C-terminal-containing $x$, $y$, and $z$ fragments[19] or (ii) an internal fragment ion, where two cleavage events occur generating $ax$, $ay$, $az$, $bx$, $by$, $bz$, $cx$, $cy$, and $cz$ fragment ions, depending on the activation method occurring.[20−23] The number of theoretical internal products that can be generated is significantly greater than the number of possible terminal fragments that can be generated.[20] Traditionally, internal fragments have been largely ignored due to the inability to reliably assign internal fragments.[24] Due to this, a plethora of information that can be accessed has largely been ignored.

The analysis of internal fragments previously has been limited to peptides and small molecules.[25,26] Assignment of internal fragments for intact proteins has been relatively limiting, owing to the complexity of the fragmentation spectra. Kelleher and co-workers showed that internal fragments from CID fragmentation of the common test protein ubiquitin (8.6 kDa) can be assigned to result in significantly greater protein sequence coverage.[20] Similarly, for other intact proteins, the inclusion of internal fragments that can be generated by CID could result in greater explanation of the protein sequence.[27,28] Our laboratory demonstrated that internal product ions can be generated from top-down MS of large, native protein complexes.[29] These examples suggest that the inclusion of internal fragments in top-down protein sequencing experiments could significantly enhance the protein sequence coverage and the efficiency of top-down mass spectrometry experiments.

Here, we investigate the utility of EID and the inclusion of internal fragments for top-down protein sequencing experiments. By using EID for top-down MS of ubiquitin (Ubq) and 29 kDa carbonic anhydrase II (CAII), the number of product ions is significantly higher compared to ECD, and approximately 20−40% of the fragments in the mass spectra can be assigned to internal fragments. The inclusion of EID-generated internal fragment ions yields nearly complete sequence coverage for CAII.

## EXPERIMENTAL SECTION

**Materials.** Bovine carbonic anhydrase II and bovine ubiquitin were purchased from Sigma-Aldrich (St. Louis, MO, USA) and used without further purification. LC/MS grade water and methanol were obtained from Fisher Chemical (Hampton, NH, USA). For electrospray ionization, aqueous solutions containing 10−20 $\mu$M protein, 49.5% water, 49.5% methanol, and 1% formic acid ($v/v$) were prepared.

**Mass Spectrometry.** All experiments were conducted on a 15-T Bruker SolariX Fourier transform ion cyclotron

resonance (FTICR)-MS instrument equipped with an infinity ICR cell (Bruker Daltonics, Billerica, MA, USA). Protein solutions were loaded into metal-coated borosilicate capillaries (Au/Pd-coated, 1 $\mu$M inner diameter, Thermo Fisher Scientific, Waltham, MA, USA), and electrospray was initiated by applying a voltage between 0.9 and 1.4 kV on the ESI capillary. Charge states were isolated in the quadrupole, with an isolation window of 10 $m/z$, before EID/ECD fragmentation. For ECD fragmentation, the pulse length was set at 0.05 s, with a lens voltage of 50 V and an ECD bias voltage of 2 V. For EID fragmentation, the pulse length and lens voltage were kept constant, and the bias voltage was altered between 20 and 30 V. For each spectrum, 200 scans were obtained.

**Data Analysis.** *Peak Assignments.* Deconvoluted mass lists were obtained with Bruker Data Analysis software, using the SNAP algorithm. Deconvoluted mass lists were uploaded into our in-house-developed python program that calculates all possible terminal and internal fragments and compares them to the experimental deconvoluted masses. A future report will describe the details of the program and the user interface. The error for matching was set at an error of 1 ppm, and 42.0105603 for the mass of acetylation was added to all the N-terminal fragments for carbonic anhydrase II. Internal fragments searched and assigned were only for $cz$ internal fragments.

*Protein Sequence Confirmation.* The elucidated protein sequences were calculated by eq 1:

$$\text{sequence confirmation (\%)} = (\text{AA}_{\text{det}}/\text{AA}_{\text{tot}}) \times 100 \qquad (1)$$

where $\text{AA}_{\text{det}}$ is the number of times an amino acid residue was detected. For internal fragments, amino acids were detected in at least 5 different fragments to ensure an accurate detection similar to previous thresholds of Kelleher and co-workers.[20] $\text{AA}_{\text{tot}}$ is the total number of amino acids in the protein. The sequence elucidated should give an indication of how much of the protein sequence can be defined by the fragments assigned.

*Protein Fragment Coverage.* Protein fragment coverages were calculated by identifying the number of observed inter-residue sites divided by the total number of inter-residue cleavages on the protein backbone. For example, Ubq and CAII have 75 and 259 inter-residue cleavage sites, respectively.

## RESULTS AND DISCUSSION

EID for protein tandem-MS experiments can result in greater fragmentation compared to ECD (Figure 1). The EID (25 eV) of $[\text{CAII, 25H}]^{25+}$ (i.e., $[\text{M+25H}]^{25+}$) resulted in the formation of the $[\text{CAII, 25H}]^{26+*}$ ion with a measurable abundance (Figure 1b), which is in good agreement with Zubarev and co-workers' observation that EID results in the formation of the oxidized species.[30] Representative mass spectral signals for some product ions identifiable within the $m/z$ 500−700 range are shown within the insets of Figure 1.

The product ions formed by EID have higher signal-to-noise ratios (S/N) compared to the fragments formed by ECD. For example, for the $c_{12}^{3+}$ ion detected in both the ECD and EID spectra, the fragment within the EID spectrum has ~30% higher S/N compared to the fragment detected in the ECD spectrum. On average, there is a ~30−40% increase in S/N of fragments formed by EID compared to ECD, suggesting that EID is a more efficient fragmentation method.

The number of fragment ions that are generated and detected by EID is also greater than that generated by ECD. For example, EID of $[\text{CAII, 25H}]^{25+}$ resulted in the formation
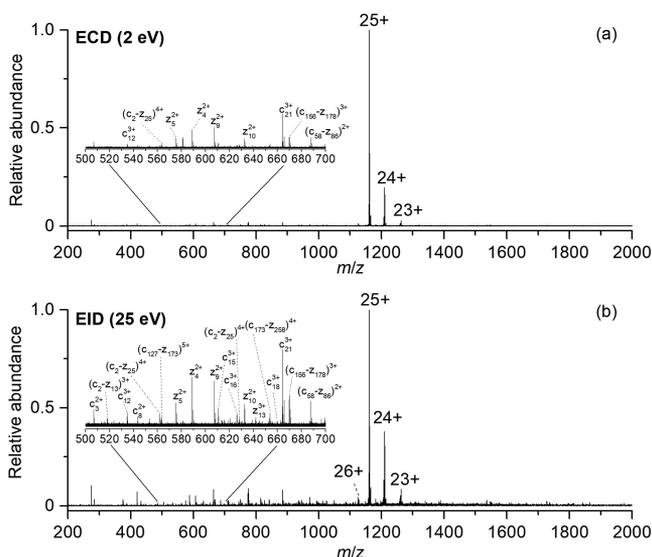
**Figure 1.** Representative top-down mass spectra of isolated $[CAII, 25H]^{25+}$ formed from 10 $\mu$M carbonic anhydrase II in 50:50 water:methanol and 1% formic acid using (a) ECD (2 eV) and (b) EID (25 eV).



**Figure 2.** Effect of ECD/EID energy for $[CAII, 25H]^{25+}$ for (a) the number of fragments deconvoluted and (b) the total number of fragments identified (open triangles), the total number of terminal fragments identified (closed circles), and the total number of internal fragments identified (open squares). (c) Fraction of fragments matched, where the percentage of all fragments identified (open triangles), percentage of terminal fragments identified (closed circles), and the percentage of internal fragments identified (open squares) are shown.

---

of 145 unique fragments, whereas ECD of $[CAII, 25H]^{25+}$ resulted in the formation of 100 unique fragments (Figure 2a). Similarly, for $[Ubq, 10H]^{10+}$, more fragments are formed by EID (159 unique fragments) compared to ECD (102 unique fragments) (Figure S1). Previous data from our lab demonstrated that EID generated more fragments than ECD for native SOD-1 proteins.[17]

Product ions generated for proteins can be classified as either terminal fragments (fragments containing the N-terminus or C-terminus) or internal fragments (fragments containing neither the N-terminus nor C-terminus). The data shown here and in other reports suggest that *internal* fragments can account for many of the mass spectral signals within a mass spectrum regardless of the fragmentation method.[31] Interestingly, the ion abundances of internal fragments are not significantly lower compared to the low-abundance terminal fragments formed. For example, the fragment assigned to the internal fragment $(c_{58}-z_{86})^{2+}$ is similar in ion abundance to low-abundance $c$- and $z$-fragment ions detected within the spectrum. On average, the total intensity of terminal fragments is ~$1.11 \times 10^7$ compared to ~$6.74 \times 10^6$ for internal fragments (Table S1). This suggests that these fragments are identifiable and could be assigned. Therefore, the inclusion of these internal fragments could result in greater sequence information for the protein compared to terminal fragments alone.

Conventionally, internal fragments have been ignored, owing to the inability to reliably assign them due to instrument and computational limitations.[24,32] However, many ion signals within tandem mass spectra can potentially be explained by internal fragments (Figure S2b). Here, we attempt to assign previously unassigned mass spectral signals by calculating internal fragment masses using an in-house written program (see Experimental Section); $cz$ internal fragments can be uniquely assigned to the known protein sequence if MS2 mass calibration is achieved to ≤1 ppm error to limit false positives. For the ECD of $[CAII, 25H]^{25+}$, there were 100 unique fragments formed, of which 80 (80%) are assigned to terminal fragments and 20 (20%) are assigned to internal fragments. For EID, there were 145 unique fragments formed for $[CAII,$

$25H]^{25+}$: 113 terminal (78%) and 32 internal fragments (22%). The number of internal fragment ions that are formed for CAII and Ubq can account for 20−40% of the ion signals within a mass spectrum (Figure 2 and Figure S1).

The effect of the electron energy on the ECD/EID efficiency of $[CAII, 25H]^{25+}$ was investigated (Figure 2). At a conventional ECD energy (2 eV), 101 unique fragments were identified. However, at greater than 20 eV of EID energies, there are more fragments generated (>125 unique fragments), with an energy range between 20 and 26 eV being optimal for generating the largest number of fragments (Figure 2a). Interestingly, as electron energy increases, the number of internal fragments formed also increases (Figure 2b); the number of internal fragments formed at 24 eV was 33 compared to only 20 at 2 eV. However, the percentage of fragments that are assigned as internal fragments does not increase significantly at 24 eV compared to 2 eV (18 to 24%, respectively). Similarly, for ubiquitin, the percentage of fragments that can be explained by internal fragments formed at ECD energies is similar to the percentage of internal fragments formed at EID energies (Figure S1). The formation of internal fragments from EID of Ubq is in good agreement with previous data from Kelleher and co-workers, who

demonstrated that the CID of Ubq results in the formation of internal fragments, with approximately 30% of the mass spectral signals attributed to internal fragments.[20] As internal fragments are generated within conventional dissociation experiments (e.g., CID, UVPD, ECD/ETD), the assignment of internal fragments should be beneficial for characterizing the protein sequence because more of the mass spectral signals can be assigned to protein fragments.

The efficiency of ECD fragmentation has been shown to correlate with the charge state of the precursor ion.[14,33−35] Similarly, for EID, the fragmentation efficiency increases with the charge (Figure 3a). For CAII, as the precursor charge
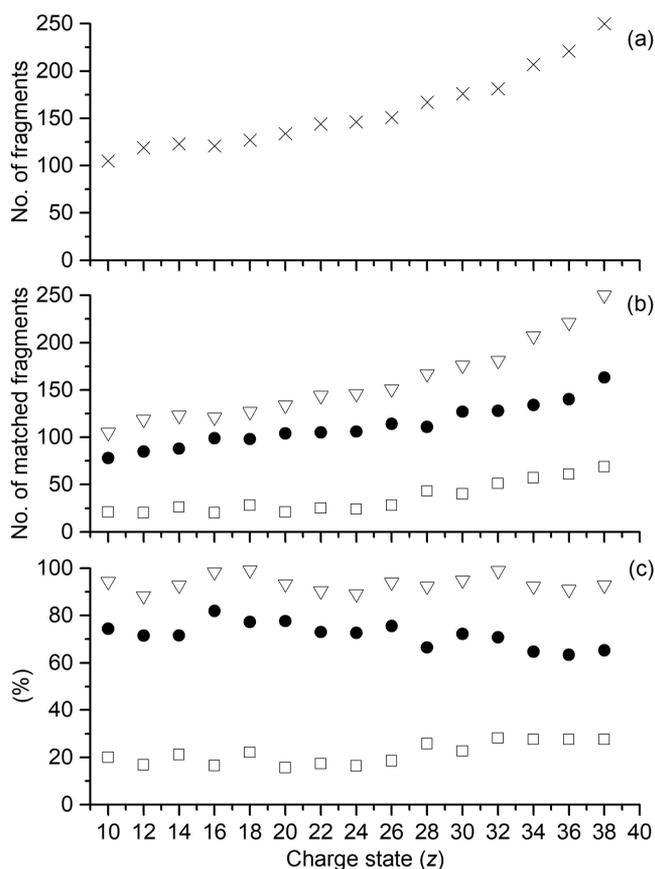


**Figure 3.** Performance of EID-MS for isolated [CAII, $z$H]$^{z+}$ ($z$ = 10+ to 38+). (a) Number of fragments deconvoluted. (b) Total number of fragments identified (open triangles), total number of terminal fragments identified (closed circles), and total number of internal fragments identified (open squares). (c) Percentage of the fragments matched, where the percentage of all fragments identified (open triangles), percentage of terminal fragments identified (closed circles), and percentage of internal fragments identified (open squares) are depicted.

increases from 10+ to 38+, the number of fragments increases from 105 to 250, respectively. A similar trend is observed for Ubq; the number of fragments increases from 69 (7+ precursor ion charge) to 175 (13+) (Figure S1). EID's greater fragmentation efficiency for higher charge states could be due to lowering the barriers for dissociation and larger reaction cross sections.

The theoretical total number of fragments, including internal fragments, larger than 4 residues scales with a second-order polynomial with increasing number of amino acids (Figure

S2). For example, the number of fragments that can be generated theoretically by fragmentation between each amino acid residue, based on cleavages of the C$\alpha$-carbonyl-C, N-C$\alpha$, and N-carbonyl-C bonds, ranges from 4 amino acid residues to one less than the entire protein length. For Ubq (76 amino acid residues), the total is 25429 compared to 299929 for CAII (259 amino acid residues), with the vast majority originating from internal fragments; for Ubq and CAII, the number of theoretical deisotoped internal fragment masses are 24975 and 298377, respectively. This is in close agreement with previous calculations by Kelleher and co-workers, who demonstrated that the number of internal fragments that can theoretically be generated outnumber the number of terminal fragments generated.[20] In addition, the formation of internal fragments should be especially beneficial for larger proteins because internal fragments are generally lighter than terminal fragments (Figure S2). For example, the average mass of a terminal fragment for CAII is 14.4 kDa, whereas the average mass of an internal fragment is 9.5 kDa (Figure S2b). This trend becomes significantly more important for larger proteins; the difference between the average mass of terminal fragments and internal fragments increases as the protein size increases. By generating internal fragments that have smaller masses than terminal fragments, these internal fragments should fall within the mass range of most mass spectrometers, which should allow for greater confirmation of these amino acid sequence regions within the protein. A top-down MS analysis suffers from an upper mass limit;[36−38] the inclusion of internal fragments could potentially help to overcome this limitation.

Internal fragment ion masses can be readily calculated and can be assigned to a target protein sequence if MS2 mass calibration is achieved to be ≤1 ppm error to ensure precise assignments and limit false positive assignments. As an example, to test the validity of these assignments, the theoretical fragment isotopic distribution was fitted over the corresponding fragment peak in the mass spectrum (Figure S3).[39,40] For both terminal and internal fragments, the theoretical isotopic distributions of the corresponding fragment are in good agreement with the peaks observed in the mass spectrum (Figure S3).

To estimate the false discovery rate of the internal fragment assignments, internal fragment masses of [CAII, 25H]$^{25+}$ were shifted with a given ppm error ranging −500 to 500 to produce null data sets. These null data sets were then searched against the CAII sequence to estimate the amount of random matching to other internal fragments.[20] The percentages of internal fragments that were assigned to a different internal fragment for the null data set are shown in Figure S4. When all possible internal fragments were considered (*ax, ay, az, bx, by, bz, cx, cy,* and *cz*), ∼ 20% of the internal fragment null sets was assigned to a different internal fragment. However, considering that ExD fragmentation was utilized, internal fragments formed should be *cz* fragments that stem from multiple fragmentation events. When only *cz* internal fragments were considered for the null data sets, only 3.4% of the internal fragments was assigned to other *cz* internal fragments from the CAII sequence, which indicates a low false discovery rate.

By including both internal fragments and terminal fragments, a larger fraction of the protein sequence can be explained (Figure 4). Heatmaps of the number of times a residue is represented by a product ion are plotted to give an indication of hotspots within the polypeptide sequence corresponding to where fragments are formed and detected.
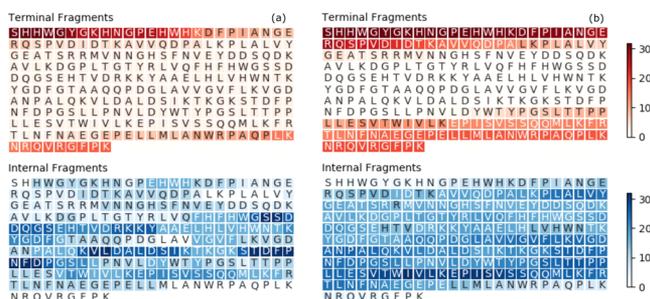
**Figure 4.** Heatmap depicting the number of times each residue is covered by a terminal fragment (top) or an internal fragment (bottom) for the (a) ECD of $[CAII, 25H]^{25+}$ and (b) EID of $[CAII, 25H]^{25+}$. Darker colors indicate greater coverage.

For ECD of CAII (Figure 4a), terminal fragments only account for a small fraction of the protein sequence (28%). Interestingly, the inclusion of internal fragments yields nearly complete sequence confirmation of carbonic anhydrase II (~90%). For EID, a larger percentage of the protein sequence can be explained by terminal fragments (47%), and similarly, the inclusion of internal fragments resulted in near-complete protein sequence confirmation (~99%). The inclusion of internal fragments was also beneficial for the percentage of the inter-residue cleavages that was observed. For ECD, inclusion of the internal fragments showed that 200 of the 259 inter-residue cleavage sites were fragmented, and for EID, 234 of the 259 inter-residue cleavage sites were fragmented, indicating that a majority of the protein backbone was cleaved to form fragment ions (Figure S5). For Ubq, similar trends are observed, with many internal fragments assigned to fragments that contain the center of the protein backbone (Figure S6). Furthermore, for Ubq, the majority of inter-residue cleavage sites was cleaved and assigned to protein fragments (Figure S7).

## CONCLUSIONS

EID for top-down MS can significantly enhance the efficiency of protein fragmentation. From this study, EID outperforms ECD with larger numbers of fragments generated as well as higher fragment ion abundances. Most significantly, the use of internal fragment assignments resulted in the confirmation of a larger fraction of a given protein sequence. Because of some ambiguity in assigning internal fragments due to the large number of theoretical internal products, terminal fragments are ideal for protein identification and internal fragments are useful for sequence confirmation. Future work will define the types of internal fragments generated by ExD and other activation/dissociation methods and the limits of assigning internal fragments for larger proteins beyond 30 kDa. As the internal fragments formed contain amino acid sequences that are complementary to the terminal fragments, these internal fragments should be useful for localizing post-translational modifications[41,42] and protein−ligand binding sites[43,44] and for characterizing large protein complexes[45] and membrane proteins[46,47] with native top-down MS. In general, incorporating the previously unassigned internal product ions generated by all activation/dissociation techniques, especially with EID, should greatly enhance the utility of top-down MS for protein sequence analysis to larger proteins.

## ASSOCIATED CONTENT

### Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/jasms.0c00160.

Internal fragments generated by electron ionization dissociation enhances protein top-down mass spectrometry (PDF)

## AUTHOR INFORMATION

### Corresponding Author

**Joseph A. Loo** − *Department of Chemistry and Biochemistry and Department of Biological Chemistry, University of California, Los Angeles, Los Angeles, California 90095, United States;* orcid.org/0000-0001-9989-1437; Email: jloo@chem.ucla.edu

### Authors

**Muhammad A. Zenaidee** − *Department of Chemistry and Biochemistry, University of California, Los Angeles, Los Angeles, California 90095, United States;* orcid.org/0000-0003-3188-1472

**Carter Lantz** − *Department of Chemistry and Biochemistry, University of California, Los Angeles, Los Angeles, California 90095, United States*

**Taylor Perkins** − *Department of Chemistry and Biochemistry, University of California, Los Angeles, Los Angeles, California 90095, United States*

**Wonhyuek Jung** − *Department of Chemistry and Biochemistry, University of California, Los Angeles, Los Angeles, California 90095, United States*

**Rachel R. Ogorzalek Loo** − *Department of Biological Chemistry, University of California, Los Angeles, Los Angeles, California 90095, United States*

Complete contact information is available at:
https://pubs.acs.org/10.1021/jasms.0c00160

### Notes

The authors declare no competing financial interest.

## REFERENCES

(1) Kelleher, N. L.; Lin, H. Y.; Valaskovic, G. A.; Aaserud, D. J.; Fridriksson, E. K.; McLafferty, F. W. Top Down versus Bottom Up Protein Characterization by Tandem High-Resolution Mass Spectrometry. *J. Am. Chem. Soc.* **1999**, *121*, 806−812.

(2) Lermyte, F.; Tsybin, Y. O.; O'Connor, P. B.; Loo, J. A. Top or Middle? Up or Down? Toward a Standard Lexicon for Protein Top-Down and Allied Mass Spectrometry Approaches. *J. Am. Soc. Mass Spectrom.* **2019**, *30*, 1149−1157.

(3) Zubarev, R. A.; Horn, D. M.; Fridriksson, E. K.; Kelleher, N. L.; Kruger, N. A.; Lewis, M. A.; Carpenter, B. K.; McLafferty, F. W.

Electron Capture Dissociation for Structural Characterization of Multiply Charged Protein Cations. *Anal. Chem.* 2000, 72, 563−573.

(4) Zubarev, R. A.; Kelleher, N. L.; McLafferty, F. W. Electron capture dissociation of multiply charged protein cations. A nonergodic process. *J. Am. Chem. Soc.* 1998, 120, 3265−3266.

(5) Good, D. M.; Wirtala, M.; McAlister, G. C.; Coon, J. J. Performance Characteristics of Electron Transfer Dissociation Mass Spectrometry. *Mol. Cell. Proteomics* 2007, 6, 1942−1951.

(6) Katta, V.; Chowdhury, S. K.; Chait, B. T. Use of a single-quadrupole mass spectrometer for collision-induced dissociation studies of multiply charged peptide ions produced by electrospray ionization. *Anal. Chem.* 1991, 63, 174−178.

(7) McCormack, A. L.; Jones, J. L.; Wysocki, V. H. Surface-Induced Dissociation of Multiply Protonated Peptides. *J. Am. Soc. Mass Spectrom.* 1992, 3, 859−862.

(8) Brodbelt, J. S. Photodissociation mass spectrometry: new tools for characterization of biological molecules. *Chem. Soc. Rev.* 2014, 43, 2757−2783.

(9) Brodbelt, J. S. Ion Activation Methods for Peptides and Proteins. *Anal. Chem.* 2016, 88, 30−51.

(10) Zhurov, K. O.; Fornelli, L.; Wodrich, M. D.; Laskay, U. A.; Tsybin, Y. O. Principles of electron capture and transfer dissociation mass spectrometry applied to peptide and protein structure analysis. *Chem. Soc. Rev.* 2013, 42, 5014−5030.

(11) Syrstad, E. A.; Tureček, F. Toward a general mechanism of electron capture dissociation. *J. Am. Soc. Mass Spectrom.* 2005, 16, 208−224.

(12) Zubarev, R. A.; Haselmann, K. F.; Budnik, B.; Kjeldsen, F.; Jensen, F. Towards an understanding of the mechanism of electron-capture dissociation: a historical perspective and modern ideas. *Eur. J. Mass Spectrom.* 2002, 8, 337−350.

(13) Lermyte, F.; Valkenborg, D.; Loo, J. A.; Sobott, F. Radical solutions: Principles and application of electron-based dissociation in mass spectrometry-based analysis of protein structure. *Mass Spectrom. Rev.* 2018, 37, 750−771.

(14) Iavarone, A. T.; Paech, K.; Williams, E. R. Effects of Charge State and Cationizing Agent on the Electron Capture Dissociation of a Peptide. *Anal. Chem.* 2004, 76, 2231−2238.

(15) Zubarev, R. A. Electron-capture dissociation tandem mass spectrometry. *Curr. Opin. Biotechnol.* 2004, 15, 12−16.

(16) Kaczorowska, M. A.; Cooper, H. J. Electron induced dissociation (EID) tandem mass spectrometry of octaethylporphyrin and its iron(iii) complex. *Chem. Commun.* 2011, 47, 418−420.

(17) Li, H.; Sheng, Y.; McGee, W.; Cammarata, M.; Holden, D.; Loo, J. A. Structural Characterization of Native Proteins and Protein Complexes by Electron Ionization Dissociation-Mass Spectrometry. *Anal. Chem.* 2017, 89, 2731−2738.

(18) Zubarev, R. A.; Yang, H. Multiple Soft Ionization of Gas-Phase Proteins and Swift Backbone Dissociation in Collisions with ≤ 99 eV Electrons. *Angew. Chem., Int. Ed.* 2010, 49, 1439−1441.

(19) Zubarev, R. Protein primary structure using orthogonal fragmentation techniques in Fourier transform mass spectrometry. *Expert Rev. Proteomics* 2006, 3, 251−261.

(20) Durbin, K. R.; Skinner, O. S.; Fellers, R. T.; Kelleher, N. L. Analyzing internal fragmentation of electrosprayed ubiquitin ions during beam-type collisional dissociation. *J. Am. Soc. Mass Spectrom.* 2015, 26, 782−787.

(21) Holden, D. D.; Brodbelt, J. S. Improving Performance Metrics of Ultraviolet Photodissociation Mass Spectrometry by Selective Precursor Ejection. *Anal. Chem.* 2017, 89, 837−846.

(22) Michalski, A.; Neuhauser, N.; Cox, J.; Mann, M. A Systematic Investigation into the Nature of Tryptic HCD Spectra. *J. Proteome Res.* 2012, 11, 5479−5491.

(23) Zinnel, N. F.; Pai, P. J.; Russell, D. H. Ion mobility-mass spectrometry (IM-MS) for top-down proteomics: increased dynamic range affords increased sequence coverage. *Anal. Chem.* 2012, 84, 3390−3397.

(24) Xu, C.; Ma, B. Complexity and scoring function of MS/MS peptide de novo sequencing. *Computational Systems Bioinformatics Conference Proceedings* 2006, 361−369.

(25) Ballard, K. D.; Gaskell, S. J. Sequential mass spectrometry applied to the study of the formation of "internal" fragment ions of protonated peptides. *Int. J. Mass Spectrom. Ion Processes* 1991, 111, 173−189.

(26) Barran, P. E.; Polfer, N. C.; Campopiano, D. J.; Clarke, D. J.; Langridge-Smith, P. R. R.; Langley, R. J.; Govan, J. R. W.; Maxwell, A.; Dorin, J. R.; Millar, R. P.; Bowers, M. T. Is it biologically relevant to measure the structures of small peptides in the gas-phase? *Int. J. Mass Spectrom.* 2005, 240, 273−284.

(27) Chen, J.; Shiyanov, P.; Green, K. B. Top-down mass spectrometry of intact phosphorylated β-casein: Correlation between the precursor charge state and internal fragments. *J. Mass Spectrom.* 2019, 54, 527−539.

(28) Cobb, J. S.; Easterling, M. L.; Agar, J. N. Structural characterization of intact proteins is enhanced by prevalent fragmentation pathways rarely observed for peptides. *J. Am. Soc. Mass Spectrom.* 2010, 21, 949−959.

(29) Li, H.; Nguyen, H. H.; Ogorzalek Loo, R. R.; Campuzano, I. D. G.; Loo, J. A. An integrated native mass spectrometry and top-down proteomics method that connects sequence to structure and function of macromolecular complexes. *Nat. Chem.* 2018, 10, 139−148.

(30) Fung, Y. M. E.; Adams, C. M.; Zubarev, R. A. Electron Ionization Dissociation of Singly and Multiply Charged Peptides. *J. Am. Chem. Soc.* 2009, 131, 9977−9985.

(31) Lyon, Y. A.; Riggs, D.; Fornelli, L.; Compton, P. D.; Julian, R. R. The Ups and Downs of Repeated Cleavage and Internal Fragment Production in Top-Down Proteomics. *J. Am. Soc. Mass Spectrom.* 2018, 29, 150−157.

(32) Zamdborg, L.; LeDuc, R. D.; Glowacz, K. J.; Kim, Y.-B.; Viswanathan, V.; Spaulding, I. T.; Early, B. P.; Bluhm, E. J.; Babai, S.; Kelleher, N. L. ProSight PTM 2.0: improved protein identification and characterization for top down mass spectrometry. *Nucleic Acids Res.* 2007, 35, W701−W706.

(33) Iavarone, A. T.; Jurchen, J. C.; Williams, E. R. Effects of Solvent on the Maximum Charge State and Charge State Distribution of Protein Ions Produced by Electrospray Ionization. *J. Am. Soc. Mass Spectrom.* 2000, 11, 976−985.

(34) Zenaidee, M. A.; Donald, W. A. Electron capture dissociation of extremely supercharged protein ions formed by electrospray ionisation. *Anal. Methods* 2015, 7, 7132−7139.

(35) Zenaidee, M. A.; Donald, W. A. Extremely supercharged proteins in mass spectrometry: profiling the pH of electrospray generated droplets, narrowing charge state distributions, and increasing ion fragmentation. *Analyst* 2015, 140, 1894−1905.

(36) Catherman, A. D.; Skinner, O. S.; Kelleher, N. L. Top Down proteomics: facts and perspectives. *Biochem. Biophys. Res. Commun.* 2014, 445, 683−693.

(37) Reid, G. E.; McLuckey, S. A. 'Top down' protein characterization via tandem mass spectrometry. *J. Mass Spectrom.* 2002, 37, 663−675.

(38) Riley, N. M.; Westphall, M. S.; Coon, J. J. Sequencing Larger Intact Proteins (30−70 kDa) with Activated Ion Electron Transfer Dissociation. *J. Am. Soc. Mass Spectrom.* 2018, 29, 140−149.

(39) Kaur, P.; O'Connor, P. B. Algorithms for Automatic Interpretation of High Resolution Mass Spectra. *J. Am. Soc. Mass Spectrom.* 2006, 17, 459−468.

(40) Beu, S. C.; Senko, M. W.; Quinn, J. P.; Wampler, F. M.; McLafferty, F. W. Fourier-transform electrospray instrumentation for tandem high-resolution mass spectrometry of large molecules. *J. Am. Soc. Mass Spectrom.* 1993, 4, 557−565.

(41) Lin, Z.; Guo, F.; Gregorich, Z. R.; Sun, R.; Zhang, H.; Hu, Y.; Shanmuganayagam, D.; Ge, Y. Comprehensive Characterization of Swine Cardiac Troponin T Proteoforms by Top-Down Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* 2018, 29, 1284−1294.

(42) Wu, Z.; Jin, Y.; Chen, B.; Gugger, M. K.; Wilkinson-Johnson, C. L.; Tiambeng, T. N.; Jin, S.; Ge, Y. Comprehensive Characterization

of the Recombinant Catalytic Subunit of cAMP-Dependent Protein Kinase by Top-Down Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **2019**, *30*, 2561−2570.

(43) Nshanian, M.; Lantz, C.; Wongkongkathep, P.; Schrader, T.; Klärner, F.-G.; Blümke, A.; Despres, C.; Ehrmann, M.; Smet-Nocca, C.; Bitan, G.; Loo, J. A. Native Top-Down Mass Spectrometry and Ion Mobility Spectrometry of the Interaction of Tau Protein with a Molecular Tweezer Assembly Modulator. *J. Am. Soc. Mass Spectrom.* **2019**, *30*, 16−23.

(44) Wongkongkathep, P.; Han, J. Y.; Choi, T. S.; Yin, S.; Kim, H. I.; Loo, J. A. Native Top-Down Mass Spectrometry and Ion Mobility MS for Characterizing the Cobalt and Manganese Metal Binding of α-Synuclein Protein. *J. Am. Soc. Mass Spectrom.* **2018**, *29*, 1870−1880.

(45) Zhou, M.; Yan, J.; Romano, C. A.; Tebo, B. M.; Wysocki, V. H.; Pasa-Tolic, L. Surface Induced Dissociation Coupled with High Resolution Mass Spectrometry Unveils Heterogeneity of a 211 kDa Multicopper Oxidase Protein Complex. *J. Am. Soc. Mass Spectrom.* **2018**, *29*, 723−733.

(46) Lippens, J. L.; Nshanian, M.; Spahr, C.; Egea, P. F.; Loo, J. A.; Campuzano, I. D. G. Fourier Transform-Ion Cyclotron Resonance Mass Spectrometry as a Platform for Characterizing Multimeric Membrane Protein Complexes. *J. Am. Soc. Mass Spectrom.* **2018**, *29*, 183−193.

(47) Susa, A. C.; Lippens, J. L.; Xia, Z.; Loo, J. A.; Campuzano, I. D. G.; Williams, E. R. Submicrometer Emitter ESI Tips for Native Mass Spectrometry of Membrane Proteins in Ionic and Nonionic Detergents. *J. Am. Soc. Mass Spectrom.* **2018**, *29*, 203−206.