New Mass Spectrometry Approaches to Investigate Proteins and Protein Interactions

Daniele Canzani

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2021

Reading Committee:

Matthew Bush, Chair

František Tureček

Miklos Guttman

Program Authorized to Offer Degree:

Chemistry

University of Washington

Abstract

New Mass Spectrometry Approaches to Investigate Proteins and Protein Interactions

Daniele Canzani

Chair of the Supervisory Committee:

Associate Professor Matthew F. Bush

Department of Chemistry

Mass spectrometry (MS) is essential for understanding the composition, structures, and interactions of proteins. There are various forms of MS that can be used to obtain valuable information about proteins and protein interactions. In native MS, intact protein ions are generated from a solution that maintains the non-covalent protein interactions into the gas-phase environment of a mass spectrometer. MS is frequently coupled liquid chromatography (LC), which is an analytical technique that is used to separate proteins or peptides before they enter the mass spectrometer. LC can improve the sensitivity of an MS analysis and increase the amount of information that can be obtained from a sample. Ion mobility (IM) is a gas-phase separation technique that occurs after a protein sample has been ionized. IM-MS measurements can provide simultaneous details about the mass and size of a protein ion, which is useful for structural biology applications. In the following chapters of this dissertation, various MS, LC, and IM

techniques were used to study protein composition, interactions, and structure. In Chapter 2, native anion exchange chromatography (AEX), which uses a low-pressure and bioinert workflow with ammonium acetate mobile phases, is used to enhance native MS by removing common contaminants that interfere with MS, separating protein variants based on their charge, and isolating proteins from complex mixtures. Those capabilities demonstrate that native AEX is a valuable tool for overcoming common challenges that are associated with native MS, and that native AEX has distinct advantages over other conventional sample preparation methods. In Chapter 3, an integrative MS strategy was developed to identify binding interactions that essential proteins called E3 ligases have in cells. Eukaryotic cells use hundreds of unique E3 ligases to control cellular functions through protein degradation. E3 ligases selectively bind to protein substrates through recognition motifs known as degrons. The integrative MS strategy, which uses a combination of native MS, native top-down MS, MS of destabilized samples, and LC-MS, is called “degronomics” and will be useful for uncovering degron motifs that other E3s can recognize in cells. Results from that study demonstrate that an E3 ligase called KLHDC2 binds to C-terminus diglycine degrons by uncovering KLHDC2-peptide binding interactions from cells. In Chapter 4, a combination of IM-MS experiments and computational tools was used to characterize the effects of charge state, charge distribution, and structure on the ion mobility of proteins in nitrogen gas. The results from Chapter 4 have implications for interpreting protein structures from IM-MS experiments and for comparing nitrogen gas-based measurements against measurements taken with helium gas.

Acknowledgements

There are many people to thank for my success in graduate. First, I would like to thank my advisor, Professor Matthew F. Bush, who provided me with invaluable advice, encouragement, and independence. Matt provided me with a wonderful research environment where I was able to explore many areas of research and grow as a scientist. I will always appreciate Matt's amazing ability and advice on communicating science. I would also like to thank František Tureček, Miklos Guttman, Shao-En Ong, Dusty Maly, Rob Synovec, and Ashleigh Theberge for serving as committee members throughout graduate school.

I must thank Domiţa-Valeria Rusnac and Ning Zheng for being amazing collaborators and allowing me to work with them on such exciting areas of research. Our collaboration spanned several years, and I always looked forward to receiving new samples to analyze from Domniţa. I was also very fortunate to collaborate with Rachel Klevitt and Chris Woods on a project that allowed me to explore crosslinking mass spectrometry, which was always an interest of mine.

Thank you to all the Bush lab members that I worked with over the years, for your advice on presentations, proofreading, and help in lab. Thank you to Sam, Ken, and Kim for helping me to get started in the Bush lab. I always appreciated Meagan's sense of humor and the interesting snacks CeCe brought to group meetings. Ben, thanks for all the good times outside of the office. Thanks to Casey for the memes and for the help in lab. Chapter 2 would not have been possible without her. Lyndsey, Theresa, AnneClaire, Addison, and Alice: I wish you the best of luck in grad school and look forward to seeing your accomplishments.

To Calvin, Sam, Jacob, Nora, Julia, Claire, Emily, and Zach, thanks for the support, encouragement, and adventures. Thanks to Wyatt and Vin for being great roommates and to Jay, Lou, Hannah, and Conner for the great company during camping, skiing, and dinners together.

Thank you to my family for supporting me throughout my entire education. Mom, Babbo, and Jeff, thank you for all your love and encouragement. Ash, thanks for moving across the globe, for always being there for me, and for always pushing me to explore the world around me.

Contents

Chapter 1. Introduction

1.1. Overview

Proteins are essential molecules that enable the biological functions of all organisms. Proteins are comprised of amino acid chains that are defined by genes. The amino acid chain folds, giving each protein a unique overall structure that provides the basis for its function. The roles of many proteins also require interactions with other proteins or molecules and can require special chemical modifications called post-translational modifications (PTMs). Since the amino acid sequence of a protein can be predicted from the corresponding gene, and the masses of the amino acids are known, mass determination is one of the best tools for studying proteins, interactions that proteins have with other molecules, and chemical modifications to a protein. Masses are determined using mass spectrometry (MS), which can be used to study a single protein at a time or tens of thousands in a single experiment. Knowledge of a protein's structure is critical for understanding the nature of its biological function, and many tools have been developed to determine protein structure, including X-ray crystallography, nuclear magnetic resonance (NMR), and cryo-electron microscopy (cryo-EM).[1] MS has also become a valuable tool for studying structures of proteins when combined with orthogonal tools like ion mobility (IM).[2] Thus, MS-based analysis of proteins has become crucial for understanding the composition of proteins and protein interactions, unraveling the complexity of protein systems, and for guiding and enhancing structural biology studies. The work in this dissertation utilizes various modes of mass spectrometry and complementary tools to investigate the composition and interactions of proteins.

1.2. Mass Spectrometry

Mass spectrometry (MS) is an analytical tool that measures the mass-to-charge (m/z) ratio of a molecular ion in the gas-phase. Measurements of m/z can be used to determine the molecular weight, elemental composition, and chemical structure of a compound. A mass spectrometer consists of an ion source, an analyzer, and a detector.[3] An ion source converts molecules into positively or negatively charged ions, which is necessary because ions are moved and stored by electrical fields within an MS instrument. Ions are separated by time or space their m/z ratios in the mass analyzer, which include time-of-flight, quadrupole, and ion traps.[4]

1.3. Electrospray Ionization and nano Electrospray Ionization

One of the most widely used ion sources for MS is electrospray ionization (ESI), which uses a high voltage to aerosolize a liquid sample that is contained in a capillary.[5,6] The high electrical potential causes a Taylor cone to form at the tip of the capillary, which emits a plume of charged droplets that contain the analytes from the sample.[7] Although the mechanisms behind ESI are not fully understood, there are several models that explain how different classes of analytes are ionized. Small molecules are thought to ionize through the ion evaporation model (IEM), which states that an ion will be emitted from a droplet when the field strength at the surface is sufficiently large to overcome solvation energy of the ion.[8] Large molecules, such as proteins, are through to ionize through the charged-residue model (CRM), where neutral solvent molecules evaporate from the initial charged droplet until the repulsion between charges exceeds the surface tension of the droplet, at which point Coulomb fission occurs.[9,10] This causes many smaller progeny droplets to form, and this process repeats until a single desolvated and charged gas-phase ion is left.

There are some disadvantages to using ESI, which can lead to reduced sensitivity or an inaccurate representation of the abundances of the analytes in a sample. Competition for charge occurs when different analytes are present in a sample and make it into the same electrospray droplet. Certain physical properties, such as hydrophobicity, can enhance or suppress the ionization efficiency of a molecule, leading to discrepancies between the observed intensity of an ion in a mass spectrum compared to its concentration in a sample.[11,12] Additionally, adduction of contaminant salts such as Na^+, K^+, or Ca^{2+} onto a molecular ion can disrupt measurements and limit sensitivity.[13,14] Many of these problems stem from the large droplets that are produced from conventional ESI approaches that use high flow rates, high inner diameter capillaries, and high voltages. A partial solution to these challenges is to use nanoelectrospray ionization (nESI), which uses capillaries with smaller inner diameters (<75 μm), lower applied voltages (<1.5 kV), and low flow rates.[7,15] Those changes reduce the size of the initial droplets, which limits non-specific adduction,[13,16] reduces ion suppression effects,[17] and increases overall ionization efficiency.[15] In most ESI and nESI applications, a sample is continually pumped into the emitter, but some iterations of nESI use electroosmotic flow without a pump to produce a steam of ions. This variant is termed "electrokinetic nESI" and enables flow rates of less than 10 nL min^{-1}, and is most commonly used for intact protein analysis with MS (Figure 1).[16]

1.4. Analysis of Proteins with Mass Spectrometry

1.4.1 Intact Proteins and Native Mass Spectrometry. The earliest iterations of intact protein MS used conventional ESI sources and solutions that contained acidified organic solvents that denatured or unfolded the proteins in the sample.[18] Although denaturing approaches are useful for determining the molecular weight of a protein, they disrupt protein structure and any interactions that a protein has with other molecules in solution. Ammonium acetate, a non-

denaturing and volatile solution that is compatible with ESI, was determined to enable intact protein MS and maintain non-covalent interactions between protein subunits,[19–21] and between small molecule ligands.[22] Those early experiments led to the widespread use of native MS (Figure 1),[23] which has become an essential tool in molecular and structural biology.[23,24] Native MS can be used for determining the composition and stoichiometry of proteins and protein complexes,[25–28] to identify protein cofactors,[29,30] and proteoforms.[29,31,32] Native MS is also becoming an important tool in industry, where it has been used characterize small molecule drugs that act as molecular glues between E3 ligases and substrate proteins,[33] or to characterize *de novo* proteins[34] and antibodies.[35]

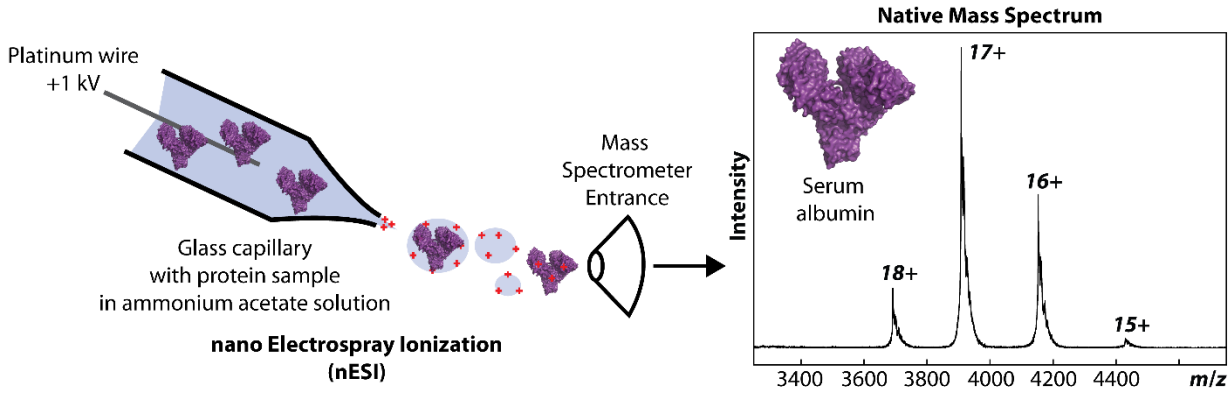

Figure 1.1. Native mass spectrometry uses nanoelectrospray ionization (nESI) to produce bare protein ions from an ammonium acetate solution. A native mass spectrum of the protein bovine serum albumin that was acquired on a time-of-flight instrument is shown as an example. The individual protein molecules can acquire different amounts of charge during the ionization process, leading to multiple peaks that appear across a range of mass-to-charge (*m/z)* values.

1.4.2 Analysis of Peptides. The alternative to studying intact proteins with MS is called a "bottom-up" approach; proteins are extracted from cell lysates or tissues, denatured, and then digested with a protease like trypsin to produce peptides.[36,37] An ionized peptide can be measured directly by a mass spectrometer, which is called a precursor spectrum or MS1. The peptide ion can be isolated and subjection to collision-induced dissociation (CID), where it is accelerated into a neutral gas (argon or nitrogen) in order to fragment the peptide into characteristic *b*- and *y*-ions that can be used to reconstruct its amino acid sequence.[38,39] These fragmentation spectra are known as tandem mass spectra (also MS/MS, MS2). Tandem mass spectra can be matched to a peptide sequence using database searching, where the experimental spectra are compared to theoretical tandem spectra that are produced from all possible enzymatic peptides in a protein sequence database.[40,41] Spectral library searching is also commonly used, in which experimental tandem spectra are compared with previously acquired and sequenced reference spectra.[42,43] Another option is to use *de novo* sequencing, where the amino acid sequence of a spectrum is assigned from the fragment ions without the use of a database.[44] Some *de novo* sequencing platforms use database searching after a preliminary peptide sequence is identified in order to increase confidence and accuracy in the assignments.[45,46] Most algorithms to interpret tandem mass spectra were developed to identify enzymatic peptides, but many tools have options for non-enzymatic searches,[41,47] which are useful in peptidomics[48] experiments that analyze naturally occurring peptides.[49]

1.4.3 Liquid Chromatography – Mass Spectrometry. Liquid chromatography (LC) is an essential analytical technique that separates the molecules in a sample based on their relative affinities for the stationary phase and the mobile phase. The stationary phase is immobilized inside of a column, and a liquid mobile phase is pumped through the column, leading to

separation (Figure 2). LC coupled to MS (LC-MS) has become indispensable for protein and peptide analysis. Perhaps the most common iteration of LC-MS uses reversed-phase chromatography, which is based on a hydrophobic stationary phase (usually C18, or octadecyl carbon chains bonded to silica beads) and an aqueous mobile phase that is modified with an organic solvent such as acetonitrile. Reversed-phase LC is standard for peptide separations, which typically utilize nano-scale flow rates of 0.1 to 1.0 μL min^{-1}.[50] Peptides bind to the stationary phase and elute off the column based on their hydrophobicity as the percentage of organic solvent in the mobile phase increases, which is called a gradient. A similar approach is used for denatured proteins, but C4 stationary phases (butyl carbon chains bonded to silica beads) are more common due to the increased hydrophobicity of a full protein compared to a peptide.[51]

LC applications for native MS are not as mature as the reversed-phase separations described above, which are denaturing. Efforts to couple LC to native MS include size-exclusion chromatography,[52–54] ion-exchange chromatography,[55–57] and hydrophobic interaction chromatography.[58,59] One major challenge has been coupling native separations with nESI due to the high flow rates that are commonly used. A recent solution has been to use commercially available multi-tip nESI emitters that split the LC flow into multiple electrospray channels to improve performance and senstivity.[59]

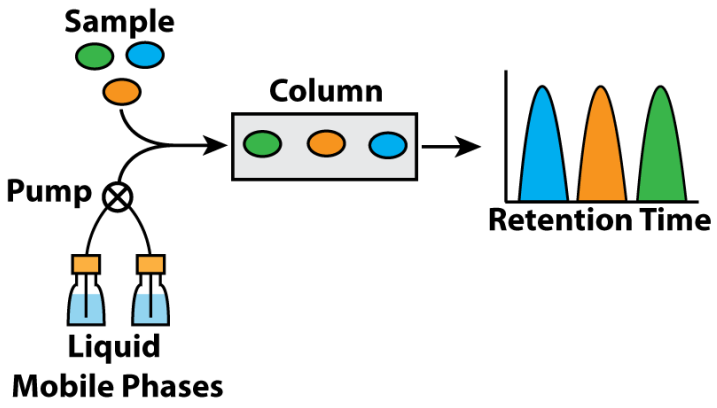

Figure 1.2. Liquid chromatography separates molecules in a sample based on their affinity for a stationary phase that is embedded in the column, and for the liquid mobile phases that are pumped through the column. As molecules elute from the column, they are detected using mass spectrometry, ultraviolet absorbance, or other methods. The time it takes for a molecule to elute off the column is called the retention time.

1.5. Ion mobility

1.5.1 Ion Mobility. Ion mobility (IM) is a gas-phase technique that separates ions based on their size and charge. IM experiments measure the time it takes for an ion to pass through a drift region that is filled with a neutral background gas (typically He or N_2) under an applied electric field (E). The measured drift time (t_D) can be used to calculate the mobility (K) of an ion:

$$K = \frac{L}{t_D E} \qquad \text{Equation 1.1}$$

where L is the length of the drift region. K can be used determine a collision cross section (Ω), which is the momentum transfer integral of many ion-background gas collisions that occur

throughout an IM separation.[60] Ω provides an estimate of the size of an ion. *K* is related to Ω through the Mason-Schamp equation:[60]

$$\Omega = \frac{3ez}{16N}\sqrt{\frac{2\pi}{\mu k_B T}}\frac{1}{K} \qquad \text{Equation 1.2}$$

where *N* is the number density of the drift gas, *e* is the elementary charge, *z* is the charge state of the ion, μ is the reduced mass of the ion and neutral gas pair, k_B is Boltzmann's constant, and *T* is the temperature of the drift-gas.

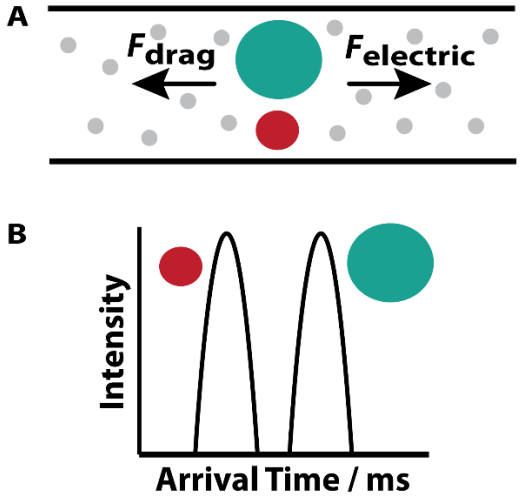

Figure 1.3. (A) Ion mobility (IM) separates ions by their mobility (*K*) through a drift region that is under an applied electric field and filled with a neutral drift gas (gray circles). The ions are pulled through the drift region by the electric field but experience an opposing drag force due to collisions with the drift gas. Larger ions experience more collisions, resulting in reduced mobility compared to smaller ions. (B) The arrival time of ions are measured by a mass spectrometer as they pass through the IM cell. Smaller ions have greater mobilities, leading to shorter arrival times. Mobility values that are derived from

arrival times are used to calculate collision cross section (Ω) values, which provide estimates of ion size.

1.5.2. Ion Mobility-Mass Spectrometry of Proteins. IM usage has seen a boom in recent years thanks to commercial implementations that unify IM separations with time-of-flight (TOF) mass spectrometers.[61–63] TOFs acquire spectra in ~ 100 microseconds, allowing hundred to thousands of measurements to be made throughout an IM separation, which occur on millisecond timescales. The combination of IM-MS allows simultaneous determination of m/z and Ω, or information on mass and size of an ion. This proved to be valuable for characterizing many types of analytes in biology, including carbohydrates,[64,65] lipids, and metabolites. It has also achieved significant success for characterizing native-like protein ions in structural biology applications.[66–69] IM-MS is particularly useful for studying proteins that resist conventional structural techniques such as X-ray crystallography, cryo-electron microscopy, or nuclear magnetic resonance. Examples of this include amyloid[70,71] and small heat shock proteins[72] that have unstructured regions which are difficult to study with other methods.

1.5.3. Ω Calculations. Experimentally determined Ω values of proteins can be used for structural analysis through comparison to Ω values that have been calculated for candidate structural models.[73,74] There are several methods for calculating Ω, including the projection approximation (PA), which is the simplest implementation that uses the rotationally-averaged project area of a protein.[75] The exact hard-spheres scattering method (EHSS) calculates Ω based on the estimated momentum transfer between elastic collisions between an ion and the neutral drift gas, with all atoms in the model treated as hard spheres.[76] The diffuse hard sphere scattering (DHSS) method is similar to EHSS, except it incorporates inelastic and diffuse collisions that may more accurately model ion collisions with polar drift gases like N_2.[77,78] The trajectory

method (TJM) accounts for the charge of an ion by calculating Ω values using an accurate potential between the ion-gas pair.[79] The structure relaxation approximation (SRA) was developed to predict Ω from an ensemble of protein structures, rather than a single structure like in the earlier methods.[80] Other methods to use experimental Ω values to guide protein structure prediction or modelling are available. The Integrative Modeling Platform[81] now incorporates course-grained protein complex topology using molecular radii that are determined from experimental Ω values. Experimental Ω values can also be combined with molecular dynamics simulations, which has been used to interpret protein unfolding trajectories.[82]

1.6. Outline of this Dissertation

In this dissertation, I demonstrate how various modes of mass spectrometry and chromatography can be used to examine the composition of proteins and protein complexes. I also explore the fundamentals of collision cross section values that are obtained for protein ions with nitrogen gas.

Chapter 2 describes the development of ammonium acetate-based anion exchange chromatography, or native AEX, as a tool to enhance native MS studies. Native AEX can be used for rapid sample clean-up and preparation for native MS and is a promising alternative to standard methods that native MS researchers use for sample preparation. Native AEX can also be used to separate proteins and protein variants from complex mixtures, and to enhance proteoform or protein cofactor identification with native MS.

Chapter 3 describes the development of an integrative MS strategy to study the interactions that essential enzymes called E3 ligases have with peptides in cell culture.[83] The strategy, called degronomics, uses native top-down MS, LC-MS/MS and destabilizing MS to determine the degron binding motifs that E3 ligases use to identify and bind to their substrates.

This research was based on a collaboration with Prof. Ning Zheng and Dr. Domniţa-Valeria Rusnac in the Department of Pharmacology at the University of Washington and our previous work to characterize an E3 ligase called KLHDC2.[84]

Chapter 4 describes the effects of charge state, structure, and shape on Ω values of protein ions that are measured with N_2 gas.[85] This study used results from IM-MS experiments and Ω values that were determined using trajectory method calculations. This work demonstrates that long-range ion-induced dipole interactions strongly increase Ω values with N_2 gas, but not with He gas. Those results have implications for reporting Ω values and converting between Ω values that are measured with different drift gases, even when calibrants are used.[86]

1.7. References

(1) Dobson, C. M. Biophysical Techniques in Structural Biology. *Annu. Rev. Biochem.* **2019**, *88* (1), 25–33. https://doi.org/10.1146/annurev-biochem-013118-111947.

(2) Ruotolo, B. T.; Benesch, J. L. P.; Sandercock, A. M.; Hyung, S.-J.; Robinson, C. V. Ion Mobility–Mass Spectrometry Analysis of Large Protein Complexes. *Nat. Protoc.* **2008**, *3* (7), 1139–1152. https://doi.org/10.1038/nprot.2008.78.

(3) Haag, A. M. Mass Analyzers and Mass Spectrometers. In *Modern Proteomics – Sample Preparation, Analysis and Practical Applications*; Mirzaei, H., Carrasco, M., Eds.; Springer International Publishing: Cham, 2016; pp 157–169. https://doi.org/10.1007/978-3-319-41448-5_7.

(4) Glish, G. L.; Vachet, R. W. The Basics of Mass Spectrometry in the Twenty-First Century. *Nature Reviews Drug Discovery* **2003**, *2* (2), 140–150. https://doi.org/10.1038/nrd1011.

(5) Dole, M.; Mack, L. L.; Hines, R. L.; Mobley, R. C.; Ferguson, L. D.; Alice, M. B. Molecular Beams of Macroions. *J. Chem. Phys.* **1968**, *49* (5), 2240–2249. https://doi.org/10.1063/1.1670391.

(6) Yamashita, M.; Fenn, J. B. Electrospray Ion Source. Another Variation on the Free-Jet Theme. *J. Phys. Chem.* **1984**, *88* (20), 4451–4459. https://doi.org/10.1021/j150664a002.

(7) Wilm, M. S.; Mann, M. Electrospray and Taylor-Cone Theory, Dole's Beam of Macromolecules at Last? *International Journal of Mass Spectrometry and Ion Processes* **1994**, *136* (2), 167–180. https://doi.org/10.1016/0168-1176(94)04024-9.

(8) Kebarle, P.; Peschke, M. On the Mechanisms by Which the Charged Droplets Produced by Electrospray Lead to Gas Phase Ions. *Analytica Chimica Acta* **2000**, *406* (1), 11–35. https://doi.org/10.1016/S0003-2670(99)00598-X.

(9) Gomez, A.; Tang, K. Charge and Fission of Droplets in Electrostatic Sprays. *Physics of Fluids* **1994**, *6* (1), 404–414. https://doi.org/10.1063/1.868037.

(10) Fernandez de la Mora, J. Electrospray Ionization of Large Multiply Charged Species Proceeds via Dole's Charged Residue Mechanism. *Analytica Chimica Acta* **2000**, *406* (1), 93–104. https://doi.org/10.1016/S0003-2670(99)00601-7.

(11) Iribarne, J. V.; Dziedzic, P. J.; Thomson, B. A. Atmospheric Pressure Ion Evaporation-Mass Spectrometry. *International Journal of Mass Spectrometry and Ion Physics* **1983**, *50* (3), 331–347. https://doi.org/10.1016/0020-7381(83)87009-0.

(12) Hirabayashi, A.; Ishimaru, M.; Manri, N.; Yokosuka, T.; Hanzawa, H. Detection of Potential Ion Suppression for Peptide Analysis in Nanoflow Liquid Chromatography/Mass Spectrometry. *Rapid Communications in Mass Spectrometry* **2007**, *21* (17), 2860–2866. https://doi.org/10.1002/rcm.3157.

(13) Juraschek, R.; Dülcks, T.; Karas, M. Nanoelectrospray—More than Just a Minimized-Flow Electrospray Ionization Source. *Journal of the American Society for Mass Spectrometry* **1999**, *10* (4), 300–308. https://doi.org/10.1016/S1044-0305(98)00157-3.

(14) Felitsyn, N.; Peschke, M.; Kebarle, P. Origin and Number of Charges Observed on Multiply-Protonated Native Proteins Produced by ESI. *International Journal of Mass Spectrometry* **2002**, *219* (1), 39–63. https://doi.org/10.1016/S1387-3806(02)00588-2.

(15) Wilm, M.; Mann, M. Analytical Properties of the Nanoelectrospray Ion Source. *Anal. Chem.* **1996**, *68* (1), 1–8. https://doi.org/10.1021/ac9509519.

(16) Davidson, K. L.; Oberreit, D. R.; Hogan, C. J.; Bush, M. F. Nonspecific Aggregation in Native Electrokinetic Nanoelectrospray Ionization. *International Journal of Mass Spectrometry* **2017**, *420*, 35–42. https://doi.org/10.1016/j.ijms.2016.09.013.

(17) Gangl, E. T.; Annan, M.; Spooner, N.; Vouros, P. Reduction of Signal Suppression Effects in ESI-MS Using a Nanosplitting Device. *Anal. Chem.* **2001**, *73* (23), 5635–5644. https://doi.org/10.1021/ac010501i.

(18) Fenn, J.; Mann, M.; Meng, C.; Wong, S.; Whitehouse, C. Electrospray Ionization for Mass Spectrometry of Large Biomolecules. *Science* **1989**, *246* (4926), 64. https://doi.org/10.1126/science.2675315.

(19) Light-Wahl, K. J.; Winger, B. E.; Smith, R. D. Observation of the Multimeric Forms of Concanavalin A by Electrospray Ionization Mass Spectrometry. *J. Am. Chem. Soc.* **1993**, *115* (13), 5869–5870. https://doi.org/10.1021/ja00066a083.

(20) Schwartz, B. L.; Bruce, J. E.; Anderson, G. A.; Hofstadler, S. A.; Rockwood, A. L.; Smith, R. D.; Chilkoti, A.; Stayton, P. S. Dissociation of Tetrameric Ions of Noncovalent Streptavidin Complexes Formed by Electrospray Ionization. *Journal of the American*

Society for Mass Spectrometry **1995**, *6* (6), 459–465. https://doi.org/10.1016/1044-0305(95)00191-F.

(21) Loo, J. A. Observation of Large Subunit Protein Complexes by Electrospray Ionization Mass Spectrometry. *Journal of Mass Spectrometry* **1995**, *30* (1), 180–183. https://doi.org/10.1002/jms.1190300127.

(22) Eckart, K.; Spiess, J. Electrospray Ionization Mass Spectrometry of Biotin Binding to Streptavidin. *Journal of the American Society for Mass Spectrometry* **1995**, *6* (10), 912–919. https://doi.org/10.1016/1044-0305(95)00480-2.

(23) Heck, A. J. R. Native Mass Spectrometry: A Bridge between Interactomics and Structural Biology. *Nature Methods* **2008**, *5*, 927–933. https://doi.org/10.1038/nmeth.1265.

(24) Robinson, C. V.; Sali, A.; Baumeister, W. The Molecular Sociology of the Cell. *Nature* **2007**, *450* (7172), 973–982. https://doi.org/10.1038/nature06523.

(25) Zhou, M.; Sandercock, A. M.; Fraser, C. S.; Ridlova, G.; Stephens, E.; Schenauer, M. R.; Yokoi-Fong, T.; Barsky, D.; Leary, J. A.; Hershey, J. W.; Doudna, J. A.; Robinson, C. V. Mass Spectrometry Reveals Modularity and a Complete Subunit Interaction Map of the Eukaryotic Translation Factor EIF3. *Proc Natl Acad Sci USA* **2008**, *105* (47), 18139–18144. https://doi.org/10.1073/pnas.0801313105.

(26) Stengel, F.; Baldwin, A. J.; Bush, M. F.; Hilton, G. R.; Lioe, H.; Basha, E.; Jaya, N.; Vierling, E.; Benesch, J. L. P. Dissecting Heterogeneous Molecular Chaperone Complexes Using a Mass Spectrum Deconvolution Approach. *Chemistry & Biology* **2012**, *19* (5), 599–607. https://doi.org/10.1016/j.chembiol.2012.04.007.

(27) Snijder, J.; van de Waterbeemd, M.; Damoc, E.; Denisov, E.; Grinfeld, D.; Bennett, A.; Agbandje-McKenna, M.; Makarov, A.; Heck, A. J. R. Defining the Stoichiometry and

Cargo Load of Viral and Bacterial Nanoparticles by Orbitrap Mass Spectrometry. *J. Am. Chem. Soc.* **2014**, *136* (20), 7295–7299. https://doi.org/10.1021/ja502616y.

(28) Lu, C.; Turley, S.; Marionni, S. T.; Park, Y.-J.; Lee, K. K.; Patrick, M.; Shah, R.; Sandkvist, M.; Bush, M. F.; Hol, W. G. J. Hexamers of the Type II Secretion ATPase GspE from Vibrio Cholerae with Increased ATPase Activity. *Structure* **2013**, *21* (9), 1707–1717. https://doi.org/10.1016/j.str.2013.06.027.

(29) Skinner, O. S.; Haverland, N. A.; Fornelli, L.; Melani, R. D.; Do Vale, L. H. F.; Seckler, H. S.; Doubleday, P. F.; Schachner, L. F.; Srzentić, K.; Kelleher, N. L.; Compton, P. D. Top-down Characterization of Endogenous Protein Complexes with Native Proteomics. *Nature Chemical Biology* **2018**, *14*, 36–41. https://doi.org/10.1038/nchembio.2515.

(30) Wang, H.; Shi, H.; Rajan, M.; Canarie, E. R.; Hong, S.; Simoneschi, D.; Pagano, M.; Bush, M. F.; Stoll, S.; Leibold, E. A.; Zheng, N. FBXL5 Regulates IRP2 Stability in Iron Homeostasis via an Oxygen-Responsive [2Fe2S] Cluster. *Mol Cell* **2020**, *78* (1), 31-41.e5. https://doi.org/10.1016/j.molcel.2020.02.011.

(31) Smith, L. M.; Kelleher, N. L.; Linial, M.; Goodlett, D.; Langridge-Smith, P.; Ah Goo, Y.; Safford, G.; Bonilla, L.; Kruppa, G.; Zubarev, R.; Rontree, J.; Chamot-Rooke, J.; Garavelli, J.; Heck, A.; Loo, J.; Penque, D.; Hornshaw, M.; Hendrickson, C.; Pasa-Tolic, L.; Borchers, C.; Chan, D.; Young, N.; Agar, J.; Masselon, C.; Gross, M.; McLafferty, F.; Tsybin, Y.; Ge, Y.; Sanders, I.; Langridge, J.; Whitelegge, J.; Marshall, A.; The Consortium for Top Down Proteomics. Proteoform: A Single Term Describing Protein Complexity. *Nature Methods* **2013**, *10* (3), 186–187. https://doi.org/10.1038/nmeth.2369.

(32) 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. *Nature Chemistry* **2018**, *10*, 139–148. https://doi.org/10.1038/nchem.2908.

(33) Beveridge, R.; Kessler, D.; Rumpel, K.; Ettmayer, P.; Meinhart, A.; Clausen, T. Native Mass Spectrometry Can Effectively Predict PROTAC Efficacy. *ACS Cent. Sci.* **2020**, *6* (7), 1223–1230. https://doi.org/10.1021/acscentsci.0c00049.

(34) Linsky, T. W.; Vergara, R.; Codina, N.; Nelson, J. W.; Walker, M. J.; Su, W.; Barnes, C. O.; Hsiang, T.-Y.; Esser-Nobis, K.; Yu, K.; Reneer, Z. B.; Hou, Y. J.; Priya, T.; Mitsumoto, M.; Pong, A.; Lau, U. Y.; Mason, M. L.; Chen, J.; Chen, A.; Berrocal, T.; Peng, H.; Clairmont, N. S.; Castellanos, J.; Lin, Y.-R.; Josephson-Day, A.; Baric, R. S.; Fuller, D. H.; Walkey, C. D.; Ross, T. M.; Swanson, R.; Bjorkman, P. J.; Gale, M.; Blancas-Mejia, L. M.; Yen, H.-L.; Silva, D.-A. De Novo Design of Potent and Resilient HACE2 Decoys to Neutralize SARS-CoV-2. *Science* **2020**, *370* (6521), 1208. https://doi.org/10.1126/science.abe0075.

(35) Campuzano, I. D. G.; Netirojjanakul, C.; Nshanian, M.; Lippens, J. L.; Kilgour, D. P. A.; Van Orden, S.; Loo, J. A. Native-MS Analysis of Monoclonal Antibody Conjugates by Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. *Anal. Chem.* **2018**, *90* (1), 745–751. https://doi.org/10.1021/acs.analchem.7b03021.

(36) Aebersold, R.; Mann, M. Mass Spectrometry-Based Proteomics. *Nature* **2003**, *422*, 198. https://doi.org/10.1038/nature01511.

(37) Chait, B. T. Mass Spectrometry: Bottom-Up or Top-Down? *Science* **2006**, *314* (5796), 65. https://doi.org/10.1126/science.1133987.

(38) Mitchell Wells, J.; McLuckey, S. A. Collision-Induced Dissociation (CID) of Peptides and Proteins. In *Methods in Enzymology*; Academic Press, 2005; Vol. 402, pp 148–185. https://doi.org/10.1016/S0076-6879(05)02005-7.

(39) Roepstorff, P.; Fohlman, J. Proposal for a Common Nomenclature for Sequence Ions in Mass Spectra of Peptides. *Biomedical Mass Spectrometry* **1994**, *11* (11), 601.

(40) Eng, J. K.; McCormack, A. L.; Yates, J. R., 3rd. An Approach to Correlate Tandem Mass Spectral Data of Peptides with Amino Acid Sequences in a Protein Database. *Journal of The American Society for Mass Spectrometry* **1994**, *5* (11), 976–989. https://doi.org/10.1016/1044-0305(94)80016-2.

(41) Eng, J. K.; Jahan, T. A.; Hoopmann, M. R. Comet: An Open-Source MS/MS Sequence Database Search Tool. *PROTEOMICS* **2013**, *13* (1), 22–24. https://doi.org/10.1002/pmic.201200439.

(42) Craig, R.; Cortens, J. P.; Beavis, R. C. The Use of Proteotypic Peptide Libraries for Protein Identification. *Rapid Communications in Mass Spectrometry* **2005**, *19* (13), 1844–1850. https://doi.org/10.1002/rcm.1992.

(43) Griss, J. Spectral Library Searching in Proteomics. *PROTEOMICS* **2016**, *16* (5), 729–740. https://doi.org/10.1002/pmic.201500296.

(44) Dančík, V.; Addona, T. A.; Clauser, K. R.; Vath, J. E.; Pevzner, P. A. De Novo Peptide Sequencing via Tandem Mass Spectrometry. *Journal of Computational Biology* **1999**, *6* (3–4), 327–342. https://doi.org/10.1089/106652799318300.

(45) Taylor, J. A.; Johnson, R. S. Sequence Database Searches via de Novo Peptide Sequencing by Tandem Mass Spectrometry. *Rapid Communications in Mass Spectrometry* **1997**, *11* (9), 1067–1075. https://doi.org/10.1385/MB:22:3:301.

(46) Zhang, J.; Xin, L.; Shan, B.; Chen, W.; Xie, M.; Yuen, D.; Zhang, W.; Zhang, Z.; Lajoie, G. A.; Ma, B. PEAKS DB: De Novo Sequencing Assisted Database Search for Sensitive and Accurate Peptide Identification. *Mol Cell Proteomics* **2012**, *11* (4), M111.010587. https://doi.org/10.1074/mcp.M111.010587.

(47) Kong, A. T.; Leprevost, F. V.; Avtonomov, D. M.; Mellacheruvu, D.; Nesvizhskii, A. I. MSFragger: Ultrafast and Comprehensive Peptide Identification in Mass Spectrometry–Based Proteomics. *Nature Methods* **2017**, *14*, 513.

(48) Baggerman, G.; Verleyen, P.; Clynen, E.; Huybrechts, J.; De Loof, A.; Schoofs, L. Peptidomics. *Journal of Chromatography B* **2004**, *803* (1), 3–16. https://doi.org/10.1016/j.jchromb.2003.07.019.

(49) Dallas, D. C.; Guerrero, A.; Parker, E. A.; Robinson, R. C.; Gan, J.; German, J. B.; Barile, D.; Lebrilla, C. B. Current Peptidomics: Applications, Purification, Identification, Quantification, and Functional Analysis. *Proteomics* **2015**, *15* (5–6), 1026–1038. https://doi.org/10.1002/pmic.201400310.

(50) Zhang, Y.; Fonslow, B. R.; Shan, B.; Baek, M.-C.; Yates, J. R., 3rd. Protein Analysis by Shotgun/Bottom-up Proteomics. *Chemical reviews* **2013**, *113* (4), 2343–2394. https://doi.org/10.1021/cr3003533.

(51) Neville, B. Reversed-Phase Chromatography of Proteins. In *Protein Purification Protocols*; Doonan, S., Ed.; Humana Press: Totowa, NJ, 1996; pp 277–292. https://doi.org/10.1385/0-89603-336-8:277.

(52) Shen, M. L.; Benson, L. M.; Johnson, K. L.; Lipsky, J. J.; Naylor, S. Effect of Enzyme Inhibitors on Protein Quaternary Structure Determined by On-Line Size Exclusion Chromatography-Microelectrospray Ionization Mass Spectrometry. *Journal of the*

American Society for Mass Spectrometry **2001**, *12* (1), 97–104. https://doi.org/10.1016/S1044-0305(00)00190-2.

(53) Muneeruddin, K.; Thomas, J. J.; Salinas, P. A.; Kaltashov, I. A. Characterization of Small Protein Aggregates and Oligomers Using Size Exclusion Chromatography with Online Detection by Native Electrospray Ionization Mass Spectrometry. *Anal. Chem.* **2014**, *86* (21), 10692–10699. https://doi.org/10.1021/ac502590h.

(54) Ehkirch, A.; Hernandez-Alba, O.; Colas, O.; Beck, A.; Guillarme, D.; Cianférani, S. Hyphenation of Size Exclusion Chromatography to Native Ion Mobility Mass Spectrometry for the Analytical Characterization of Therapeutic Antibodies and Related Products. *Journal of Chromatography B* **2018**, *1086*, 176–183. https://doi.org/10.1016/j.jchromb.2018.04.010.

(55) Yan, Y.; Liu, A. P.; Wang, S.; Daly, T. J.; Li, N. Ultrasensitive Characterization of Charge Heterogeneity of Therapeutic Monoclonal Antibodies Using Strong Cation Exchange Chromatography Coupled to Native Mass Spectrometry. *Anal. Chem.* **2018**, *90* (21), 13013–13020. https://doi.org/10.1021/acs.analchem.8b03773.

(56) Ma, F.; Raoufi, F.; Bailly, M. A.; Fayadat-Dilman, L.; Tomazela, D. Hyphenation of Strong Cation Exchange Chromatography to Native Mass Spectrometry for High Throughput Online Characterization of Charge Heterogeneity of Therapeutic Monoclonal Antibodies. *mABs* **2020**, *12* (1), e1763762. https://doi.org/10.1080/19420862.2020.1763762.

(57) Matsuda, Y.; Kliman, M.; Mendelsohn, B. A. Application of Native Ion Exchange Mass Spectrometry to Intact and Subunit Analysis of Site-Specific Antibody–Drug Conjugates

Produced by AJICAP First Generation Technology. *J. Am. Soc. Mass Spectrom.* **2020**, *31* (8), 1706–1712. https://doi.org/10.1021/jasms.0c00129.

(58) Wei, B.; Han, G.; Tang, J.; Sandoval, W.; Zhang, Y. T. Native Hydrophobic Interaction Chromatography Hyphenated to Mass Spectrometry for Characterization of Monoclonal Antibody Minor Variants. *Anal. Chem.* **2019**, *91* (24), 15360–15364. https://doi.org/10.1021/acs.analchem.9b04467.

(59) Yan, Y.; Xing, T.; Wang, S.; Daly, T. J.; Li, N. Online Coupling of Analytical Hydrophobic Interaction Chromatography with Native Mass Spectrometry for the Characterization of Monoclonal Antibodies and Related Products. *Journal of Pharmaceutical and Biomedical Analysis* **2020**, *186*, 113313. https://doi.org/10.1016/j.jpba.2020.113313.

(60) Mason, E.; McDaniel. *Transport Properties of Ions in Gases*; John Wiley and Sons, Inc.: New York, New York, 1988.

(61) Giles, K.; Williams, J. P.; Campuzano, I. Enhancements in Travelling Wave Ion Mobility Resolution. *Rapid Communications in Mass Spectrometry* **2011**, *25* (11), 1559–1566. https://doi.org/10.1002/rcm.5013.

(62) May, J. C.; Goodwin, C. R.; Lareau, N. M.; Leaptrot, K. L.; Morris, C. B.; Kurulugama, R. T.; Mordehai, A.; Klein, C.; Barry, W.; Darland, E.; Overney, G.; Imatani, K.; Stafford, G. C.; Fjeldsted, J. C.; McLean, J. A. Conformational Ordering of Biomolecules in the Gas Phase: Nitrogen Collision Cross Sections Measured on a Prototype High Resolution Drift Tube Ion Mobility-Mass Spectrometer. *Anal. Chem.* **2014**, *86* (4), 2107–2116. https://doi.org/10.1021/ac4038448.

(63) Silveira, J. A.; Ridgeway, M. E.; Laukien, F. H.; Mann, M.; Park, M. A. Parallel Accumulation for 100% Duty Cycle Trapped Ion Mobility-Mass Spectrometry. *International Journal of Mass Spectrometry* **2017**, *413*, 168–175. https://doi.org/10.1016/j.ijms.2016.03.004.

(64) Gaye, M. M.; Nagy, G.; Clemmer, D. E.; Pohl, N. L. B. Multidimensional Analysis of 16 Glucose Isomers by Ion Mobility Spectrometry. *Anal. Chem.* **2016**, *88* (4), 2335–2344. https://doi.org/10.1021/acs.analchem.5b04280.

(65) Huang, Y.; Dodds, E. D. Ion Mobility Studies of Carbohydrates as Group I Adducts: Isomer Specific Collisional Cross Section Dependence on Metal Ion Radius. *Anal. Chem.* **2013**, *85* (20), 9728–9735. https://doi.org/10.1021/ac402133f.

(66) Uetrecht, C.; Barbu, I. M.; Shoemaker, G. K.; van Duijn, E.; Heck, A. J. R. Interrogating Viral Capsid Assembly with Ion Mobility–Mass Spectrometry. *Nature Chemistry* **2011**, *3* (2), 126–132. https://doi.org/10.1038/nchem.947.

(67) Pacholarz, K. J.; Porrini, M.; Garlish, R. A.; Burnley, R. J.; Taylor, R. J.; Henry, A. J.; Barran, P. E. Dynamics of Intact Immunoglobulin G Explored by Drift-Tube Ion-Mobility Mass Spectrometry and Molecular Modeling. *Angewandte Chemie International Edition* **2014**, *53* (30), 7765–7769. https://doi.org/10.1002/anie.201402863.

(68) Song, Y.; Nelp, M. T.; Bandarian, V.; Wysocki, V. H. Refining the Structural Model of a Heterohexameric Protein Complex: Surface Induced Dissociation and Ion Mobility Provide Key Connectivity and Topology Information. *ACS Cent. Sci.* **2015**, *1* (9), 477–487. https://doi.org/10.1021/acscentsci.5b00251.

(69) Vimer, S.; Ben-Nissan, G.; Morgenstern, D.; Kumar-Deshmukh, F.; Polkinghorn, C.; Quintyn, R. S.; Vasil'ev, Y. V.; Beckman, J. S.; Elad, N.; Wysocki, V. H.; Sharon, M.

Comparative Structural Analysis of 20S Proteasome Ortholog Protein Complexes by Native Mass Spectrometry. *ACS Cent. Sci.* **2020**, *6* (4), 573–588. https://doi.org/10.1021/acscentsci.0c00080.

(70) Bernstein, S. L.; Dupuis, N. F.; Lazo, N. D.; Wyttenbach, T.; Condron, M. M.; Bitan, G.; Teplow, D. B.; Shea, J.-E.; Ruotolo, B. T.; Robinson, C. V.; Bowers, M. T. Amyloid-β Protein Oligomerization and the Importance of Tetramers and Dodecamers in the Aetiology of Alzheimer's Disease. *Nature Chemistry* **2009**, *1* (4), 326–331. https://doi.org/10.1038/nchem.247.

(71) Young, L. M.; Cao, P.; Raleigh, D. P.; Ashcroft, A. E.; Radford, S. E. Ion Mobility Spectrometry–Mass Spectrometry Defines the Oligomeric Intermediates in Amylin Amyloid Formation and the Mode of Action of Inhibitors. *J. Am. Chem. Soc.* **2014**, *136* (2), 660–670. https://doi.org/10.1021/ja406831n.

(72) Baldwin, A. J.; Lioe, H.; Hilton, G. R.; Baker, L. A.; Rubinstein, J. L.; Kay, L. E.; Benesch, J. L. P. The Polydispersity of AB-Crystallin Is Rationalized by an Interconverting Polyhedral Architecture. *Structure* **2011**, *19* (12), 1855–1863. https://doi.org/10.1016/j.str.2011.09.015.

(73) D'Atri, V.; Porrini, M.; Rosu, F.; Gabelica, V. Linking Molecular Models with Ion Mobility Experiments. Illustration with a Rigid Nucleic Acid Structure. *Journal of Mass Spectrometry* **2015**, *50* (5), 711–726. https://doi.org/10.1002/jms.3590.

(74) Lee, J. W.; Davidson, K. L.; Bush, M. F.; Kim, H. I. Collision Cross Sections and Ion Structures: Development of a General Calculation Method via High-Quality Ion Mobility Measurements and Theoretical Modeling. *Analyst* **2017**, *142*, 4289–4298. https://doi.org/10.1039/C7AN01276D.

(75) Mack, E. Average Cross-Sectional Areas of Molecules by Gaseous Diffusion Methods. *J. Am. Chem. Soc.* **1925**, *47* (10), 2468–2482. https://doi.org/10.1021/ja01687a007.

(76) Shvartsburg, A. A.; Jarrold, M. F. An Exact Hard-Spheres Scattering Model for the Mobilities of Polyatomic Ions. *Chemical Physics Letters* **1996**, *261* (1), 86–91. https://doi.org/10.1016/0009-2614(96)00941-4.

(77) Larriba, C.; Hogan, C. J. Ion Mobilities in Diatomic Gases: Measurement versus Prediction with Non-Specular Scattering Models. *J. Phys. Chem. A* **2013**, *117* (19), 3887–3901. https://doi.org/10.1021/jp312432z.

(78) Larriba, C.; Hogan, C. J. Free Molecular Collision Cross Section Calculation Methods for Nanoparticles and Complex Ions with Energy Accommodation. *Journal of Computational Physics* **2013**, *251* (Supplement C), 344–363. https://doi.org/10.1016/j.jcp.2013.05.038.

(79) Mesleh, M. F.; Hunter, J. M.; Shvartsburg, A. A.; Schatz, G. C.; Jarrold, M. F. Structural Information from Ion Mobility Measurements: Effects of the Long-Range Potential. *J. Phys. Chem.* **1996**, *100* (40), 16082–16086. https://doi.org/10.1021/jp961623v.

(80) Bleiholder, C.; Liu, F. C. Structure Relaxation Approximation (SRA) for Elucidation of Protein Structures from Ion Mobility Measurements. *J. Phys. Chem. B* **2019**, *123* (13), 2756–2769. https://doi.org/10.1021/acs.jpcb.8b11818.

(81) Webb, B.; Viswanath, S.; Bonomi, M.; Pellarin, R.; Greenberg, C. H.; Saltzberg, D.; Sali, A. Integrative Structure Modeling with the Integrative Modeling Platform. *Protein Science* **2018**, *27* (1), 245–258. https://doi.org/10.1002/pro.3311.

(82) Kulesza, A.; Marklund, E. G.; MacAleese, L.; Chirot, F.; Dugourd, P. Bringing Molecular Dynamics and Ion-Mobility Spectrometry Closer Together: Shape Correlations, Structure-

Based Predictors, and Dissociation. *J. Phys. Chem. B* **2018**, *122* (35), 8317–8329. https://doi.org/10.1021/acs.jpcb.8b03825.

(83) Canzani, D.; Rusnac, D.-V.; Zheng, N.; Bush, M. F. Degronomics: Mapping the Interacting Peptidome of a Ubiquitin Ligase Using an Integrative Mass Spectrometry Strategy. *Anal. Chem.* **2019**, *91* (20), 12775–12783. https://doi.org/10.1021/acs.analchem.9b02331.

(84) Rusnac, D.-V.; Lin, H.-C.; Canzani, D.; Tien, K. X.; Hinds, T. R.; Tsue, A. F.; Bush, M. F.; Yen, H.-C. S.; Zheng, N. Recognition of the Diglycine C-End Degron by CRL2KLHDC2 Ubiquitin Ligase. *Molecular Cell* **2018**, *72* (5), 813-822.e4. https://doi.org/10.1016/j.molcel.2018.10.021.

(85) Canzani, D.; Laszlo, K. J.; Bush, M. F. Ion Mobility of Proteins in Nitrogen Gas: Effects of Charge State, Charge Distribution, and Structure. *J. Phys. Chem. A* **2018**, *122* (25), 5625–5634. https://doi.org/10.1021/acs.jpca.8b04474.

(86) Gabelica, V.; Shvartsburg, A. A.; Afonso, C.; Barran, P.; Benesch, J. L. P.; Bleiholder, C.; Bowers, M. T.; Bilbao, A.; Bush, M. F.; Campbell, J. L.; Campuzano, I. D. G.; Causon, T.; Clowers, B. H.; Creaser, C. S.; De Pauw, E.; Far, J.; Fernandez-Lima, F.; Fjeldsted, J. C.; Giles, K.; Groessl, M.; Hogan Jr, C. J.; Hann, S.; Kim, H. I.; Kurulugama, R. T.; May, J. C.; McLean, J. A.; Pagel, K.; Richardson, K.; Ridgeway, M. E.; Rosu, F.; Sobott, F.; Thalassinos, K.; Valentine, S. J.; Wyttenbach, T. Recommendations for Reporting Ion Mobility Mass Spectrometry Measurements. *Mass Spectrometry Reviews* **2019**, *38* (3), 291–320. https://doi.org/10.1002/mas.21585.

Chapter 2. Native Anion-Exchange Chromatography: Rapid Sample Clean-Up and Fractionation for Native Mass Spectrometry

Canzani, D.; Chen, C.; Kozemchak, C.; Rathod, P.K.; Bush, M.F. *In Preparation.*

2.1. Abstract

Despite the success of mass spectrometry (MS) in answering a wide range of questions related to protein structure, interactions, and quality, there are several common scenarios that make sample preparation challenging and/or decrease the quality of the resulting native MS data: (1) the presence of solution additives or contaminants that decrease the quality of native mass spectra, (2) various copies of a protein with different post-translational modifications and/or cofactors that cannot be resolved simultaneously by native MS, or (3) mixtures of proteins that yield highly congested mass spectra. Various methods have been developed to address those challenges, but additional complementary tools are needed. This work demonstrates that native anion-exchange chromatography (AEX) is a rapid and facile tool for overcoming each of those common scenarios. Native AEX uses a low-pressure, bioinert workflow that is based on a polymeric quaternary amine stationary phase and aqueous ammonium acetate mobile phases. Native AEX yields clean native-like ions and requires less time and materials than typical native MS sample preparation workflows. We anticipate that many native MS experiments would benefit significantly from using native AEX to replace one or more steps in their sample preparation workflow. Therefore, this strategy offers great potential for benefiting native MS, including applications to biophysics and the characterization of biotherapeutics.

2.2. Introduction

The detailed characterization of proteins and protein complexes by native mass spectrometry (MS) has become essential for the fields of structural and molecular biology.[1,2]

Native MS is widely used for determining the composition and stoichiometry of proteins and protein complexes,[3–6] especially when multiple forms of a complex exist simultaneously. Native MS is also becoming increasingly used for the identification of protein cofactors,[7,8] substrates,[9,10] and proteoforms,[7,11,12] and to unravel the complexity of glycosylated proteins.[13,14]

Most samples for native MS are first prepared using standard protein purification strategies, which often are time-intensive, require multiple steps, and make use of specialized solutions that include components (*e.g.*, nonvolatile salts, buffering agents, reducing agents, and detergents) that negatively impact the performance of native MS experiments.[15–17] For a successful native MS experiment, the proteins in a sample are usually then exchanged from the solution used for purification into aqueous ammonium acetate,[18] which increases the time to analysis and is not always straightforward. Various methods have been developed to make protein samples compatible with native MS and reduce the preparation time for analysis. For example, solution additives were shown to reduce the extent of sodium adduction during electrospray ionization,[19,20] but those additives might not be appropriate for all native MS experiments, especially if they have supercharging effects or disrupt non-covalent interactions. Another advancement was electrospray emitters that have sub-micron diameter openings, which reduce salt adduction and interferences from buffering agents,[21,22] but can also result in surface-induced unfolding of the protein.[23] Perhaps one of the easiest to implement methods is buffer-loading, in which a protein sample is diluted into concentrated (1-7 M) ammonium acetate, but this requires a high concentration of starting material[24] and the stability of many noncovalent interactions depend on ionic strength.[3]

Although the methods mentioned above are useful in certain situations, they will not overcome all challenges associated with native MS analysis. In addition, researchers are

interested in improving automation during sample preparation, so more general methods are needed. Due to those considerations, chromatography is now being adopted for native MS analysis. For example, size-exclusion chromatography (SEC) with native MS has been used for various purposes, including characterizing protein aggregates and oligomeric state,[25] determining the effects of small molecules on protein quaternary structure,[26] and enabling therapeutic antibody analysis.[27] Online buffer exchange (OBE) using miniature SEC devices have been used to rapidly screen recombinant proteins and guide protein expression and purification.[28] These SEC methods are useful in a wide variety of applications, but lack the resolution to separate proteoforms and can dilute samples. Cation-exchange chromatography has been used to separate proteins based on natural charge variants, including therapeutic monoclonal antibodies.[29–31] However, cation-exchange is limited to separating proteins that are positively charged in solution, so it will not be appropriate for all native MS applications. The limitations of SEC and cation exchange highlight the need for complementary tools.

Here, we demonstrate that native anion-exchange chromatography (AEX) is an effective method for rapid clean-up and fractionation of proteins. In this study, native AEX refers to separations performed using a polymeric, quaternary amine stationary phase, aqueous ammonium acetate mobile phases, modest pressures (~500 PSI), and a bioinert workflow (e.g., no metal or silica). Although AEX has been used to separate proteins for use with MS, those methods used conventional HPLC systems[32,33] and mobile phases that were modified using formic acid.[33] We demonstrate that selected salts, detergents, and common additive can be removed from protein samples using native AEX. The ability to reduce sample heterogeneity with native AEX protein separations will enhance proteoform and protein complex identification with native MS. Therefore, we anticipate that this strategy will be highly valuable for a wide

range of workflows that use native MS to answer questions in biophysics and to characterize biotherapeutics.

2.3. Methods

2.3.1. Materials. Bovine serum albumin (BSA), human apo-transferrin, and ammonium acetate were purchased from Sigma-Aldrich (St. Louis, MO)**.** Chicken egg whites were separated from yolk and 10 mL of egg white was added to 30 mL of deionized H_2O and then adjusted to pH 6. The sample was left at 4 ºC overnight on a rotator to precipitate ovomucin and decrease the viscosity of the sample.[34] 20 uL of that sample was added to 980 uL of aqueous 100 mM ammonium acetate, pH 9.7, and centrifuged at 20,000 *g* for 10 minutes. *Escherichia coli* cells that overexpressed superfolder green fluorescent protein with a with a hexa histidine tag (sfGFP-His6, Addgene plasmid #85942)[35] were grown overnight in Luria Broth at 37 °C. The cells were spun down and the growth media was removed. 150 mg of the cell pellet was resuspended in 1000 µL of aqueous 50 mM ammonium acetate and lysed on ice using a Branson Sonifier 200, then centrifuged at 20,000 *g* for 10 minutes at 4 ºC. Immediately after centrifugation, 50 µL of the cleared lysate was loaded onto the AEX column. Additional details about the molecular biology and affinity purification of sfGFP-His6 protein can be found in the Supporting Information.

2.3.2. Native Anion-Exchange Chromatography. A BioRad (Hercules, CA) BioLogic DuoFlow 10 fast protein liquid chromatography (FPLC) system equipped with a UV detector (280 nm) was used to achieve separations on a BioRad Continuous Bed Uno Q-1 anion-exchange column (7 mm x 35 mm). This system was operated using gradients of either ionic strength or pH. The ionic strength gradient consisted of solution A: deionized H_2O and solution B: aqueous 1 M ammonium acetate, both of which were adjusted to pH 8 with ammonium hydroxide. Linear

gradients were run at 0.6 mL min^{-1} from 5 to 95% solvent B. The pH mode of operation consisted of solvent A: 100 mM ammonium acetate, pH 9.7 and solvent B: 100 mM ammonium acetate, pH 4.5. 15 or 20-minute linear gradients were run at 0.6 mL min^{-1} from 0 to 100% solvent B. Fractions of 0.5 mL were collected over the entire run for all methods. All native AEX separations were performed at 4 ºC.

2.3.3. Native Mass Spectrometry. The fractions collected during native AEX were analyzed individually. 3 µL of a single fraction was added into the tip of a pulled borosilicate glass capillary. A platinum wire electrode was inserted into the capillary to achieve contact with the sample. The sample was ionized by applying a potential of between 0.5 to 1.0 kV to the platinum wire electrode. The resulting ions were analyzed using a hybrid electrospray/quadrupole/ion-mobility/time-of-flight mass spectrometer (Waters Synapt G2 HDMS, Milford, MA). A 30 mg mL^{-1} solution of aqueous CsI was also analyzed and the resulting mass spectrum was used for external calibration of all native mass spectra.

2.4. Results and Discussion

Despite the success of native MS in answering a wide range of questions related to protein structure, interactions, and quality,[3,7] there are several common scenarios that make sample preparation challenging and/or decrease the quality of the resulting native MS data: the sample includes (1) molecules and/or ions, which were often added intentionally during early stages of purification, that decrease the quality of native mass spectra, (2) various copies of a protein with different post-translational modifications and/or interactions with cofactors that are not resolved simultaneously by native MS, or (3) mixtures of proteins that yield highly congested mass spectra. In this study, we will use separations performed using a polymeric, quaternary amine stationary phase, aqueous ammonium acetate mobile phases, modest pressures (~500 PSI),

and a bioinert workflow (e.g., no metal or silica), which we will refer to as native AEX, to demonstrate the rapid and facile method for overcoming each of those common scenarios.

2.4.1. Rapid Sample Clean-up using Native AEX. Various molecules and ions are often introduced throughout the process of lysing cells and purifying proteins. Detergents are used frequently to enhance protein solubility and stability during purification protocols,[15,16] but many of those detergents are incompatible with MS, even at low concentrations.[36] For example, Figure 1A shows a spectrum of bovine serum albumin (BSA) in a solution containing 0.1% NP-40, a common detergent used in many affinity-purification schemes. This spectrum exhibits intense signals for ions of detergent molecules at low-*m*/*z* values and very broad features for the protein ions at high-*m*/*z* values. Spectra measured for BSA in solutions containing either 0.05% Triton X-100 (Figure A1A) or 0.05% Tween 20 (Figure A1B) exhibited intense signals for ions of detergent molecules and negligible signals for protein ions. Even low concentrations of detergents can strongly interfere with and limit the interpretation of the protein components of samples.

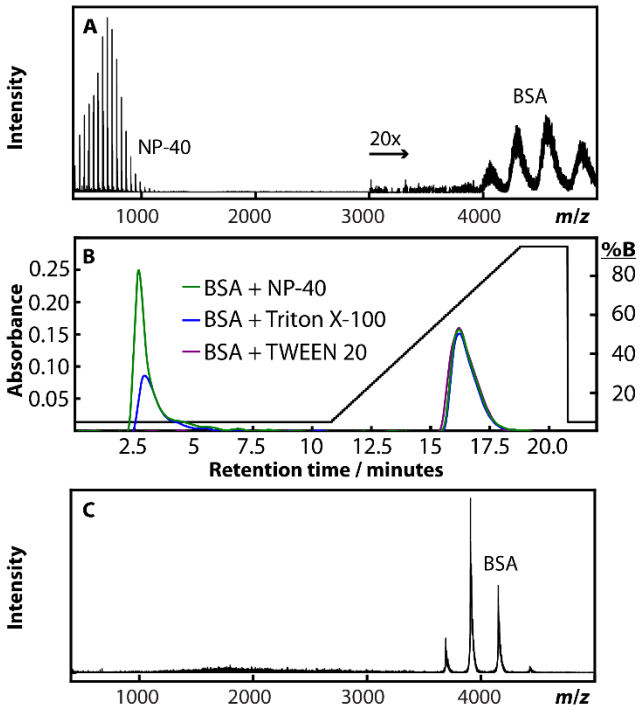

Figure 2.1. (A) Native mass spectrum of a BSA in an aqueous solution of 0.1% NP-40. Strong interference from NP-40 limited the detection of BSA, and the region above 3000 *m*/*z* is magnified 20-fold to aid in visualization. (B) Native AEX chromatograms of samples containing BSA and one of three common non-ionic detergents: 0.1% NP-40, 0.05% Triton X-100, or 0.05% Tween 20. Solution A was aqueous 50 mM ammonium acetate and solution B was 950 mM ammonium acetate. (C) Native mass spectrum of a fraction collected near 16 minutes during the native AEX separation shown in panel B. This spectrum exhibits no interference from the NP-40 detergent.

Unfortunately, most detergents are difficult to remove from protein samples; NP-40 and Triton X-100 are non-dialyzable and can be difficult to separate from proteins by size-exclusion or molecular-weight cut-off methods due to the large micelles they form in solution (between 60

and 90 kDa).[37] There are single-use devices that can remove detergents but these can be expensive,[38] and other methods can denature proteins and are therefore unsuitable for native MS.[39–41] MS-compatible detergents have been developed, but can form non-specific adducts with proteins during ESI,[42] which can complicate or prevent complete interpretation of native mass spectra. Some detergents can be removed using high extents of collisional activation in the gas phase, but that activation could also cause protein fragmentation or unfolding.[43]

Native AEX is well suited for removing non-ionic or cationic detergents, since they are not retained on the positively charged stationary phase.[37,44] Additionally, many anion-exchangers are compatible with aqueous ammonium acetate, which is the primary solution that is used for native MS. Figure 1B shows the native AEX chromatograms of BSA mixed with 0.1% NP-40, 0.05% Triton X-100, or 0.05% Tween 20. The non-ionic detergents eluted during the wash step at the beginning of the separation, whereas BSA was retained on the stationary phase and was then eluted using a 10-minute gradient from 50 to 950 mM ammonium acetate. Figure 1C, shows the native mass spectrum of the eluted fraction that contained BSA (collected at approximately minute 16 of the separation), which demonstrate that NP-40 is effectively removed from the protein sample. Triton X-100 and Tween 20 were also effectively removed from the BSA sample using native AEX, and the corresponding mass spectra are shown in Figures S1C and S1D.

Other molecules are often used in protein purification to control the properties of solutions (*e.g.*, ionic strength and pH of solutions) and improve the stability of proteins during storage; most molecules used in molecular biology workflows are nonvolatile and will decrease the quality of native mass spectra. For example, Figure A2A shows the native mass spectrum of a sample of α-lactalbumin prepared in solution of aqueous 50 mM NaCl; the ions observed in this spectrum exhibit a very high degree of sodium adduction, with up to 28 sodium ions bound

to the 7+ ion (Figure A2A). In contrast, subjecting that sample to native AEX prior to native MS results in spectrum that exhibits a very low extent of sodium adduction (Figure A2B), which enabled the identification of calcium-bound α-lactalbumin in that spectrum. Figure A3 shows a series of native mass spectra measured for BSA prepared in solutions with either 5% dimethyl sulfoxide, 50 mM imidazole, or 5% glycerol. All three samples characterized without native AEX yielded poor quality native mass spectra, with evidence of protein unfolding, poor signals for protein ions relative to non-protein ions, and/or wide peaks for protein ions. In contrast, all three samples characterized after native AEX yielded similar, high-quality native mass spectra. In addition to rapid and effective detergent removal, these results demonstrate that native AEX is highly effective for de-salting and removal of other nonvolatile components of solutions. Overall, these results demonstrate a robust and fast (less than 25-minute chromatography run) method to prepare proteins for native MS from solutions that contain salts, detergents, and other common additives.

2.4.2. Fractionating Proteoforms. Transferrin is the major iron-binding protein of blood plasma and has been shown to be modified by glycans containing variable numbers of sialic acid groups in the form of *N*-acetylneuraminic acid.[45] Transferrin was separated using native AEX with a linear pH gradient (Figure 2A), and six fractions were collected and analyzed with native MS (Figure 2B). The native mass spectra for the early fractions exhibited features assigned to transferrin with 2, 3, or 4 sialic acid groups, whereas those for the later fractions exhibited features assigned to transferrin with 5 or 6 sialic acid groups. That elution order is consistent with the expected order based on the decrease of pI of the proteoforms with increasing numbers of sialic acids.[46]

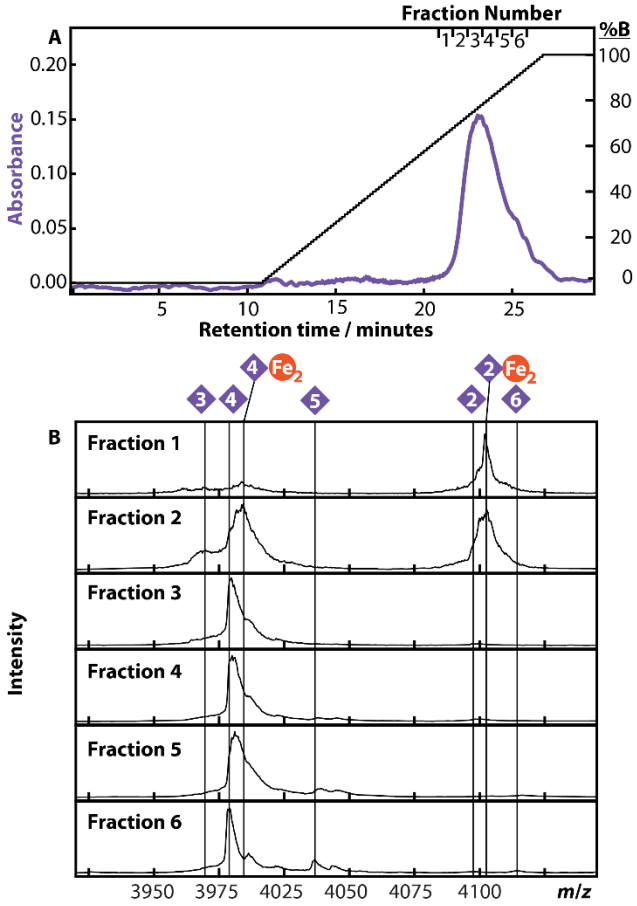

Figure 2.2. (A) Native AEX chromatogram of transferrin separated using a linear, pH gradient of aqueous 100 mM ammonium acetate from pH 9.7 (solution A) to 4.5 (solution B). (B) Six fractions were collected and analyzed using native MS. This region exhibits features that are assigned to 19+ ions with 2 sialic acid groups and 20+ ions with 3 to 6 sialic acids groups (the number of groups is indicated in purple diamonds; the structures of the glycans are shown in Figure A5). These assignments indicate that the retention time of these proteoforms increases with the number of sialic acid groups, consistent with the effect of the number of sialic acid groups on the isoelectric point. Di-ferric, glycosylated holo-transferrin was also observed in the early fractions (orange circles).

Although the material used for these experiments was sold as apo-transferrin, the first two fractions yielded mass spectra with peaks assigned to di-ferric holo-transferrin ions ($[M + 2Fe^{3+} + (z\text{-}6)H]^{z+}$) in addition to having 2 or 4 sialic acid groups. Holo-transferrin was not observed in the native mass spectrum of the sample acquired prior to native AEX (Figure A4), but it may have been present at low abundance. To confirm the assignment of the holo-transferrin peaks in fractions 1 and 2 of the native AEX experiments, the material sold as apo-transferrin was suspended in a solution containing 150 mM ammonium acetate, 50 mM ammonium bicarbonate, and 0.5 mM $FeCl_3$ (Figure A4). The observed *m*/*z* values corroborate the identities of $[M+19H]^{19+}$ and $[M + 2Fe^{3+} + 13H]^{19+}$ ions.

In addition to the rapid separation of transferrin glycoforms, the observation of holo-transferrin in the apo sample demonstrates the value of fractionation to enhance cofactor identification with native MS. For comparison, an ionic-strength gradient from 50 to 950 mM ammonium acetate was also used to separate transferrin (Figure A5) but was not as effective in separating the holo- and sialoforms as the 15-minute, pH gradient.

2.4.4. Separation of Proteins from Egg White. Egg whites are a complex matrix with a high concentration of proteins. Despite the presence of many high-abundance proteins, "buffer loading"[24] analysis enabled by 20-fold dilution of egg white into aqueous 1 M ammonium acetate (Figure A6A) only resulted in the identification of lysozyme; some broad, unidentified features were also observed at higher *m*/*z* values. To demonstrate the capabilities of native AEX to rapidly prepare proteins for native MS, egg whites were separated on a 20-minute, linear gradient of 100 mM ammonium acetate from pH 9.7 to 4.5 (Figure 3A). Four distinct peaks were observed in the chromatogram (Figure A4B), which are assigned to lysozyme, ovotransferrin,

ovomucoid, and ovalbumin, in order of increasing retention time. These assignments are consistent with the decreasing isoelectric points of those proteins. Although egg whites contain a relatively high electrolyte concentration (including sodium, potassium, and other ions), the proteins were effectively desalted and various glycoforms of ovotransferrin (Figure 3B) and ovalbumin (Figure 3C and S7) were observed in the native mass spectra. The predominate features observed for ovotransferrin are consistent with modifications by either $GlcNAc_6Man_3$ or $GlcNAc_5Man_3$; these were identified previously as the primary glycoforms of ovotransferrin.[47,48] Although HPLC using an AEX stationary phase has been used to separate glycoforms present in an commercially prepared sample of ovalbumin,[33] the present experiments demonstrate the potential for similar molecular characterization using a bioinert workflow and minimal preparation of the original biological source. Lysozyme eluted during the loading of the sample, consistent with the high pI of that protein and a positive charge under the initial conditions of the mobile phase (Figure A4B). These results indicate that native AEX is most suitable for preparing and separating negatively charged proteins, consistent with the mechanism of the separation.

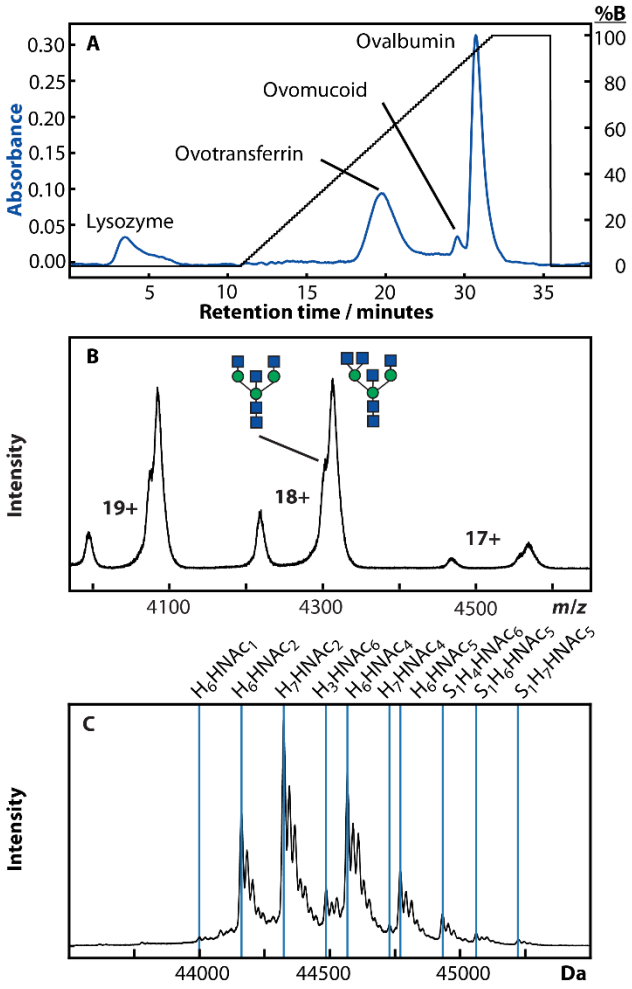

Figure 2.3. (A) Native AEX chromatogram of egg white using a 20-minute, linear gradient of aqueous 100 mM ammonium acetate from pH 9.7 (solution A) to 4.5 (solution B). (B) Native mass spectrum of ovotransferrin isolated from the separation in panel A. Unmodified transferrin and two probable glycoforms, $GlcNAc_6Man_3$ and $GlcNAc_5Man_3$ were identified. The glycoforms are labeled based on assignments reported previously.[48] (C) Zero-charge mass spectrum of ovalbumin ; the corresponding *m*/*z* spectrum is shown in Figure A7. Various glycoforms were assigned based on the masses of those identified in previous studies.[13,33] Abbreviations used: Hexose (H), N-acetyl hexose (HNAc), Sialic acid (S).

2.4.5. Separation of an Overexpressed Protein from a Cell Lysate. Since most protein purification schemes are time-intensive and require specialized buffers, efforts have been made to bypass affinity purification and bring the protein directly from cell culture into a native MS-compatible solution.[49–51] Those methods are promising, but can require a protein to be secreted and might not sufficiently clear all cell debris or solution components that are present in the in cell lystate.

To evaluate the use of native AEX for the analysis of samples without any previous purification, we used native AEX to analyze a lysate from *E. coli* that overexpressed superfolder green fluorescent protein with a hexa-histidine tag (sfGFP-His6). The cells were homogenized in aqueous 50mM ammonium acetate using an ultrasonic device and the resulting lysate was cleared using centrifugation; no affinity purification was performed. Figure 4A shows a native AEX chromatogram of the lysate. The bound fraction was eluted with an ionic strength gradient from 50 to 950 mM ammonium acetate. A fraction collected around 12 minutes yielded a native mass spectrum that exhibits peaks that are consistent with the overexpressed protein (Figure 4B). This assignment was then corroborated through native MS (Figure 4C) of the same native-AEX fraction number collected for a sample that was affinity purified (see Supporting Information) prior to native AEX. Although there are features in Figure 4B that are attributed to unidentified lysate components below 1500 *m*/*z* and other unidentified *E. coli* proteins, sfGFP-His6 was clearly distinguishable. These results demonstrate the potential of using native AEX to analyze complex mixtures without affinity purification, which saves time and material. Only a small quantity of cell lysate was loaded onto the column (derived from 7.5 mg of cell pellet; a liter of *E. coli* culture grown to an OD_{600} of 1 typically yields a ~3 g pellet) and only 3 uL of each 0.5 mL fraction from the separation was used for native MS, so this method is applicable to

screening small-scale cell cultures. More generally, these results illustrate the potential for using native MS to guide protein purification without the use of affinity tags, which can both remove steps and completely remove unintended complications associated with introducing affinity tags using molecular biology, including applications that require high throughput and/or automation.

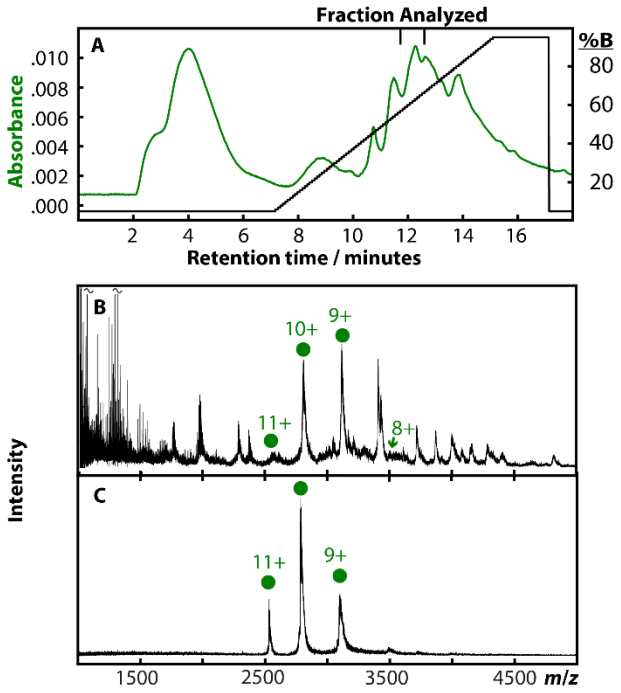

Figure 2.4. (A) *E.coli* cells that overexpressed sfGFP-His6 were lysed in 50 mM ammonium acetate and cleared by centrifugation. The cleared lysate was loaded onto an AEX column and eluted using a 10-minute, linear gradient from aqueous 50 (solution A) to 950 (solution B) mM ammonium acetate. (B) A native mass spectrum of a fraction that was collected around 12 minutes from the chromatogram in panel A. Peaks assigned to sfGFP-His6 (green circles) are the predominant protein signal, even though other lysate components were present in the sample. (C) A native mass spectrum of the same fraction number acquired for a

sample that was affinity purified prior to native AEX, which corroborate the assignments in panel B.

2.5. Conclusions

Native AEX using aqueous ammonium acetate mobile phases is a rapid and facile tool for preparing proteins for native MS and overcoming several common challenges that are associated with native MS experiments. Native AEX: (1) performs well as a sample preparation tool by simultaneously exchanging a protein into an MS-compatible solution and eliminating detergents (Figures 1 and S1), salts (Figure A2), and other common solution additives from a protein sample (Figure A3), (2) can improve the identification of post-translational modifications and protein co-factors, especially when mass spectrometer resolving power isn't sufficient on its own (Figures 2, S2B, and S4), and (3) can enhance the identification of proteins that are in complex mixtures or dense matrices (Figures 3 and S7), including cell lysates (Figure 4).

These experiments suggest that many native MS workflows would benefit from incorporating native AEX either as an additional step or as a replacement for earlier purification steps that would introduce molecules or ions that require removal prior to MS. Native AEX is a valuable tool that can enable successful native MS preparation, separation, and analysis. An intrinsic limitation of AEX is that it is limited to separating proteins that are negatively charged, so protein pI must be taken into consideration for each sample and in the selection of mobile phases. Solution pH or ionic strength can affect the stability of proteins and protein complexes, so the use of pH or ionic strength gradients should also be considered for each protein in order to maintain native-like conditions during the AEX separation. In these experiments, pH gradients were used from 9.7 to 4.5, but the range of pH could be narrowed for specific proteins that have instability at high or low pH. Native AEX has limited utility for removing solution additives or

contaminants that are anionic since they will be retained by the column. The peak capacities of native AEX separations are modest compared to some other modes of separation, but this level of separation is adequate for most native MS applications. Overall, the selectivity that native AEX provides is complementary to forms of chromatography, like size-exclusion or cation-exchange, and that selectivity makes it a valuable option for native MS applications. In the future, we will explore a wider range of column chemistries, and combinations of those column chemistries, to provide a broader selection of tools that can meet the needs of any native MS application.

2.6. Supporting Information. Detail about the molecular biology and protein purification, and Figures A1 to A7 can be found in Appendix A.

2.7. Acknowledgements. This research was supported by the National Science Foundation under award number CHE-1807382 (M. F. B.); the National Institutes of Health under award numbers AI093380 (P.K.R.), AI089688 (P.K.R.), and T32GM008268 (D. C.); and the ACS Division of Analytical Chemistry and Agilent Technologies (fellowship to D.C.).

2.8. References

(1) Robinson, C. V.; Sali, A.; Baumeister, W. The Molecular Sociology of the Cell. *Nature* **2007**, *450* (7172), 973–982. https://doi.org/10.1038/nature06523.

(2) Heck, A. J. R. Native Mass Spectrometry: A Bridge between Interactomics and Structural Biology. *Nat. Methods* **2008**, *5*, 927–933. https://doi.org/10.1038/nmeth.1265.

(3) Zhou, M.; Sandercock, A. M.; Fraser, C. S.; Ridlova, G.; Stephens, E.; Schenauer, M. R.; Yokoi-Fong, T.; Barsky, D.; Leary, J. A.; Hershey, J. W.; Doudna, J. A.; Robinson, C. V.

Mass Spectrometry Reveals Modularity and a Complete Subunit Interaction Map of the Eukaryotic Translation Factor EIF3. *Proc. Natl. Acad. Sci. USA* **2008**, *105* (47), 18139–18144. https://doi.org/10.1073/pnas.0801313105.

(4) Stengel, F.; Baldwin, A. J.; Bush, M. F.; Hilton, G. R.; Lioe, H.; Basha, E.; Jaya, N.; Vierling, E.; Benesch, J. L. P. Dissecting Heterogeneous Molecular Chaperone Complexes Using a Mass Spectrum Deconvolution Approach. *Chem. Biol.* **2012**, *19* (5), 599–607. https://doi.org/10.1016/j.chembiol.2012.04.007.

(5) Snijder, J.; van de Waterbeemd, M.; Damoc, E.; Denisov, E.; Grinfeld, D.; Bennett, A.; Agbandje-McKenna, M.; Makarov, A.; Heck, A. J. R. Defining the Stoichiometry and Cargo Load of Viral and Bacterial Nanoparticles by Orbitrap Mass Spectrometry. *J. Am. Chem. Soc.* **2014**, *136* (20), 7295–7299. https://doi.org/10.1021/ja502616y.

(6) Lu, C.; Turley, S.; Marionni, S. T.; Park, Y.-J.; Lee, K. K.; Patrick, M.; Shah, R.; Sandkvist, M.; Bush, M. F.; Hol, W. G. J. Hexamers of the Type II Secretion ATPase GspE from Vibrio Cholerae with Increased ATPase Activity. *Structure* **2013**, *21* (9), 1707–1717. https://doi.org/10.1016/j.str.2013.06.027.

(7) Skinner, O. S.; Haverland, N. A.; Fornelli, L.; Melani, R. D.; Do Vale, L. H. F.; Seckler, H. S.; Doubleday, P. F.; Schachner, L. F.; Srzentić, K.; Kelleher, N. L.; Compton, P. D. Top-down Characterization of Endogenous Protein Complexes with Native Proteomics. *Nat. Chem. Biol.* **2018**, *14*, 36–41. https://doi.org/10.1038/nchembio.2515.

(8) Wang, H.; Shi, H.; Rajan, M.; Canarie, E. R.; Hong, S.; Simoneschi, D.; Pagano, M.; Bush, M. F.; Stoll, S.; Leibold, E. A.; Zheng, N. FBXL5 Regulates IRP2 Stability in Iron Homeostasis via an Oxygen-Responsive [2Fe2S] Cluster. *Mol. Cell* **2020**, *78* (1), 31-41.e5. https://doi.org/10.1016/j.molcel.2020.02.011.

(9) Canzani, D.; Rusnac, D.-V.; Zheng, N.; Bush, M. F. Degronomics: Mapping the Interacting Peptidome of a Ubiquitin Ligase Using an Integrative Mass Spectrometry Strategy. *Anal. Chem.* **2019**, *91* (20), 12775–12783. https://doi.org/10.1021/acs.analchem.9b02331.

(10) Ma, C.; Hu, Y.; Townsend, J. A.; Lagarias, P. I.; Marty, M. T.; Kolocouris, A.; Wang, J. Ebselen, Disulfiram, Carmofur, PX-12, Tideglusib, and Shikonin Are Nonspecific Promiscuous SARS-CoV-2 Main Protease Inhibitors. *ACS Pharmacol. Transl. Sci.* **2020**, *3* (6), 1265–1277. https://doi.org/10.1021/acsptsci.0c00130.

(11) Smith, L. M.; Kelleher, N. L.; Linial, M.; Goodlett, D.; Langridge-Smith, P.; Ah Goo, Y.; Safford, G.; Bonilla, L.; Kruppa, G.; Zubarev, R.; Rontree, J.; Chamot-Rooke, J.; Garavelli, J.; Heck, A.; Loo, J.; Penque, D.; Hornshaw, M.; Hendrickson, C.; Pasa-Tolic, L.; Borchers, C.; Chan, D.; Young, N.; Agar, J.; Masselon, C.; Gross, M.; McLafferty, F.; Tsybin, Y.; Ge, Y.; Sanders, I.; Langridge, J.; Whitelegge, J.; Marshall, A.; The Consortium for Top Down Proteomics. Proteoform: A Single Term Describing Protein Complexity. *Nat. Methods* **2013**, *10* (3), 186–187. https://doi.org/10.1038/nmeth.2369.

(12) 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. https://doi.org/10.1038/nchem.2908.

(13) Yang, Y.; Barendregt, A.; Kamerling, J. P.; Heck, A. J. R. Analyzing Protein Micro-Heterogeneity in Chicken Ovalbumin by High-Resolution Native Mass Spectrometry Exposes Qualitatively and Semi-Quantitatively 59 Proteoforms. *Anal. Chem.* **2013**, *85* (24), 12037–12045. https://doi.org/10.1021/ac403057y.

(14) 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.* **2018**, *115* (35), 8763–8768. https://doi.org/10.1073/pnas.1807439115.

(15) Gerace, E.; Moazed, D. Affinity Purification of Protein Complexes Using TAP Tags. *Methods Enzymol.* **2015**, *559*, 37–52. https://doi.org/10.1016/bs.mie.2014.11.007.

(16) Spriestersbach, A.; Kubicek, J.; Schäfer, F.; Block, H.; Maertens, B. Chapter One - Purification of His-Tagged Proteins. In *Methods in Enzymology*; Lorsch, J. R., Ed.; Academic Press, 2015; Vol. 559, pp 1–15. https://doi.org/10.1016/bs.mie.2014.11.003.

(17) Smith, S. M. Strategies for the Purification of Membrane Proteins. *Methods Mol. Biol. Clifton NJ* **2017**, *1485*, 389–400. https://doi.org/10.1007/978-1-4939-6412-3_21.

(18) Hernández, H.; Robinson, C. V. Determining the Stoichiometry and Interactions of Macromolecular Assemblies from Mass Spectrometry. *Nat. Protoc.* **2007**, *2*, 715–726. https://doi.org/10.1038/nprot.2007.73.

(19) Cassou, C. A.; Williams, E. R. Desalting Protein Ions in Native Mass Spectrometry Using Supercharging Reagents. *Analyst* **2014**, *139* (19), 4810–4819. https://doi.org/10.1039/C4AN01085J.

(20) Flick, T. G.; Cassou, C. A.; Chang, T. M.; Williams, E. R. Solution Additives That Desalt Protein Ions in Native Mass Spectrometry. *Anal. Chem.* **2012**, *84* (17), 7511–7517. https://doi.org/10.1021/ac301629s.

(21) Susa, A. C.; Xia, Z.; Williams, E. R. Small Emitter Tips for Native Mass Spectrometry of Proteins and Protein Complexes from Nonvolatile Buffers That Mimic the Intracellular Environment. *Anal. Chem.* **2017**, *89* (5), 3116–3122. https://doi.org/10.1021/acs.analchem.6b04897.

(22) Nguyen, G. T. H.; Tran, T. N.; Podgorski, M. N.; Bell, S. G.; Supuran, C. T.; Donald, W. A. Nanoscale Ion Emitters in Native Mass Spectrometry for Measuring Ligand–Protein Binding Affinities. *ACS Cent. Sci.* **2019**, *5* (2), 308–318. https://doi.org/10.1021/acscentsci.8b00787.

(23) Mortensen, D. N.; Williams, E. R. Surface-Induced Protein Unfolding in Submicron Electrospray Emitters. *Anal. Chem.* **2016**, *88* (19), 9662–9668. https://doi.org/10.1021/acs.analchem.6b02499.

(24) Iavarone, A. T.; Udekwu, O. A.; Williams, E. R. Buffer Loading for Counteracting Metal Salt-Induced Signal Suppression in Electrospray Ionization. *Anal. Chem.* **2004**, *76* (14), 3944–3950. https://doi.org/10.1021/ac049724+.

(25) Muneeruddin, K.; Thomas, J. J.; Salinas, P. A.; Kaltashov, I. A. Characterization of Small Protein Aggregates and Oligomers Using Size Exclusion Chromatography with Online Detection by Native Electrospray Ionization Mass Spectrometry. *Anal. Chem.* **2014**, *86* (21), 10692–10699. https://doi.org/10.1021/ac502590h.

(26) Shen, M. L.; Benson, L. M.; Johnson, K. L.; Lipsky, J. J.; Naylor, S. Effect of Enzyme Inhibitors on Protein Quaternary Structure Determined by On-Line Size Exclusion Chromatography-Microelectrospray Ionization Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **2001**, *12* (1), 97–104. https://doi.org/10.1016/S1044-0305(00)00190-2.

(27) Ehkirch, A.; Hernandez-Alba, O.; Colas, O.; Beck, A.; Guillarme, D.; Cianférani, S. Hyphenation of Size Exclusion Chromatography to Native Ion Mobility Mass Spectrometry for the Analytical Characterization of Therapeutic Antibodies and Related Products. *J. Chromatogr. B* **2018**, *1086*, 176–183. https://doi.org/10.1016/j.jchromb.2018.04.010.

(28) VanAernum, Z. L.; Busch, F.; Jones, B. J.; Jia, M.; Chen, Z.; Boyken, S. E.; Sahasrabuddhe, A.; Baker, D.; Wysocki, V. H. Rapid Online Buffer Exchange for Screening of Proteins, Protein Complexes and Cell Lysates by Native Mass Spectrometry. *Nat. Protoc.* **2020**, *15* (3), 1132–1157. https://doi.org/10.1038/s41596-019-0281-0.

(29) Yan, Y.; Liu, A. P.; Wang, S.; Daly, T. J.; Li, N. Ultrasensitive Characterization of Charge Heterogeneity of Therapeutic Monoclonal Antibodies Using Strong Cation Exchange Chromatography Coupled to Native Mass Spectrometry. *Anal. Chem.* **2018**, *90* (21), 13013–13020. https://doi.org/10.1021/acs.analchem.8b03773.

(30) Ma, F.; Raoufi, F.; Bailly, M. A.; Fayadat-Dilman, L.; Tomazela, D. Hyphenation of Strong Cation Exchange Chromatography to Native Mass Spectrometry for High Throughput Online Characterization of Charge Heterogeneity of Therapeutic Monoclonal Antibodies. *mAbs* **2020**, *12* (1), e1763762. https://doi.org/10.1080/19420862.2020.1763762.

(31) Matsuda, Y.; Kliman, M.; Mendelsohn, B. A. Application of Native Ion Exchange Mass Spectrometry to Intact and Subunit Analysis of Site-Specific Antibody–Drug Conjugates Produced by AJICAP First Generation Technology. *J. Am. Soc. Mass Spectrom.* **2020**, *31* (8), 1706–1712. https://doi.org/10.1021/jasms.0c00129.

(32) Leblanc, Y.; Bihoreau, N.; Chevreux, G. Characterization of Human Serum Albumin Isoforms by Ion Exchange Chromatography Coupled On-Line to Native Mass Spectrometry. *J. Chromatogr. B* **2018**, *1095*, 87–93. https://doi.org/10.1016/j.jchromb.2018.07.014.

(33) Füssl, F.; Criscuolo, A.; Cook, K.; Scheffler, K.; Bones, J. Cracking Proteoform Complexity of Ovalbumin with Anion-Exchange Chromatography–High-Resolution Mass

Spectrometry under Native Conditions. *J. Proteome Res.* **2019**, *18* (10), 3689–3702. https://doi.org/10.1021/acs.jproteome.9b00375.

(34) Guérin-Dubiard, C.; Pasco, M.; Hietanen, A.; Quiros del Bosque, A.; Nau, F.; Croguennec, T. Hen Egg White Fractionation by Ion-Exchange Chromatography. *J. Chromatogr. A* **2005**, *1090* (1), 58–67. https://doi.org/10.1016/j.chroma.2005.06.083.

(35) Peeler, J. C.; Mehl, R. A. Site-Specific Incorporation of Unnatural Amino Acids as Probes for Protein Conformational Changes. In *Unnatural Amino Acids: Methods and Protocols*; Pollegioni, L., Servi, S., Eds.; Humana Press: Totowa, NJ, 2012; pp 125–134. https://doi.org/10.1007/978-1-61779-331-8_8.

(36) Ogorzalek Loo, R. R.; Dales, N.; Andrews, P. C. The Effect of Detergents on Proteins Analyzed by Electrospray Ionization. In *Protein and Peptide Analysis by Mass Spectrometry*; Chapman, J. R., Ed.; Humana Press: Totowa, NJ, 1996; pp 141–160. https://doi.org/10.1385/0-89603-345-7:141.

(37) Linke, D. Chapter 34 Detergents: An Overview. In *Methods in Enzymology*; Burgess, R. R., Deutscher, M. P., Eds.; Academic Press, 2009; Vol. 463, pp 603–617. https://doi.org/10.1016/S0076-6879(09)63034-2.

(38) Antharavally, B. S.; Mallia, K. A.; Rosenblatt, M. M.; Salunkhe, A. M.; Rogers, J. C.; Haney, P.; Haghdoost, N. Efficient Removal of Detergents from Proteins and Peptides in a Spin Column Format. *Anal. Biochem.* **2011**, *416* (1), 39–44. https://doi.org/10.1016/j.ab.2011.05.013.

(39) Yeung, Y.-G.; Nieves, E.; Angeletti, R. H.; Stanley, E. R. Removal of Detergents from Protein Digests for Mass Spectrometry Analysis. *Anal. Biochem.* **2008**, *382* (2), 135–137. https://doi.org/10.1016/j.ab.2008.07.034.

(40) Yeung, Y.-G.; Stanley, E. R. Rapid Detergent Removal from Peptide Samples with Ethyl Acetate for Mass Spectrometry Analysis. *Curr. Protoc. Protein Sci.* **2010**, *59* (1), 16.12.1-16.12.5. https://doi.org/10.1002/0471140864.ps1612s59.

(41) Rey, M.; Mrázek, H.; Pompach, P.; Novák, P.; Pelosi, L.; Brandolin, G.; Forest, E.; Havlíček, V.; Man, P. Effective Removal of Nonionic Detergents in Protein Mass Spectrometry, Hydrogen/Deuterium Exchange, and Proteomics. *Anal. Chem.* **2010**, *82* (12), 5107–5116. https://doi.org/10.1021/ac100171m.

(42) Urner, L. H.; Maier, Y. B.; Haag, R.; Pagel, K. Exploring the Potential of Dendritic Oligoglycerol Detergents for Protein Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **2019**, *30* (1), 174–180. https://doi.org/10.1007/s13361-018-2063-2.

(43) Borysik, A. J.; Robinson, C. V. The 'Sticky Business' of Cleaning Gas-Phase Membrane Proteins: A Detergent Oriented Perspective. *Phys. Chem. Chem. Phys.* **2012**, *14* (42), 14439–14449. https://doi.org/10.1039/C2CP41687E.

(44) Gebauer, B.; Šimić, M.; Branović, K.; Treščec, A.; Benko, B. Ion-Exchange Chromatography Separation of the Detergent and the Solvent from Immunoglobulins after Solvent–Detergent Treatment. *J. Chromatogr. A* **1999**, *852* (1), 83–86. https://doi.org/10.1016/S0021-9673(99)00230-7.

(45) Wada, Y. Mass Spectrometry of Transferrin Glycoforms to Detect Congenital Disorders of Glycosylation: Site-Specific Profiles and Pitfalls. *Proteomics* **2016**, *16* (24), 3105–3110. https://doi.org/10.1002/pmic.201500551.

(46) Arizaga Rodríguez, S.; Blanco González, E.; Alvarez Llamas, G.; Montes-Bayón, M.; Sanz-Medel, A. Detection of Transferrin Isoforms in Human Serum: Comparison of UV

and ICP–MS Detection after CZE and HPLC Separations. *Anal. Bioanal. Chem.* **2005**, *383* (3), 390–397. https://doi.org/10.1007/s00216-005-3217-1.

(47) Spik, G.; Coddeville, B.; Montreuil, J. Comparative Study of the Primary Structures of Sero-, Lacto- and Ovotransferrin Glycans from Different Species. *Biochimie* **1988**, *70* (11), 1459–1469. https://doi.org/10.1016/0300-9084(88)90283-0.

(48) Dasilva, M. L. C.; Tamura, T.; Rice, K. G. Derivatization and Purification of Bisecting Tyrosinamide-Oligosaccharides from Ovotransferrin. *Arch. Biochem. Biophys.* **1994**, *315* (2), 460–466. https://doi.org/10.1006/abbi.1994.1525.

(49) Gan, J.; Ben-Nissan, G.; Arkind, G.; Tarnavsky, M.; Trudeau, D.; Noda Garcia, L.; Tawfik, D. S.; Sharon, M. Native Mass Spectrometry of Recombinant Proteins from Crude Cell Lysates. *Anal. Chem.* **2017**, *89* (8), 4398–4404. https://doi.org/10.1021/acs.analchem.7b00398.

(50) Ben-Nissan, G.; Vimer, S.; Warszawski, S.; Katz, A.; Yona, M.; Unger, T.; Peleg, Y.; Morgenstern, D.; Cohen-Dvashi, H.; Diskin, R.; Fleishman, S. J.; Sharon, M. Rapid Characterization of Secreted Recombinant Proteins by Native Mass Spectrometry. *Commun. Biol.* **2018**, *1* (1), 213. https://doi.org/10.1038/s42003-018-0231-3.

(51) Vimer, S.; Ben-Nissan, G.; Sharon, M. Direct Characterization of Overproduced Proteins by Native Mass Spectrometry. *Nat. Protoc.* **2020**, *15*, 236–265. https://doi.org/10.1038/s41596-019-0233-8.

Chapter 3. Degronomics: Mapping the Interacting Peptidome of a Ubiquitin Ligase Using an Integrative Mass Spectrometry Strategy

3.1. Abstract

Human cells make use of hundreds of unique ubiquitin E3 ligases to ensure proteome fidelity and control cellular functions by promoting protein degradation. These processes require exquisite selectivity, but the individual roles of most E3s remain poorly characterized in part due to the challenges associated with identifying, quantifying, and validating substrates for each E3. We report an integrative mass spectrometry (MS) strategy for characterizing protein fragments that interact with KLHDC2, a human E3 that recognizes the extreme C-terminus of substrates. Using a combination of native MS, native top-down MS, MS of destabilized samples, and liquid chromatography MS, we identified and quantified a near complete fraction of the KLHDC2-binding peptidome in *E. coli* cells. This degronome includes peptides that originate from a variety of proteins. Although all identified protein fragments are terminated by diglycine, the preceding amino acids are diverse. These results significantly expand our understanding the sequences that can be recognized by KLHDC2, which provides insights into the potential substrates of this E3 in humans. We anticipate that this integrative MS strategy could be leveraged more broadly to characterize the degronomes of other E3 ligase substrate receptors, including those that adhere to the more common N-end rule for substrate recognition. Therefore, this work advances "degronomics," *i.e.*, identifying, quantifying, and validating functional E3:peptide interactions in order to determine the individual roles of each E3.

3.2. Introduction

The ubiquitin-proteasome system (UPS) regulates intracellular protein degradation.[1] E3 ubiquitin ligases are essential UPS enzymes that transfer ubiquitin from an E2 conjugating enzyme to a protein substrate[2] and ensure proteome fidelity by selectively eliminating aberrant proteins that may have translation, folding, or other errors.[3] There are hundreds of E3s in humans, but the individual roles of most remain poorly characterized.[4] One tremendous hurdle for determining the roles of individual E3s has been identifying their substrates. Several techniques have been adapted for characterizing E3-substrate pairing, including affinity-based capturing approaches,[5] yeast two-hybrid methods,[6] high-throughput microscopy screens,[7] and phage display assays.[8] Bottom-up, mass spectrometry (MS) based proteomics,[9] in which proteins are enzymatically digested into peptides, peptides are separated using liquid chromatography (LC), and peptides are sequenced using tandem MS, has been used to investigate several aspects of the UPS. For example, targets of the entire UPS have been determined by identifying ubiquitination sites across the proteome.[10,11] Quantitative proteomics has been used to characterize protein turnover in mammalian cells in response to expression of a selected E3 ligase, which identified several substrates for those E3s.[12,13]

One emerging technique for characterizing the UPS is global protein stability profiling (GPS), which is a high throughput approach for characterizing the stability of individual proteins.[14] Recently, GPS was used to identify hundreds of potential substrates of the Skp1–CUL1–F-box (SCF) family of ubiquitin ligases. Specific E3-substrate interactions are often dictated by a short linear sequence, known as a degron, in the substrate that is specifically recognized by the E3.[15] Identification of degrons can help predict and validate substrate targeting by an E3. For instance, GPS revealed that one family of E3 ligases, the CUL2-RING ligases (CRL2), recognizes degrons at the extreme C-terminus of substrates.[16,17] GPS analysis of

KLHDC2, a member of the CRL2 family, revealed binding to selenoprotein fragments that contain diglycine at the C terminus.[18] Subsequent characterization using *in vitro* binding assays and X-ray crystallography showed that KLHDC2 can recognize degron peptides with nanomolar affinities via the top surface pocket of its β-propeller domain.[19]

Despite substantial and sustained efforts to identify E3 ligase substrates, substrates have been validated for relatively few E3s. One challenge is that established methods, including conventional MS-based proteomics and GPS, do not directly probe E3:substrate binding. Native MS, in which samples are prepared in aqueous solutions with physiological pH and ionic strength, can preserve noncovalent interactions during transfer into the gas phase.[20] Due to this advantage, native MS has been used to identify protein cofactors[21–23] and to characterize the stoichiometry and composition of protein complexes.[24–26] More recently, native top-down MS, in which ions of intact proteins and protein complexes are subjected to fragmentation to yield sequence information, has been used to identify proteoforms[27] and link that information to higher-order protein structure.[28,29] Native MS has been used to construct structural models of the $CRL5^{SOCS2}$ ligase[30] and to characterize cofactor binding to the SCF^{FBXL3} and SCF^{TIR1} E3s.[31,32] We recently used native top-down MS to corroborate interactions between KLHDC2 and a selenoprotein fragment identified through GPS analysis;[19] otherwise native MS and native top-down MS have not been used previously to characterize E3:substrate interactions.

In this study, we report an integrative MS strategy that combines results from native and non-native experiments to characterize KLHDC2:peptide complexes that were copurified from cells. We describe the first application of native top-down MS to identify unknown copurified peptides. Our results provide direct evidence that KLHDC2 recognizes cellular peptides containing C-terminal glycylglycine or glycylalanine, and demonstrates high-affinity binding for

a diverse profile of peptides that contain those features. These results significantly expand our understanding of degrons that can be recognized by KLHDC2, which aids in identifying potential substrates of this E3 in humans. More generally, this integrative MS strategy advances "degronomics," *i.e.*, identifying, quantifying, and validating functional E3:peptide interactions in order to determine the individual roles of each E3.

3.3. Methods

3.3.1. Molecular Biology and Protein Purification for Mass Spectrometry Experiments.

The kelch repeat domain of human KLHDC2 (UniProt: Q9Y2U9, amino acids 22–362) was subcloned into the pET vector with an N-terminally fused His-elongation factor Ts (TSF) and a TEV-cleavage site. The His-TSF-KLHDC2 protein was overexpressed and purified from BL21 (DE3) *E. coli* cells. Bacterial cells transformed with the pET-based expression plasmid were grown in LB broth to an OD_{600} of 0.8-1 and induced with 0.5 mM IPTG. Cells were harvested, re-suspended and lysed in lysis buffer (20 mM Tris, pH 8.0, 200 mM NaCl, 20 mM imidazole) in the presence of protease inhibitors (1 μg mL^{-1} leupeptin, 1 μg mL^{-1} pepstatin and 100 μM phenylmethylsulfonyl fluoride) using a microfluidizer. The His-TSF-KLHDC2 protein was isolated from the soluble cell lysate by HisPur™ Ni-NTA Superflow Agarose (Thermo Fisher Scientific, Waltham, Massachusetts). After TEV cleavage of the His-TSF, KLHDC2 was further purified by Q Sepharose High Performance resin (GE Healthcare, Chicago, Illinois). The NaCl eluates were subjected to Superdex-200 size-exclusion chromatography (GE Healthcare). All samples were flash frozen in liquid nitrogen for storage prior to use.

3.3.2. Native MS. KLHDC2 purified from *E. coli* cells was buffer exchanged into aqueous ammonium acetate (200 mM, pH 7) using four cycles of dilution and re-concentration

with a 10K MWCO Spin-X UF centrifugal concentrator (Corning Inc., Corning, New York) at 14,000 g and 4 °C. The final concentration of KLHDC2 used for experiments was approximately 20 μM. 2-3 μL of the prepared solution was added into the tip of a pulled borosilicate glass capillary. A platinum wire electrode was inserted into the capillary and placed in direct contact with the solution. A potential between 0.5 and 1.0 kV was applied to the electrode to achieve ionization.[33] Those ions were then analyzed using a hybrid electrospray/quadrupole/ion-mobility/time-of-flight mass spectrometer (Waters Synapt G2 HDMS, Milford, Massachusetts). Native mass spectra were acquired using a 45 V bias between the sampling and extraction cones in the atmospheric-pressure interface that was operated at room temperature, and using a 3 V bias between the quadrupole mass filter and the trap collision cell that contained ~20 mTorr of argon gas. Activation in the trap collision cell was performed using those conditions, except that the bias between the quadrupole and trap collision cell was increased to 45 V. Activation in the atmospheric-pressure interface was achieved by increasing the bias between the sampling and extraction cones to 120 V. Tandem MS of peptide ions released during activation in the atmospheric-pressure interface was performed by increasing the bias between the quadrupole mass filter (used to isolate a released peptide ion) and the trap collision cell to 30-45 V. Mass spectra were calibrated externally using spectra obtained from electrospray ionization of 30 mg mL^{-1} aqueous CsI.

3.3.4. MS from Destabilizing Conditions. Thermal destabilization was achieved by transferring approximately 5 μL of the sample prepared for native MS into a 1.5 mL snap-cap vial, which was then placed in 55 °C heat block for 5 minutes. Chemical destabilization was achieved by adding 15% acetonitrile and 1% formic acid by volume into the sample prepared for native MS. The thermally and chemically destabilized samples were then each analyzed using

the parameters described for native MS. Tandem MS was performed by increasing the bias between the quadrupole mass filter (used to isolate the selected peptide ion) and the trap collision cell to 25-45 V.

3.3.5. LC-MS^2. The sample prepared for native MS was adjusted to pH 4 with the addition of triflouroacetic acid, loaded onto a Pierce C_{18} tip with a 10 µL bed (Thermo Fisher Scientific), washed, and eluted using an aqueous solution with 70% acetonitrile and 0.1% formic acid by volume. Peptides were separated using a nanoAcquity UPLC (Waters) using a 150 mm fused-silica emitter that was packed with reversed-phase ProntoSIL AQ 120-C18 5 µm resin (Bischoff Chromatography, Germany). A 60-minute linear gradient from 10-30% organic phase (acetonitrile with 0.1% formic acid), followed by a 1-minute gradient to 80% organic phase, was used with a flow rate of 0.3 µL/min. The eluent was analyzed using a Q Exactive Plus mass spectrometer (Thermo Fisher Scientific) and a data-dependent acquisition method. Full-scan MS spectra were acquired in positive ion mode with the resolution set to 70k and with a scan range of 400 to 2,500 *m*/*z*. MS^2 spectra were acquired on the top 20 precursors present in the MS^1 scan, and MS^2 spectra were acquired at 17.5k resolution with a 1.5 *m*/*z* isolation window.

3.3.6. Peptide Identification. Peptide sequences were identified using unrestricted database searches and downstream validation. A protein database was constructed that contained the *E. coli* BL21 proteome (Proteome ID UP000002032),[34] the His-TSF-KLHDC2 fusion protein that introduced with molecular biology (Figure S3), and reverse-sequence protein decoys. The experimental data was searched against this database using both Comet[35] and MSFragger.[36] Searches were performed using standard settings for accurate mass data and allowing for a wide range of isotope errors (-1, 0, +1, +2, +3 isotope offsets). The results of both searches were each

analyzed with PeptideProphet[37] using only the expect score as the discriminate and using the decoys to establish the negative distribution. Those analyses were then combined using iProphet.[38] This workflow was performed within the Trans-Proteomic Pipeline,[39] except the MSFragger search and the subsequent PeptideProphet analysis were performed using FragPipe.[36] Quantification based on the LC-MS data was performed using XPRESS[40] in label-free mode within the Trans-Proteomic Pipeline. PEAKS Studio 8.5 (Bioinformatics Solutions, Waterloo, Ontario) was used to sequence spectra obtained from native top-down and destabilizing MS experiments. Spectra from those experiments were signal averaged using MassLynx v4.1, converted into a single mzML file, then searched using a 0.1 Da precursor mass tolerance and a 0.1 Da fragment ion tolerance with no enzyme selected for *de novo* sequencing. For PEAKS DB analysis, the UniProt *E.coli* BL21 and K12 proteomes were combined with the FASTA entries for human KLHDC2 and the His-TSF tag.

3.3.7. Determining Relative Abundances of Bound Peptides. The native mass spectra were resampled with 4 equally spaced points between each *m*/*z* value. The signal profile for apo KLHDC2 was used to represent the contributions from each KLHDC2:peptide complex. A template was created by first isolating the feature for apo KLHDC2 in the resampled spectrum. Since the tailing edge of this feature was obscured by the peptide-bound portion of the spectrum, a tail was added by reflecting the leading edge of the profile and then reducing the intensity of this tail to reach the baseline of the spectrum. The relative intensity (I_n) for each complex was optimized by minimizing the total residual (Equation 1) using Powell's conjugate gradient method.[41] A notebook illustrating this method is included with the *Supporting Information*.

3.4. Results and Discussion

KLHDC2 is a human E3 ligase that is of considerable interest because it recognizes the C-terminus of substrates,[16–19] rather than the more common N-end rule[42] for substrate recognition. As part of our effort to characterize the molecular biophysics of KLHDC2, we used native MS to analyze this protein purified from *E. coli* cells. Several high-intensity peaks were observed at *m*/*z* values expected for the 11+ to 13+ charge states of apo KLHDC2. To our surprise, many additional features were also observed (Figure 1). The unexpected features correspond to ions that are approximately 400 to 1500 Da greater in mass than apo KLHDC2 and suggest that many different peptides had complexed with KLHDC2. Since exogenous peptides were not added at any point, these peptides must have bound to KLHDC2 during their expression or purification from *E. coli* cells. Therefore, the bound peptides represent bacteria-produced protein fragments that copurified with KLHDC2. To the best of our knowledge, these spectra show the first evidence of copurified protein fragments bound directly to an E3 using MS. Here we report the integrative mass spectrometry strategy that we developed to identify and quantify the protein fragments that comprise the interacting peptidome of this E3 ligase, *i.e.*, the degronome of KLHDC2 in these cells. We then discus the properties of this degronome and the potential for using this strategy to determine the roles of individual E3 ligases.

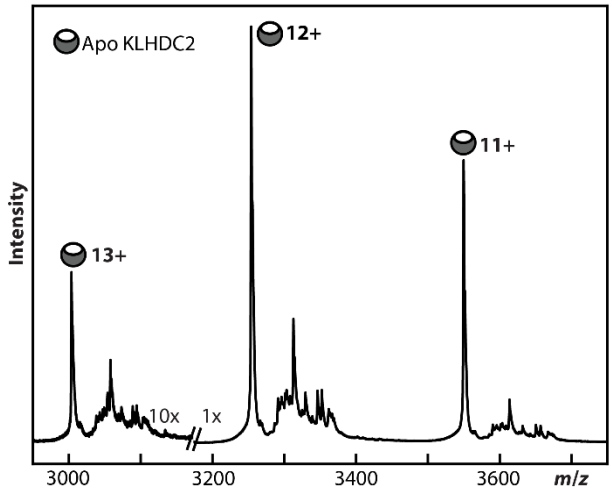

Figure 3.1. Native mass spectrum of KLHDC2 expressed in *E. coli* cells. In addition to features expected for apo KLHDC2, additional features corresponding to ions that are approximately 400-1500 Da larger in mass relative to the apo protein are also observed. These features are attributed to KLHDC2:peptide complexes. Intensities below 3200 *m*/*z* are increased tenfold to aid in visualization.

3.4.1. Native Top-Down MS. To determine the identity of the copurified peptides, we first used native top-down MS, *i.e.*, subjecting intact KLHDC2:peptide complex ions to multiple stages of activation in order to sequence the bound peptides. Peptides were released using collision-induced dissociation (CID), which was achieved at the atmospheric-pressure interface or in the trap collision cell of the mass spectrometer. When CID was performed in the trap collision cell, the transmission profile of the quadruple (positioned immediately prior to the trap cell and operated as an RF-only guide) was set to favor higher-*m*/*z* ions corresponding to apo or peptide-bound KLHDC2. This filtered away lower-*m*/*z* contaminants, including any peptides that were unbound in solution, and provides high confidence that the peptide ions observed after CID were released from complexes with KLHDC2. When CID was performed at the atmospheric-

pressure interface, it is possible that unbound peptides from solution may also be transmitted. CID performed in the trap collision cell (Figure S1) or at the atmospheric-pressure interface (Figure 2A) both yielded identical peptide ions, which corroborates that the peptide ions in both experiments originated from complexes with KLHDC2. Peptide ions released at the atmospheric-pressure interface were used for acquiring peptide fragmentation spectra because those ions could be quadrupole selected prior to subsequent fragmentation.

Seven interacting peptides were identified using native top-down MS; representative fragmentation spectra are shown in Figure 2B and S2. In order to minimize bias, the fragmentation spectra were first analyzed using unconstrained *de novo* sequencing and the candidate sequences were compared with a protein library, as described in the *Methods*. Five of the peptides are assigned to a series of N-terminal truncations of ASDEGEVIVFGG (Figure 2B), which is a fragment of KLHDC2. This assignment was validated by subjecting synthetic ASDEGEVIVFGG to tandem MS on the same instrument (Figure 2C). The origin of the N-terminal truncations are not yet understood, but are presumably the result of digestion or degradation that occurred *in vivo*. The other two peptides were assigned to EKGGGSGGG and GGGSGGG, which are fragments of the His-elongation factor Ts (TSF) tag that was fused to the N-terminus of KLHDC2 for expression and purification purposes (Figure S3). These peptides all feature C-terminal diglycine, which is present in the majority of the reported degrons of KLHDC2.[17] These seven degron-like peptides were all identified through the direct analysis of KLHDC2:peptide complexes that were formed in cells or during cell lysis and survived extensive purification.

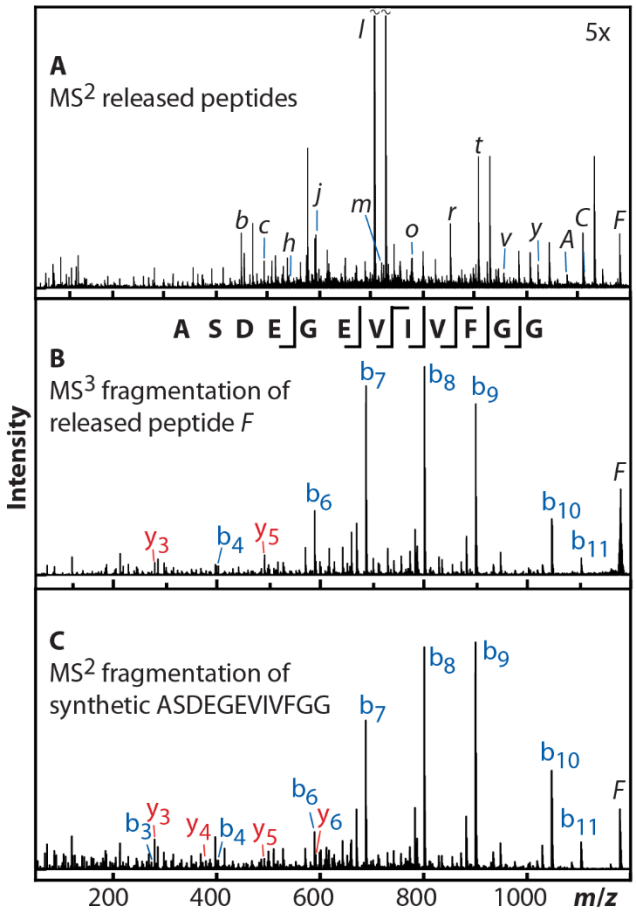

Figure 3.2. Native top-down MS of KLHDC2:peptide complexes. (A) Complexes were ionized and subjected to collision-induced dissociation (CID) at the atmospheric-pressure interface of the mass spectrometer, resulting in the release of peptide ions that are labeled using the IDs in Table 1. Intensities are increased fivefold to aid in visualization. Additional ions are annotated in Figure S2. (B) Ion *F* was isolated and subjected to CID in the trap collision cell of the instrument. CID fragments (labeled using the scheme of Roepstorff and Fohlman)[43] were used to assign ion *F* to ASDEGEVIVFGG, which is a fragment of KLHDC2. (C) Synthesized ASDEGEVIVFGG was resuspended in aqueous 200 mM ammonium acetate, ionized using electrospray, and subjected to CID in the trap collision cell of the same instrument. The fragmentation spectra in B and C are nearly identical, which corroborates the assignment of peptide *F* to ASDEGEVIVFGG.

Table 3.1. The degronome of KLHDC2 in *E. Coli* Cells based on observations (○) and sequencing (●) in *N*ative, *D*estabilizing, and *L*iquid *C*hromatography MS experiments.

Protein of Origin	Interacting Peptide	ID	RA[a]	*N*	*D*	*LC*
KLHDC2	VSDGRHMFVWGG	*J*	1.7(0.4)	–	○	●
	TVGNRGFVFGG	*A*	0.3(0.2)	○	○	●
	VGNRGFVFGG	*h*	1.4(0.4)	–	–	●
	NRGFVFGG	*D*	0.4(.4)	○	○	●
	FVFGG	*x*	0.5(0.1)	○	○	●
	SSDHLFLFGG	*r*	1.8(0.5)	○	–	●
	FLFGG	*f*	0.6(0.4)	○	-	●
	HTACASDEGEVIVFGG	*U*	0.2(0.1)	–	–	●
	ACASDEGEVIVFGG	*K*	0.3(0.4)	–	○	●
	CASDEGEVIVFGG	*H*	2.0(0.5)	○	–	●
	ASDEGEVIVFGG	*F*	5.7(1.1	●	●	●
	SDEGEVIVFGG	*C*	5.5(0.8)	●	●	●
	DEGEVIVFGG	*y*	1.7(0.2)	●	●	●
	EGEVIVFGG	*t*	4.3(0.5)	●	●	●
	GEVIVFGG	*o*	3.1(0.5)	○	○	●
	EVIVFGG	*m*	1.5(0.8)	–	○	●
	VIVFGG	*j*	5.4(1.7)	–	–	●
	IVFGG	*c*	1.9(1.1)	●	○	●
His-TSF solubility tag	AWSHPQFEKGGGSGGGSGG	*X*	0.8(0.3)	–	–	●
	WSHPQFEKGGGSGGGSGG	*W*	0.9(0.7)	–	–	●
	SHPQFEKGGGSGGGSGG	*R*	0.3(0.4)	–	–	●
	SSAWSHPQFEKGGGSGGG	*Y*	0.4(0.2)	–	–	●
	AWSHPQFEKGGGSGGG	*T*	0.5(0.1)	–	–	●
	WSHPQFEKGGGSGGG	*Q*	0.3(0.1)	–	–	●
	SHPQFEKGGGSGGG	*I*	3.7(0.8)	–	–	●
	HPQFEKGG	*s*	1.5(0.5)	○	○	●
	EKGGGSGGG	*l*	19.5(2.6)	●	●	–
	GGGSGGG	*b*	4.5(2.0)	●	○	–
UPF0441 protein YgiB	QRSATGTSSRSMGG	*N*	1.7(0.3)	–	–	●
Cell division protein FtsZ	TNDAVIKVIGVGGGG	*L*	0.6(0.3)	–	–	●
	VIKVIGVGGGG	*v*	0.8(0.1)	○	–	●
Glutamate dehydrogenase	SSAIGPYKGG	*u*	0.8(0.3)	○	–	●
	SAIGPYKGG	*q*	0.2(0.2	○	–	●
Outer membrane protein F	SDDFFVGRVGG	*E*	0.8(0.1)	–	–	●
	DDFFVGRVGG	*z*	0.8(0.5)	–	–	●
Malate dehydrogenase	MKVAVLGAAGG	*w*	0.9(0.2)	–	–	●
	KVALGAAGG	*p*	1.7(0.2)	–	–	●
Superoxide dismutase	TAFEGKLEEIIRSSEGG	*Z*	1.0(0.5)	–	–	●
30S ribosomal protein S11	TDRQGNALGWATAGG	*P*	0.2(0.1)	–	–	●
	LGWATAGG	*n*	1.1(0.9)	○	–	●
30S ribosomal protein S12	IGGEGHNLQEHSVILIRGG	*α*	0.5(0.2)	–	–	●
30S ribosomal protein S5	VFMQPASEGTGIIAGGA	*V*	0.5(0.2)	–	–	●
	MQPASEGTGIIAGGA	*M*	0.6(0.3)	–	–	●
Phage Shock operon rhodanese PspE	AEHWIDVRVPEQYQQEHVQGA	*δ*	1.6(1.8)	–	–	●
	IDVRVPEQYQQEHVQGA	*β*	0.8(0.6)	–	–	●
17 kDa surface antigen	IQGGDDSNVIGAIGGA	*O*	0.9(0.2)	–	–	●
Phenylalanine-tRNA ligase α subunit	LRELPPEERPAAGA	*S*	0.5(0.2)	–	–	●
DNA- binding protein HU-β	KSQLIDKIAAGA	*G*	0.5(0.3)	–	–	●
	SQLIDKIAAGA	*B*	0.1(0.1)	–	–	●
Ambiguous origin (See SI)	[b]LPPGG	*e*	1.2(0.6)	○	–	●
	LVYGG	*d*	2.3(0.9)	–	○	●
	LPEFGG	*k*	4.9(0.7)	–	–	●
	PEFGG	*g*	0.0(0.0)	○	○	●
	LIIGGG	*i*	3.0(0.5)	–	○	●
	[b]LYFGG	*a*	1.2(0.9)	○	–	●

[a] *R*elative *A*bundances are reported at the 95% confidence level.
[b] Additional sequences with leucine/isoleucine substitutions are also possible.

3.4.2. Integrating Complementary MS-Based Measurements. Although roughly half of the KLHDC2 containing ions observed using native MS were complexed with peptides (Figure 1), the abundance of each individual peptide was inherently lower than that of KLHDC2. This ultimately limited the number of peptides that could be observed in native MS^2 experiments (Figure 2A) and sequenced in native MS^3 experiments (Table 1). Since it was apparent that many other peptides were bound to KLHDC2 than were identified using native top-down MS alone, we developed an integrative MS-based strategy to identify additional bound peptides (Figure 3).

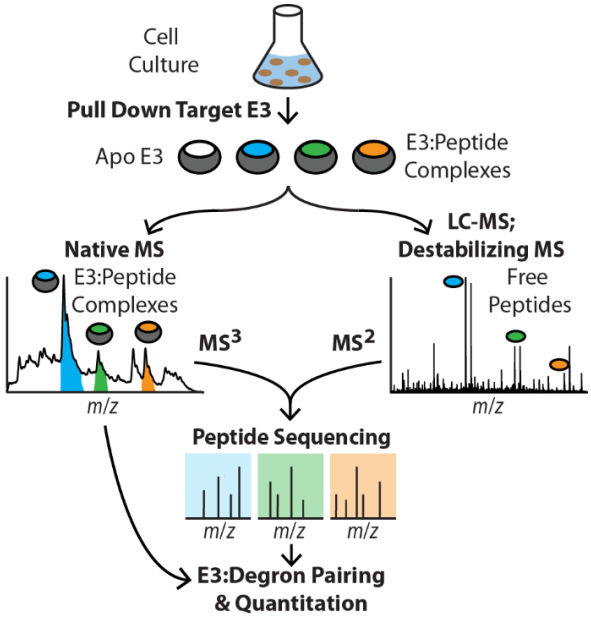

Figure 3.3. Integrative MS for degronomics. Native MS enabled the direct observation and mass determination of E3:peptide complexes that were purified from cells. Native top-down MS provided primary amino acid sequences for a subset of the bound peptides. Complementary MS-based approaches, including destabilized-sample MS^2 and LC-MS^2 provided additional peptide identifications. Together, these methods provide an integrative strategy to identify and quantify protein fragments interacting with an E3 in cells that could be used to study E3:degron interactions more broadly.

Destabilizing samples, by the addition of acetonitrile and formic acid (15% and 0.1% by volume, respectively) or heating the sample to 55 °C, caused peptides to be released from KLHDC2 in solution. MS^1 of these destabilized samples and MS^2 of the native-like complexes yielded similar spectra, but destabilized-sample MS yielded more intense signals for peptide ions than native MS^2. The highest-intensity ions from the destabilized samples were quadrupole selected for MS^2 (Figure S4), which confirmed the peptide sequences identified using native top-down MS (Table 1). However, analysis of the destabilized samples was limited by congested MS^1 spectra; some peptide ions could not be isolated completely using the quadrupole mass filter and others exhibited inadequate intensity for MS^2. Ultimately, no additional peptides were sequenced from the destabilized samples, but analysis of those samples corroborated many of the results from native top-down MS. LC-MS^2, without enzymatic digestion, was used to determine the sequences for the greatest number of interacting peptides (Table 1, probability scores from peptide-spectrum matches in Table S1).

Although the destabilized-sample MS and LC-MS^2 methods offer certain advantages, *i.e.*, greater precursor ion intensities and greater number of identified peptides, respectively, those methods both destroy the KLHDC2:peptide interactions prior to analysis. The native top-down and destabilized sample MS benefited from the use of direct infusion of the sample into the ion source. Relative to the LC-MS^2 experiments, adjusting the extent of activation for individual peptide ions contributed to higher quality fragmentation spectra and acquiring fragmentation spectra for more time enabled higher signal-to-noise spectra after averaging. Consequently, the average of the significance scores from PEAKS DB analysis ($-10\log_{10}P$) for the peptides identified using all three approaches are 51, 44, and 30 for native MS^3, destabilized-sample MS^2, and LC- MS^2, respectively. Interesting, two of the most abundant interacting peptides (*b* and *l*)

were not identified in the LC-MS^2 experiments. The absence of *b* and *l* in the LC-MS^2 data may be attributable to the hydrophobicity of each peptide, which is significantly lower than the peptides that were identified using LC-MS^2 (Table S2), and poor retention during solid-phase extraction and LC. Therefore, the data generated using all three approaches are complementary and enable the most comprehensive overview of the interacting peptidome of KLHDC2 in *E. coli* cells.

3.4.3. Relative Abundances. The peptides identified through the integrative MS strategy were used to assign features in the native mass spectrum (Figure 4A) that had been tentatively assigned to KLHDC2:peptide complexes (Figure 1). Fundamental studies indicate that native-like ions are formed through the charged-residue model for electrospray ionization, in which the final ions are the result of desolvation of charged droplets containing the biomolecule.[33,44,45] Since the ionization efficiency of KLHDC2 and its complexes should be largely independent of peptide binding,[46] we hypothesize that the relative abundances of the KLHDC2:peptide complex ions accurately reflect the relative abundance of those complexes in solution prior to ionization. In contrast, the ionization efficiency of peptides from electrospray can depend strongly on peptide length, sequence, hydrophobicity, solubility, and other factors.[47] For example, hydrophobic peptides have greater affinity for surface of the electrospray droplet,[48] which can result in their preferential ionization[49] and suppress the ionization of less hydrophobic peptides.[50] Therefore, it is normally extremely difficult to quantify peptides without using internal standards or peptide labelling, which is challenging to implement in discovery experiments. Label-free quantitation of the peptides discovered in these experiments was attempted based on their LC-MS elution profiles. The relative abundances of the peptides from that analysis are reported in Table S1, but due to the absence of highly abundant peptides such as *b* and *l* (Table 1) in the LC-

MS experiments and the predicted differences in the ionization of peptides as described above, those values are not correlated with the relative abundances determined from the native MS experiments.

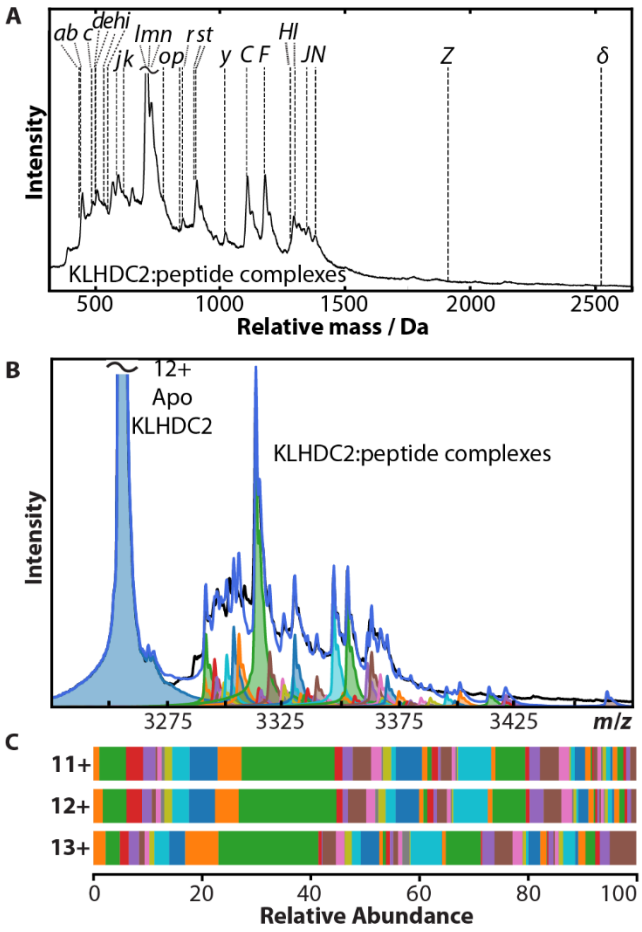

Figure 3.4. Identification and quantification of interacting peptides. (A) Native mass spectrum of the 12+ KLHDC2:peptide complex ions, plotted using a false mass axis that is relative to apo KLHDC2. Vertical bars are plotted using the masses of the interacting peptides that had a 1% or greater relative abundance and are labeled using the IDs in Table 1. For comparison, this region is plotted with all peptide IDs as a function of *m*/*z* in Figure S5. (B) The total residual (Equation

1) of the experiment (*black trace*) and the sum of modelled components (*blue trace*) was minimized by optimizing the relative intensity of each component. The contribution from each KLHDC2:peptide complex is represented using a different color. (C) The relative abundances of the interacting peptides determined using this approach for the 11+, 12+, and 13+ ions.

Congestion in these native mass spectra imposes some additional challenges in determining the relative abundance of the bound peptides. For example, the signals for apo KLHDC2 (Figure 1) are broad and appear to depend on the presence of non-specific adducts that persisted through electrospray ionization[51] as well as scattering with background gas in the time-of-flight mass analyzer. It is possible that individual peptides that bind to KLHDC2 contribute to differences in adduct formation during electrospray, but the major factors that broaden the MS signals for apo KLHDC2 are expected to be common to all KLHDC2:peptide complexes. From this, we hypothesize that the profile for apo KLHDC2 can be used as template to simulate the signals originating from each of the KLHDC2:peptide complexes. To test this hypothesis, the feature for apo KLHDC2 was extracted from the experimental spectrum as described in the *Methods*. For each of the *n* peptides in Table 1, this template was reproduced and shifted by τ, which is the neutral mass of the interacting peptide divided by the charge state of the complex. Next, the relative intensity (I_n) for each simulated feature was optimized to minimize the total residual:

$$\text{Total Residual} = \sqrt{\sum_{m/z}\left(Experiment\left(\frac{m}{z}\right) - \sum_0^n I_n \times Template\left(\frac{m}{z} + \tau_n\right)\right)^2} \qquad (1)$$

Note that I_n is a scalar that is relative to that for the original template (I_0). Additional details of this process are described in the *Methods* and *Supporting Information.*

Figure 4B shows results from this analysis of the 12+ ions. The sum of the components is strongly correlated with the experimental spectrum. The derivatives of these traces are shown in Figure S6, which highlight the fidelity and the limitations of this approach. The differences between the model and experiment are most likely due to limitations in the template and unidentified peptides; for instance, the experiment appears to include contributions from KLHDC2 bound to peptides with masses of approximately 398, 565, and 648 Da that were not identified and are therefore not included in the model. However, the model reproduces nearly 98% of signal observed for KLHDC2:peptide ions, suggesting that the integrative MS strategy identified the vast majority of the interacting peptides. The relative abundances of the interacting peptides determined from the analysis of the 11+, 12+, and 13+ KLHDC2:peptide ions are shown in Figure 4C. Overall, the relative abundances determined for different charge states and for technical replicates (Figure S6) are similar to each other, demonstrating the robustness of this approach. The relative abundances, averaged over all charge states and replicates, are reported in Table 1. The difference between the most and least abundant peptides (corresponding to 19.5 and 0.1% relative abundance) suggests that this approach has a dynamic range that is capable of quantifying peptides that differ nearly 100-fold in intensity. Note that the accuracy of this approach depends on the extent to which the original native mass spectrum (Figure 1) reports the distribution of interacting peptides in solution; different extents of peptide dissociation during protein purification or MS may bias these results. Additionally, the accuracy of the templating approach could be reduced if adduct formation differs significantly between the apo and peptide-bound components.

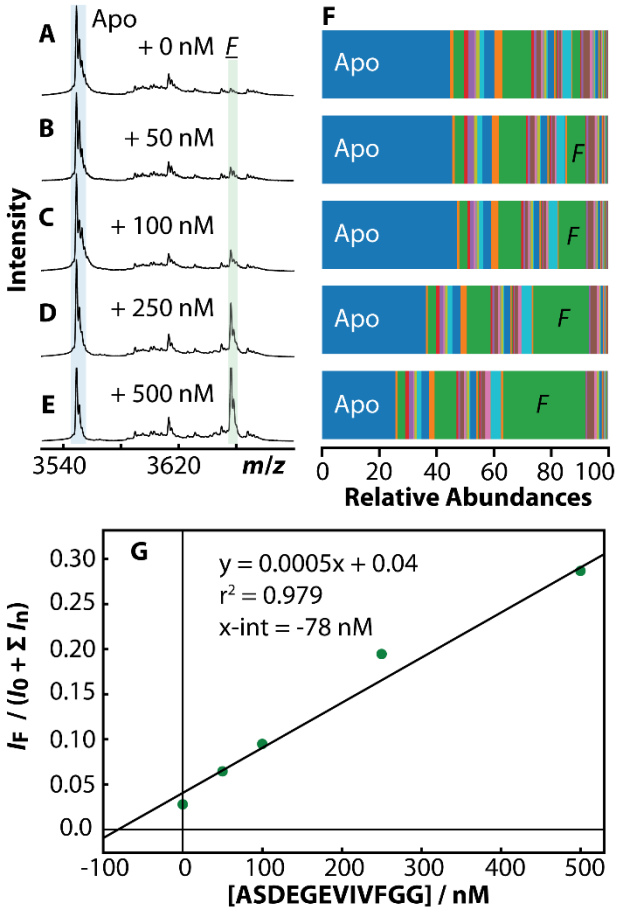

Figure 3.5. The peptide *F* (ASDEGEVIVFGG) was synthesized and added at various concentrations to the sample of KLHDC2 expressed in *E. coli* cells. (A-E) The intensity of the feature for the KLHDC2*:F* complex increases with peptide concentration, which is consistent with the intensities in the native mass spectra depending on the abundances of the corresponding complexes in solution. (F) The relative abundances of apo KLHDC2 and the KLHDC2:peptide complexes were determined for the spectra shown in A to E using the same method used for Figure 4. (G) The abundance of the KLHDC2*:F* complex relative to apo KLHDC2 and all KLHDC2:peptide complexes as function of the concentration of additional ASDEGEVIVFGG after it was spiked into the sample.

In order to characterize the relationship between the relative abundances determined using this quantitation approach and the abundances of the complexes in solution, we performed a titration experiment by adding the synthesized peptide *F* (ASDEGEVIVFGG). We added that peptide to the original sample of KLHDC2 from *E. Coli* cells and monitored the change in the resulting native mass spectra, which is shown in Figure 5A-E. The original sample was diluted twofold with the addition of the standard; the concentration of the additional standard is reported after this dilution step. The intensity of the feature assigned to the KLHDC2:*F* complex increased with the concentration of the spiked peptide. Figure 5F shows that the increase in the relative abundance of that complex is concomitant with a decrease in the relative abundance of the other components of the sample, especially apo KLHDC2. For comparison, Figure S7 shows the relative abundance of the other components, excluding apo KLHDC2 and KLHDC2:*F*. This analysis suggests that the concentration of the KLHDC2:*F* complex in sample was 164 ± 44 nM, prior to the addition of the standard (Figure 5G). These results also illustrate the advantage interpreting native mass spectra using an experimentally derived template. Even though the different experiments yield slightly different extents of nonspecific adducts following electrospray ionization (Figures 5A to 5E), those differences are inherently accounted for using this approach for quantitation.

We anticipate that this approach for quantitation could be used to quantify the relative abundance of other molecules that interact with proteins, including those that bind endogenously and/or lack standards. Additionally, this templating approach could be used to facilitate measurements of K_d values directly from native MS experiments. Although simulating native mass spectra in order to quantify the relative abundance of noncovalent complexes has been reported previously,[25,52] to the best of our knowledge this is the first report the use of an

integrative mass spectrometry strategy to inform the identities and masses of the contributing complexes as well the use of an experimentally derived template to represent peak shapes.

3.4.4. The Degronome of KLHDC2 in E. coli Cells. This degronome includes peptides that originate from a variety of proteins, including those from *E. coli* and those introduced using molecular biology (Table 1). *E. coli* doesn't have a UPS or encode KLHDC2, so the peptides that originate from *E. coli* proteins may not directly inform the role of KLHDC2 in humans. However, these *E. coli* derived peptides may provide insights into KLHDC2 function in humans. For instance, peptide *N* is the extreme C-terminal fragment of UPF0441 protein YgiB, whereas all other peptides originate from the interior or close to the N-termini of the associated proteins (Table S3). This suggests that KLHDC2 may recognize proteins that natively feature diglycine at their extreme C-terminus, in addition to protein fragments.

Interestingly, many of the peptides originated from KLHDC2. There are six interspersed diglycines within KLHDC2 (Figure 6A, one for each of the Kelch repeats of the protein) and interacting peptides terminating in four of the six diglycines were identified (Figure 6B). To determine the binding affinity of ASDEGEVIVFGG, which is the most-abundant peptide originating from KLHDC2, we used a luminescence-based competition assay (AlphaScreen, see *Supporting Information* and Figure S8). In this experiment, free ASDEGEVIVFGG peptide competed for binding to KLHDC2 against immobilized HLRGSPPPMAGG, which enables a direct comparison with the other degron peptides that have been characterized.[19] Because the concentration of immobilized components is extremely low relative to that of the free peptide, the IC_{50} values determined here are approximately equal to the corresponding K_d values.[53] Strikingly, this internal KLHDC2 peptide binds almost as tightly (IC_{50} = 5.3 nM) as the recently characterized selenoprotein fragment (HLRGSPPPMAGG, IC_{50} = 3.4 nM).[19] An intriguing,

albeit speculative, possibility is that the high affinity between KLHDC2 and its internal fragments may enable KLHDC2 to clear its own proteolytic products through the UPS.

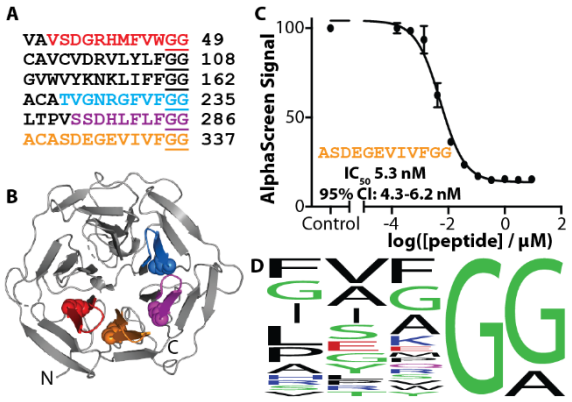

Figure 3.6. Recognition of KLHDC2 protein fragments. (A) The sequences near the six internal diglycines of KLHDC2 are shown. The colored regions represent internal peptides that were identified using integrative MS. (B) These internal fragments are highlighted on the structure of KLHDC2 and correspond by color to the sequences in panel A (diglycines represented as spheres). (C) The peptide ASDEGEVIVFGG binds to KLHDC2 with an IC_{50} value of 5.3 nM. (D) LOGO plot[54] of all peptides in Table 1.

Although all identified peptides are terminated by diglycine or glycylalanine (Table 1), the preceding amino acids in those peptides are diverse (Figure 6D). This finding is consistent with the extremely tight binding of both ASDEGEVIVFGG and HLRGSPPPMAGG, as well as the previous report that KLHDC2 binds degron peptides through non-covalent interactions between the peptide backbone carbonyls and the binding pocket of KLHDC2, rather than specific interactions with amino acid side chains of the peptide.[19] However, if only the presence of diglycine or glyclalanine were required, the *E. coli* peptides in this degronome are much less

diverse than might have been considered possible. The *E. coli* proteome contains ~4300 proteins, ~2700 of which contain a total of nearly 7000 diglycines and ~3000 of which contain ~8000 glycylalanines. The peptides identified here originate from only 12 of those proteins, which based on comparisons with a meta-analysis of results for *E. coli* K-12,[55] have cellular concentrations that likely span several orders of magnitude (Figure S9). This finding indicates that KLHDC2 has particularly high affinity for these degrons or that only a small number of the candidate proteins undergo proteolytic digestion that exposes these C-terminal motifs. Therefore, this degronome may be useful for extending the profile of sequences that can be recognized by KLHDC2, *i.e.*, treat the *E. coli* proteome as a peptide library that can be recognized by this E3. For instance, if the last 5 amino acids from each peptide originating from an *E. coli* protein represents a functional degron, KLHDC may have hundreds of potential protein substrates across the human proteome (Table S4). In the future, applying the integrative MS strategy reported here to study the degronomes of E3s in human cells may enable even more direct insights into the roles of those E3s in human biology.

3.5. Conclusions

We demonstrated an integrative MS strategy for identifying and quantifying cellular peptides that interacted with an E3 ubiquitin ligase (Figure 3). Unlike other methods for identifying E3 substrates, native MS provides direct evidence of E3:peptide binding. Combining results from native MS (Figure 1), native top-down MS (Figure 2), MS^2 of destabilized samples, and liquid chromatography MS^2 revealed a near complete fraction of the KLHDC2-binding peptidome from *E. coli* (Figure 4 and Table 1). These results demonstrate that KLHDC2 binds to peptides containing C-terminal diglycine or glycylalanine, and that a wide profile of sequences

preceding the C-terminus are recognized (Figure 6D). Using the native mass spectra and the comprehensive list of identified peptides, we demonstrated a novel and robust method for determining the relative abundance of interacting molecules (Figure 4B). The direct and quantitative data enabled by this integrative MS strategy offers many advantages relative to existing tools for characterizing E3-substrate pairing. Given the persistent challenges associated with identifying and validating substrates for the hundreds of E3 ligases in humans, we anticipate that MS-based degronomics will be leveraged more broadly to determine the individual roles of other E3s, including those that adhere to the more common N-end rule[42] for substrate recognition.

3.6. Supporting Information

The online version of this article contains supporting information, which is available to authorized users.

3.7. Acknowledgements

Research reported in this publication was supported by the National Science Foundation under award number CHE-1807382 (M. F. B.), the National Institutes of Health under award number T32GM008268 (training grant to D. C.), and the Howard Hughes Medical Institute (N. Z.). The authors thank David Shteynberg (Institute for Systems Biology) for useful discussions. and Priska Von Haller and the University of Washington Proteomics Resource for assistance with LC-MS experiments.

3.8. References

(1) Goldberg, A. L. Protein Degradation and Protection against Misfolded or Damaged Proteins. *Nature* **2003**, *426* (6968), 895–899. https://doi.org/10.1038/nature02263.

(2) Deshaies, R. J.; Joazeiro, C. A. P. RING Domain E3 Ubiquitin Ligases. *Annu. Rev. Biochem.* **2009**, *78* (1), 399–434. https://doi.org/10.1146/annurev.biochem.78.101807.093809.

(3) Glickman, M. H.; Ciechanover, A. The Ubiquitin-Proteasome Proteolytic Pathway: Destruction for the Sake of Construction. *Physiol. Rev.* **2002**, *82* (2), 373–428. https://doi.org/10.1152/physrev.00027.2001.

(4) Iconomou, M.; Saunders, D. N. Systematic Approaches to Identify E3 Ligase Substrates. *Biochem. J.* **2016**, *473* (22), 4083–4101. https://doi.org/10.1042/BCJ20160719.

(5) Mark, K. G.; Loveless, T. B.; Toczyski, D. P. Isolation of Ubiquitinated Substrates by Tandem Affinity Purification of E3 Ligase–Polyubiquitin-Binding Domain Fusions (Ligase Traps). *Nat. Protoc.* **2016**, *11*, 291. https://doi.org/10.1038/nprot.2016.008.

(6) Guo, Z.; Song, E.; Ma, S.; Wang, X.; Gao, S.; Shao, C.; Hu, S.; Jia, L.; Tian, R.; Xu, T.; et al. Proteomics Strategy to Identify Substrates of LNX, a PDZ Domain-Containing E3 Ubiquitin Ligase. *J. Proteome Res.* **2012**, *11* (10), 4847–4862. https://doi.org/10.1021/pr300674c.

(7) Benanti, J. A.; Cheung, S. K.; Brady, M. C.; Toczyski, D. P. A Proteomic Screen Reveals SCFGrr1 Targets That Regulate the Glycolytic–Gluconeogenic Switch. *Nat. Cell Biol.* **2007**, *9*, 1184. https://doi.org/10.1038/ncb1639.

(8) Guo, Z.; Wang, X.; Li, H.; Gao, Y. Screening E3 Substrates Using a Live Phage Display Library. *PloS One* **2013**, *8* (10), e76622–e76622. https://doi.org/10.1371/journal.pone.0076622.

(9) Zhang, Y.; Fonslow, B. R.; Shan, B.; Baek, M.-C.; Yates, J. R., 3rd. Protein Analysis by Shotgun/Bottom-up Proteomics. *Chem. Rev.* **2013**, *113* (4), 2343–2394. https://doi.org/10.1021/cr3003533.

(10) Kim, W.; Bennett, E. J.; Huttlin, E. L.; Guo, A.; Li, J.; Possemato, A.; Sowa, M. E.; Rad, R.; Rush, J.; Comb, M. J.; et al. Systematic and Quantitative Assessment of the Ubiquitin-Modified Proteome. *Mol. Cell* **2011**, *44* (2), 325–340. https://doi.org/10.1016/j.molcel.2011.08.025.

(11) Wagner, S. A.; Beli, P.; Weinert, B. T.; Nielsen, M. L.; Cox, J.; Mann, M.; Choudhary, C. A Proteome-Wide, Quantitative Survey of in Vivo Ubiquitylation Sites Reveals Widespread Regulatory Roles. *Mol. Cell. Proteomics* **2011**, *10* (10), M111.013284-M111.013284. https://doi.org/10.1074/mcp.M111.013284.

(12) Burande, C. F.; Heuzé, M. L.; Lamsoul, I.; Monsarrat, B.; Uttenweiler-Joseph, S.; Lutz, P. G. A Label-Free Quantitative Proteomics Strategy to Identify E3 Ubiquitin Ligase Substrates Targeted to Proteasome Degradation. *Mol. Cell. Proteomics* **2009**, *8* (7), 1719–1727. https://doi.org/10.1074/mcp.M800410-MCP200.

(13) Hör, S.; Ziv, T.; Admon, A.; Lehner, P. J. Stable Isotope Labeling by Amino Acids in Cell Culture and Differential Plasma Membrane Proteome Quantitation Identify New Substrates for the MARCH9 Transmembrane E3 Ligase. *Mol. Cell. Proteomics* **2009**, *8* (8), 1959–1971. https://doi.org/10.1074/mcp.M900174-MCP200.

(14) Yen, H.-C. S.; Xu, Q.; Chou, D. M.; Zhao, Z.; Elledge, S. J. Global Protein Stability Profiling in Mammalian Cells. *Science* **2008**, *322* (5903), 918. https://doi.org/10.1126/science.1160489.

(15) Yen, H.-C. S.; Elledge, S. J. Identification of SCF Ubiquitin Ligase Substrates by Global Protein Stability Profiling. *Science* **2008**, *322* (5903), 923. https://doi.org/10.1126/science.1160462.

(16) Koren, I.; Timms, R. T.; Kula, T.; Xu, Q.; Li, M. Z.; Elledge, S. J. The Eukaryotic Proteome Is Shaped by E3 Ubiquitin Ligases Targeting C-Terminal Degrons. *Cell* **2018**, *173* (7), 1622-1635.e14. https://doi.org/10.1016/j.cell.2018.04.028.

(17) Lin, H.-C.; Yeh, C.-W.; Chen, Y.-F.; Lee, T.-T.; Hsieh, P.-Y.; Rusnac, D. V.; Lin, S.-Y.; Elledge, S. J.; Zheng, N.; Yen, H.-C. S. C-Terminal End-Directed Protein Elimination by CRL2 Ubiquitin Ligases. *Mol. Cell* **2018**, *70* (4), 602-613.e3. https://doi.org/10.1016/j.molcel.2018.04.006.

(18) Lin, H.-C.; Ho, S.-C.; Chen, Y.-Y.; Khoo, K.-H.; Hsu, P.-H.; Yen, H.-C. S. CRL2 Aids Elimination of Truncated Selenoproteins Produced by Failed UGA/Sec Decoding. *Science* **2015**, *349* (6243), 91–95. https://doi.org/10.1126/science.aab0515.

(19) Rusnac, D.-V.; Lin, H.-C.; Canzani, D.; Tien, K. X.; Hinds, T. R.; Tsue, A. F.; Bush, M. F.; Yen, H.-C. S.; Zheng, N. Recognition of the Diglycine C-End Degron by CRL2KLHDC2 Ubiquitin Ligase. *Mol. Cell* **2018**, *72* (5), 813-822.e4. https://doi.org/10.1016/j.molcel.2018.10.021.

(20) Leney, A. C.; Heck, A. J. R. Native Mass Spectrometry: What Is in the Name? *J. Am. Soc. Mass Spectrom.* **2017**, *28* (1), 5–13. https://doi.org/10.1007/s13361-016-1545-3.

(21) Marcoux, J.; Wang, S. C.; Politis, A.; Reading, E.; Ma, J.; Biggin, P. C.; Zhou, M.; Tao, H.; Zhang, Q.; Chang, G.; et al. Mass Spectrometry Reveals Synergistic Effects of Nucleotides, Lipids, and Drugs Binding to a Multidrug Resistance Efflux Pump. *Proc. Natl. Acad. Sci. USA* **2013**, *110* (24), 9704. https://doi.org/10.1073/pnas.1303888110.

(22) Dyachenko, A.; Gruber, R.; Shimon, L.; Horovitz, A.; Sharon, M. Allosteric Mechanisms Can Be Distinguished Using Structural Mass Spectrometry. *Proc. Natl. Acad. Sci. USA* **2013**, *110* (18), 7235. https://doi.org/10.1073/pnas.1302395110.

(23) Skinner, O. S.; Haverland, N. A.; Fornelli, L.; Melani, R. D.; Do Vale, L. H. F.; Seckler, H. S.; Doubleday, P. F.; Schachner, L. F.; Srzentić, K.; Kelleher, N. L.; et al. Top-down Characterization of Endogenous Protein Complexes with Native Proteomics. *Nat. Chem. Biol.* **2017**, *14*, 36. https://doi.org/10.1038/nchembio.2515.

(24) Zhou, M.; Sandercock, A. M.; Fraser, C. S.; Ridlova, G.; Stephens, E.; Schenauer, M. R.; Yokoi-Fong, T.; Barsky, D.; Leary, J. A.; Hershey, J. W.; et al. Mass Spectrometry Reveals Modularity and a Complete Subunit Interaction Map of the Eukaryotic Translation Factor EIF3. *Proc. Natl. Acad. Sci. USA* **2008**, *105* (47), 18139. https://doi.org/10.1073/pnas.0801313105.

(25) Stengel, F.; Baldwin, A. J.; Bush, M. F.; Hilton, G. R.; Lioe, H.; Basha, E.; Jaya, N.; Vierling, E.; Benesch, J. L. P. Dissecting Heterogeneous Molecular Chaperone Complexes Using a Mass Spectrum Deconvolution Approach. *Chem. Biol.* **2012**, *19* (5), 599–607. https://doi.org/10.1016/j.chembiol.2012.04.007.

(26) Snijder, J.; van de Waterbeemd, M.; Damoc, E.; Denisov, E.; Grinfeld, D.; Bennett, A.; Agbandje-McKenna, M.; Makarov, A.; Heck, A. J. R. Defining the Stoichiometry and Cargo Load of Viral and Bacterial Nanoparticles by Orbitrap Mass Spectrometry. *J. Am. Chem. Soc.* **2014**, *136* (20), 7295–7299. https://doi.org/10.1021/ja502616y.

(27) Aebersold, R.; Agar, J. N.; Amster, I. J.; Baker, M. S.; Bertozzi, C. R.; Boja, E. S.; Costello, C. E.; Cravatt, B. F.; Fenselau, C.; Garcia, B. A.; et al. How Many Human

Proteoforms Are There? *Nat. Chem. Biol.* **2018**, *14* (3), 206–214. https://doi.org/10.1038/nchembio.2576.

(28) Skinner, O. S.; Havugimana, P. C.; Haverland, N. A.; Fornelli, L.; Early, B. P.; Greer, J. B.; Fellers, R. T.; Durbin, K. R.; Do Vale, L. H. F.; Melani, R. D.; et al. An Informatic Framework for Decoding Protein Complexes by Top-down Mass Spectrometry. *Nat. Methods* **2016**, *13*, 237. https://doi.org/10.1038/nmeth.3731.

(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. https://doi.org/10.1038/nchem.2908.

(30) Bulatov, E.; Martin, E. M.; Chatterjee, S.; Knebel, A.; Shimamura, S.; Konijnenberg, A.; Johnson, C.; Zinn, N.; Grandi, P.; Sobott, F.; et al. Biophysical Studies on Interactions and Assembly of Full-Size E3 Ubiquitin Ligase: SUPPRESSOR OF CYTOKINE SIGNALING 2 (SOCS2)-ELONGIN BC-CULLIN 5-RING BOX PROTEIN 2 (RBX2). *J. Biol. Chem.* **2015**, *290* (7), 4178–4191. https://doi.org/10.1074/jbc.M114.616664.

(31) Tan, X.; Calderon-Villalobos, L. I. A.; Sharon, M.; Zheng, C.; Robinson, C. V.; Estelle, M.; Zheng, N. Mechanism of Auxin Perception by the TIR1 Ubiquitin Ligase. *Nature* **2007**, *446*, 640. https://doi.org/10.1038/nature05731.

(32) Xing, W.; Busino, L.; Hinds, T. R.; Marionni, S. T.; Saifee, N. H.; Bush, M. F.; Pagano, M.; Zheng, N. SCFFBXL3 Ubiquitin Ligase Targets Cryptochromes at Their Cofactor Pocket. *Nature* **2013**, *496*, 64. https://doi.org/10.1038/nature11964.

(33) Davidson, K. L.; Oberreit, D. R.; Hogan, C. J.; Bush, M. F. Nonspecific Aggregation in Native Electrokinetic Nanoelectrospray Ionization. *Int. J. Mass Spectrom.* **2017**, *420*, 35–42. https://doi.org/10.1016/j.ijms.2016.09.013.

(34) Lucas, S.; Copeland, A.; Lapidus, A.; Glavina Del Rio, T.; Dalin, E.; Tice, H.; Bruce, D.; Goodwin, L.; Pitluck, S.; LaButti, K. M.; et al. Complete Sequence of Escherichia Coli BL21(DE3). EMBL/GenBank/DDBJ databases July 2009.

(35) Eng, J. K.; Jahan, T. A.; Hoopmann, M. R. Comet: An Open-Source MS/MS Sequence Database Search Tool. *PROTEOMICS* **2013**, *13* (1), 22–24. https://doi.org/10.1002/pmic.201200439.

(36) Kong, A. T.; Leprevost, F. V.; Avtonomov, D. M.; Mellacheruvu, D.; Nesvizhskii, A. I. MSFragger: Ultrafast and Comprehensive Peptide Identification in Mass Spectrometry–Based Proteomics. *Nat. Methods* **2017**, *14*, 513.

(37) Keller, A.; Nesvizhskii, A. I.; Kolker, E.; Aebersold, R. Empirical Statistical Model To Estimate the Accuracy of Peptide Identifications Made by MS/MS and Database Search. *Anal. Chem.* **2002**, *74* (20), 5383–5392. https://doi.org/10.1021/ac025747h.

(38) Shteynberg, D.; Deutsch, E. W.; Lam, H.; Eng, J. K.; Sun, Z.; Tasman, N.; Mendoza, L.; Moritz, R. L.; Aebersold, R.; Nesvizhskii, A. I. IProphet: Multi-Level Integrative Analysis of Shotgun Proteomic Data Improves Peptide and Protein Identification Rates and Error Estimates. *Mol. Amp Cell. Proteomics* **2011**, *10* (12), M111.007690. https://doi.org/10.1074/mcp.M111.007690.

(39) Deutsch, E. W.; Mendoza, L.; Shteynberg, D.; Slagel, J.; Sun, Z.; Moritz, R. L. Trans-Proteomic Pipeline, a Standardized Data Processing Pipeline for Large-Scale

Reproducible Proteomics Informatics. *PROTEOMICS – Clin. Appl.* **2015**, *9* (7–8), 745–754. https://doi.org/10.1002/prca.201400164.

(40) Han, D. K.; Eng, J.; Zhou, H.; Aebersold, R. Quantitative Profiling of Differentiation-Induced Microsomal Proteins Using Isotope-Coded Affinity Tags and Mass Spectrometry. *Nat. Biotechnol.* **2001**, *19* (10), 946–951. https://doi.org/10.1038/nbt1001-946.

(41) Powell, M. J. D. An Efficient Method for Finding the Minimum of a Function of Several Variables without Calculating Derivatives. *Comput. J.* **1964**, *7* (2), 155–162. https://doi.org/10.1093/comjnl/7.2.155.

(42) Tasaki, T.; Sriram, S. M.; Park, K. S.; Kwon, Y. T. The N-End Rule Pathway. *Annu. Rev. Biochem.* **2012**, *81* (1), 261–289. https://doi.org/10.1146/annurev-biochem-051710-093308.

(43) Roepstorff, P.; Fohlman, J. Letter to the Editors. *Biomed. Mass Spectrom.* **1984**, *11* (11), 601–601. https://doi.org/10.1002/bms.1200111109.

(44) Konermann, L.; Ahadi, E.; Rodriguez, A. D.; Vahidi, S. Unraveling the Mechanism of Electrospray Ionization. *Anal. Chem.* **2013**, *85* (1), 2–9. https://doi.org/10.1021/ac302789c.

(45) Allen, S. J.; Schwartz, A. M.; Bush, M. F. Effects of Polarity on the Structures and Charge States of Native-Like Proteins and Protein Complexes in the Gas Phase. *Anal. Chem.* **2013**, *85* (24), 12055–12061. https://doi.org/10.1021/ac403139d.

(46) Kitova, E. N.; El-Hawiet, A.; Schnier, P. D.; Klassen, J. S. Reliable Determinations of Protein–Ligand Interactions by Direct ESI-MS Measurements. Are We There Yet? *J. Am. Soc. Mass Spectrom.* **2012**, *23* (3), 431–441. https://doi.org/10.1007/s13361-011-0311-9.

(47) Mirzaei, H.; Regnier, F. Enhancing Electrospray Ionization Efficiency of Peptides by Derivatization. *Anal. Chem.* **2006**, *78* (12), 4175–4183. https://doi.org/10.1021/ac0602266.

(48) Cech, N. B.; Enke, C. G. Practical Implications of Some Recent Studies in Electrospray Ionization Fundamentals. *Mass Spectrom. Rev.* **2002**, *20* (6), 362–387. https://doi.org/10.1002/mas.10008.

(49) Iribarne, J. V.; Dziedzic, P. J.; Thomson, B. A. Atmospheric Pressure Ion Evaporation-Mass Spectrometry. *Int. J. Mass Spectrom. Ion Phys.* **1983**, *50* (3), 331–347. https://doi.org/10.1016/0020-7381(83)87009-0.

(50) Hirabayashi, A.; Ishimaru, M.; Manri, N.; Yokosuka, T.; Hanzawa, H. Detection of Potential Ion Suppression for Peptide Analysis in Nanoflow Liquid Chromatography/Mass Spectrometry. *Rapid Commun. Mass Spectrom.* **2007**, *21* (17), 2860–2866. https://doi.org/10.1002/rcm.3157.

(51) Felitsyn, N.; Peschke, M.; Kebarle, P. Origin and Number of Charges Observed on Multiply-Protonated Native Proteins Produced by ESI. *Int. J. Mass Spectrom.* **2002**, *219* (1), 39–63. https://doi.org/10.1016/S1387-3806(02)00588-2.

(52) Morgner, N.; Robinson, C. V. Massign: An Assignment Strategy for Maximizing Information from the Mass Spectra of Heterogeneous Protein Assemblies. *Anal. Chem.* **2012**, *84* (6), 2939–2948. https://doi.org/10.1021/ac300056a.

(53) Yung-Chi, C.; Prusoff, W. H. Relationship between the Inhibition Constant (KI) and the Concentration of Inhibitor Which Causes 50 per Cent Inhibition (I50) of an Enzymatic Reaction. *Biochem. Pharmacol.* **1973**, *22* (23), 3099–3108. https://doi.org/10.1016/0006-2952(73)90196-2.

(54) Crooks, G. E. WebLogo: A Sequence Logo Generator. *Genome Res.* **2004**, *14* (6), 1188–1190. https://doi.org/10.1101/gr.849004.

(55) Wang, M.; Herrmann, C. J.; Simonovic, M.; Szklarczyk, D.; von Mering, C. Version 4.0 of PaxDb: Protein Abundance Data, Integrated across Model Organisms, Tissues, and Cell-Lines. *PROTEOMICS* **2015**, *15* (18), 3163–3168. https://doi.org/10.1002/pmic.201400441.

Chapter 4. Ion Mobility of Proteins in Nitrogen Gas: Effects of Charge State, Charge Distribution, and Structure

4.1. Abstract

Ion mobility is emerging as a rapid and sensitive tool for structural characterization. Collision cross section (Ω) values determined using ion mobility are often compared to values calculated for candidate structures generated through molecular modeling. Several methods exist for calculating Ω values, but the trajectory method explicitly includes contributions from long-range, ion-neutral interactions. Recent implementations of the trajectory method have significantly reduced its expense and have made applications to proteins far more tractable. Here, we use ion mobility experiments and trajectory method calculations to characterize the effects of charge state, charge distribution, and structure on the ion mobility of proteins in nitrogen gas. These results show that ion-induced dipole interactions contribute significantly to the Ω values of these ions with nitrogen gas, even for the modestly charged ions commonly observed in native mass spectrometry experiments. Therefore, these interactions contribute significantly to the values measured in most structural biology and biophysics applications of ion mobility using nitrogen gas. Comparisons between the reciprocal mobilities of protein ions in helium gas and in nitrogen gas show that there are significant, non-correlated differences between these values. As a consequence, it is challenging to estimate the errors associated with interconverting between helium- and nitrogen-based mobilities without extensive characterization in both gases, even for ions of proteins with similar sequences. Therefore, we recommend reporting Ω and mobility

values that are based on the predominant gas present in the separation and applying additional caution when comparing results from mobility experiments performed using different gases.

4.2. Introduction

Ion mobility (IM) mass spectrometry (MS) is emerging as a rapid, sensitive, and robust tool for molecular analysis and structural characterization.[1,2] IM separates ions by their mobility (*K*) in a buffer gas under an applied electric field. When combined with MS, IM separations have proved valuable in the characterization of polymers,[3,4] nanoparticles,[5] carbohydrates,[6,7] and other analytes. Collision cross section (Ω) values determined from experimental *K* values can be used in structural analysis through comparison to Ω values calculated for candidate structures.[8,9] This approach has had a string of recent successes in structural biology.[10–13]

The recent success of IM-MS in biomolecular characterization has been fueled by improvements in IM technology and the development of commercial instruments.[14–16] Nearly all commercial instruments use N_2 as the buffer gas. The selection of N_2 can be attributed to practical reasons including cost and availability,[15] as well as the observation of improved resolution for separations using N_2 relative to other gases.[14,17] However, many foundational IM-MS investigations made use of He buffer gas[18–20] and methods for calculating Ω_{He} values are more mature.[21–23] Consequently, many studies calibrate mobilities determined from N_2-based measurements using Ω_{He} values reported in the literature.[24] One justification for this approach is the linear correlation observed for the reciprocal reduced mobilities in N_2 and He for selected ions.[15,25,26] However, deviations have been observed and attributed to differences in properties including charge state, charge distribution, molecular shape and size, and surface area.[23,25,27–30] As a consequence, several studies have reported that selecting calibrants with properties that are similar to the analyte is critical to managing errors in Ω determination.[25,27,29,31] Selecting

calibrants with similar properties may also reduce errors associated with velocity relaxation, which is more significant in IM devices that use dynamic electric fields, *e.g.*, travelling waves.[32]

The difficulty associated with converting between Ω_{He} and Ω_{N_2} values arises from the intrinsic physical properties of each gas. For example, N_2 is larger in size and mass than He, and consequently much more momentum may be transferred during collisions with N_2. The polarizability of N_2 (1.74 $Å^3$)[33] is also much larger than that of He (0.21 $Å^3$),[34] which contributes to larger Ω values through ion-induced dipole interactions.[23,35–37] Due to these differences, Bleiholder proposed that IM data "acquired in different buffer gases are functionally not rigorously related but instead to some extent complementary."[28] Therefore, errors associated with interconverting mobility data between gases or treating gas blends, *e.g.*, air or mixtures of helium and nitrogen, as pure gases can ultimately diminish the accuracy of IM results.

Several methods have been developed or adapted to calculate Ω_{N_2} values. For example, the exact hard-spheres scattering method (EHSS) calculates Ω based on the expected momentum transfer between neutrals and ions with all atoms treated as hard spheres.[22] $^{EHSS}\Omega$ values have been shown to be accurate for He, but this method fails to provide good quantitative agreement for N_2 and other diatomic gasses.[38] One challenge is that ion-N_2 collisions are likely to be inelastic and diffuse in nature, whereas the EHSS method treats all collisions as elastic and specular.[36] The diffuse hard sphere scattering (DHSS) method incorporates inelastic and diffuse collisions and yields Ω_{N_2} values that are closer to experimental values, especially for larger ions.[39,40] However, the DHSS method overestimates the Ω values of smaller ions such as iodide clusters.[41] The projection superposition approximation (PSA) is computationally inexpensive and was developed to account for shape effects through the incorporation of a shape factor that is unique to each ion.[42] PSA calculations were used to estimate the effect of temperature on Ω

values of ubiquitin ions, which showed that Ω increases with decreasing temperature and that these increases are larger with N_2 than with He.[28] Note that the EHSS, DHSS, and PSA methods do not explicitly account for the effects of charge. Therefore, it is challenging to apply any of these methods universally for all ions.

The trajectory (TJ) method explicitly accounts for the effects of charge and shape by calculating Ω values using the average momentum transfer for many gas molecule trajectories that are simulated using a potential, *U*.[21] In most implementations, *U* for collisions between the gas and each of the *n* atoms in the ion is described using a Lennard-Jones potential, which is repulsive at short distances and attractive at long distances, and an ion-induced dipole potential, which is attractive over all distances:

$$U(r) = 4\sum_i^n \varepsilon_i[(\frac{\sigma_i}{d_i})^{12} - (\frac{\sigma_i}{d_i})^6] - \frac{\alpha}{2}[(\sum_i^n \frac{q_i x_i}{{d_i}^3})^2 + (\sum_i^n \frac{q_i y_i}{{d_i}^3})^2 + (\sum_i^n \frac{q_i z_i}{{d_i}^3})^2] \quad (1)$$

where *r* is the position of the neutral relative to the ion center-of-mass. For each atom, *i*, ε is the depth of the Lennard-Jones potential well, σ is the finite distance at which the Lennard-Jones potential is zero, *d* is the Euclidian distance relative to *r*, α is the polarizability, *q* is the charge, and *x*, *y*, and *z* are the Cartesian coordinates relative to *r*. Alternative functional forms of the interaction potential have also been evaluated and have been reported to improve the accuracy of TJ method calculations.[9,43] Although the benefits of the TJ method have been appreciated since its initial development, its widespread use for predicting Ω values of protein ions was limited by its exceptionally high computational cost. Recent advances in computer hardware and faster implementations of the TJ method[38,44] have made these calculations far more tractable for structural models with many atoms, such as protein ions. Relative to ion-He interactions, ion-N_2 interactions are stronger and more challenging to represent accurately. For example, the highest-accuracy representation of ion-N_2 interactions would include ion-quadrupole interactions and

model N_2 as a linear diatomic molecule. Improved treatments of ion-N_2 interactions have been reported,[37,43,45] albeit with significantly increased computational expense. To the best of our knowledge, these improved treatments have not yet been fully integrated with high-performance implementations of the TJ method.[38,44]

Recently, we characterized the effects of protein charge state, charge distribution, and structure on the Ω_{He} values of ubiquitin (Ub_1), diubiquitin (Ub_2), and alcohol dehydrogenase (ADH) using the TJ method.[46] Over the range of charge states observed experimentally, ion-induced dipole interactions can increase the Ω_{He} values of Ub_1 and Ub_2 by up 13% and 9%, respectively. Ion-induced dipole interactions with He were shown to be important for ions with high *z*, but these effects decreased with ion size, *i.e.*, the magnitude of the increases in Ω_{He} values with increasing *z* were lower for Ub_2 than for Ub_1 and were lower for folded models than α-helical or linear models. Here, we use the TJ method and IM experiments to characterize how these factors affect Ω_{N_2} values and the consequences of interconverting between He- and N_2-based mobility data.

4.3. Methods

4.3.1. Experiments. Ub_1 was purchased from Boston Biochem (Cambridge, MA) and C-to-N terminally linked Ub_2 was purchased from Enzo Life Science (Farmingdale, NY). For denaturing conditions, proteins were dissolved into a 70:30 (v/v) mixture of methanol and water, which was adjusted to pH 2.0 with trifluoroacetic acid. For supercharging conditions, 15% by volume of propylene carbonate[47] was added to the solutions for denaturing conditions. For native-like conditions, Ub_2 was prepared in a solution of aqueous 200 mM ammonium acetate at pH 7. Final protein concentrations ranged from 8.5 to 20 μM. Electrokinetic nanoelectrospray ionization (ESI) was used to generate protein cations, as described previously.[48] IM-MS data was

acquired using a Waters Synapt G2 HDMS (Wilmslow, UK) modified with a radio-frequency confining drift cell that enables the absolute determination of Ω values.[49,50] Most IM experiments were performed using the radio-frequency confining drift cell filled with 0.8 mbar N_2 gas, and arrival-time distributions were converted to Ω distributions as described previously.[50] Note that only N_2 gas is supplied to the radio-frequency confining drift cell in these experiments; this implementation does not use a "helium cell" to assist in ion transfer.[14,50]

4.3.2. Calculations. The protein models used in these calculations have been reported previously.[46] Briefly, the native models were generated by modifying structures deposited into the Protein Data Bank for Ub_1 (1UBQ),[51] Ub_2 (3AXC),[52] and ADH (5ENV).[53] Energy-relaxed models were generated using the native models, the Amber force field, and steepest-descent energy minimization. The α-helical and linear models were built using the expected dihedral angles of the protein backbone. TJ method calculations were performed using IMoS v1.06.[36,39,54] Ω_{TJ} values determined using IMoS are within 1% of those determined using MOBCAL, but require approximately two orders of magnitude less time.[38] Charge distributions were assigned using four methods, which are summarized in Table 1 and have been described previously in detail.[46] Each calculation used 10 000 gas trajectories in 199.98 pascals of N_2 at 301.15 K and the optimized Lennard-Jones parameters reported previously.[45] IMoS parameters for a representative calculation is included in the *Supporting Information*. Confidence intervals were determined at the 95% level using at least 10 independent calculations and *t* statistics. Intervals are reported by placing the uncertainty of the least significant digit(s) of each value in parentheses, *i.e.*, a value of 12.04(11) nm^2 means that the 95% confidence interval spans from 11.93 to 12.15 nm^2.

4.4. Results and Discussion.

Ω values of ions measured with He and N_2 are correlated, which has provided researchers with a means of converting N_2-based mobilities into effective He-based Ω values.[15,55,56] For example, Ω_{He} and Ω_{N_2} values for Ub_1 and Ub_2 generated from denaturing and denaturing-supercharging conditions are shown in Figure 1. Although these values are correlated, deviations are apparent, especially for lowest- and highest-*z* ions observed. These results are consistent with reports of orthogonality between mobilities measured in different drift gases[17,57] and of concerns regarding the potential errors associated with interconverting between He- and N_2-based mobilities.[28] For multiply charged proteins, the relationship between *z* and the accuracy of the correlation between drift gasses has not been established universally. Since most proteins exhibit a wide range of charge states following electrospray ionization, a deeper understanding of the relationship between protein charge state, drift gas, and Ω would be useful for directly comparing results between different IM experiments and drawing accurate conclusions from IM-based structural biology experiments.

In this work, we use a version of the TJ method to characterize how ion charge state, charge distribution, and structure affect the Ω_{N_2} values of protein ions. The $^{TJ}\Omega_{N_2}$ values obtained here are compared directly with previously determined $^{TJ}\Omega_{He}$ values for these ions.[46] Comparing $^{TJ}\Omega_{N_2}$ and $^{TJ}\Omega_{He}$ values provides insights into the differences between mobility information obtained using N_2 and He buffer gasses, and also sheds light on the efficacy of interconverting between Ω_{He} and Ω_{N_2} values. The N_2-based TJ method calculations used here treat N_2 as a polarizable sphere. More accurate treatments of N_2 would increase the computational cost of each calculation (using existing implementations) by several orders of magnitude, which would preclude broad consideration of structure, charge state, and charge distribution. For reference, inclusion of ion-quadrupole interactions increases Ω_{N_2} values of ammonium cations and abiotic

amino acids by less than 3%.[37] Relative contributions from ion-quadrupole interactions are expected to decrease with ion size, therefore omission of those interactions is expected to have a comparatively small effect on the values calculated here for protein ions with kDa masses. Due to uncertainties in the structures of the protein ions probed in the experiments, a range of model geometries (summarized in the *Methods* section) and charge distributions methods (summarized in Table 1) are evaluated.

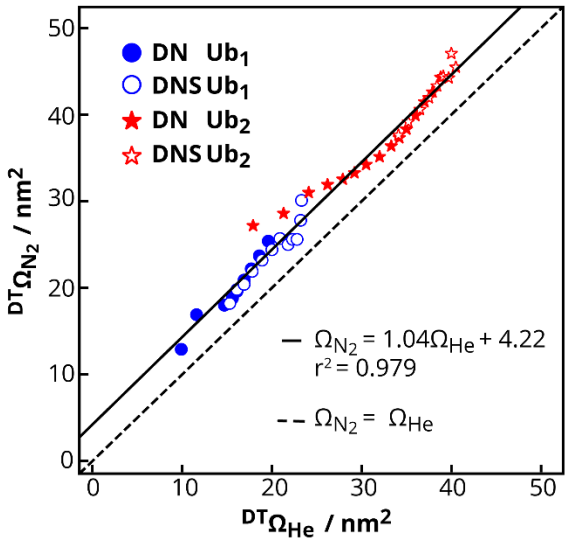

Figure 4.1. Median $^{DT}\Omega_{He}$ and $^{DT}\Omega_{N_2}$ values of ubiquitin (Ub_1, *blue circles*) and diubiquitin (Ub_2, *red stars*) ions from denaturing (DN, *solid markers*) and denaturing-supercharging (DNS, *hollow markers*) condition experiments. The identity and Ω values of these ions are reported in Table S1.

We will first consider an energy-relaxed model of Ub_1, which was initialized using a native structure as described in the *Methods* section, and distribute 1 to 22 excess charges using the Even method. $^{TJ}\Omega_{He}$ and $^{TJ}\Omega_{N_2}$ values are shown in Figure 2A. The $^{TJ}\Omega_{N_2}$ value of 12.04(11)

nm^2 for Ub_1^{1+} is 12% larger than the corresponding $^{TJ}\Omega_{He}$ value of 9.98(6) nm^2. Note that all $^{TJ}\Omega$ values reported in this work are based on at least ten independent calculations. The uncertainty of the least significant digit(s) of each value is placed in parentheses and spans the 95% confidence interval, *i.e.*, a value of 12.04(11) nm^2 means that the 95% confidence interval spans from 11.93 to 12.15 nm^2. $^{TJ}\Omega_{N_2}$ values for Ub_1^{z+} are all similar for $z \leq 4$ ($m/z \geq 2142$) and then increase monotonically by up to 5% with each subsequent increase in z (Figure 2B). By comparison, $^{TJ}\Omega_{He}$ values for Ub_1^{z+} are similar for $z \leq 10$ ($m/z \geq 857$) and then increase by only ~1% with each subsequent increase in z. As the charge state increases from 1+ to 22+, $^{TJ}\Omega_{N_2}$ values increase 91.4% from 12.04(11) to 23.05(5) nm^2. By comparison, over this range of charge states, $^{TJ}\Omega_{He}$ values only increase by approximately 14% from 9.98(6) to 11.44(14) nm^2.

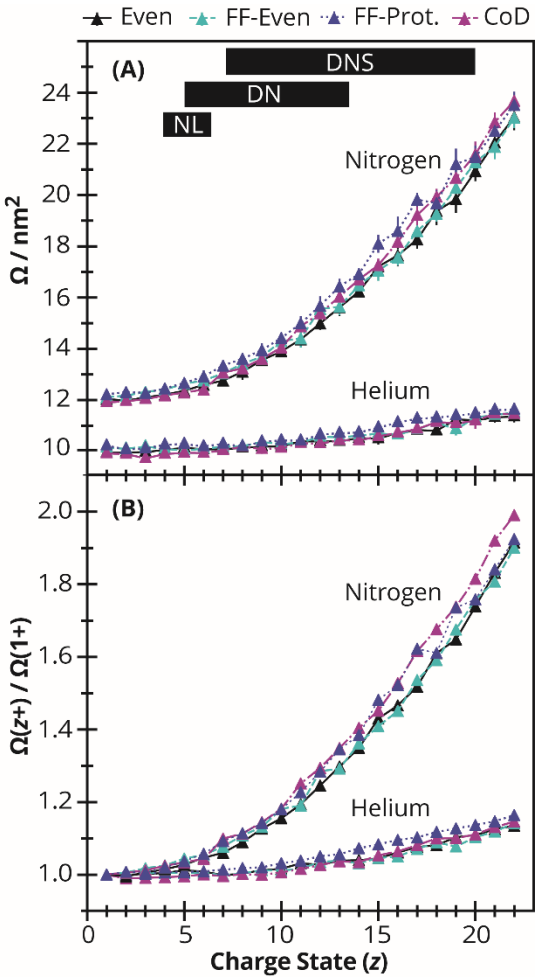

Figure 4.2. (A) $^{TJ}\Omega_{He}$ and $^{TJ}\Omega_{N_2}$ values of the energy-relaxed relaxed model of Ub_1 with charge distributed using the Even, FF-Even, FF-Protonation, and CoD methods (Table 1), as a function of charge state. The horizontal black bars show the typical range of charge states observed for Ub_1 ions generated from native-like (NL), denaturing (DN), and denaturing-supercharging (DNS) conditions. (B) Ratio of the $^{TJ}\Omega$ for a given charge state to that of the corresponding 1+ ion, as a function of charge state.

For context, these $^{TJ}\Omega$ values for Ub_1 in the limit of low charge state are generally consistent with consensus values from native IM-MS community. For example, these $^{TJ}\Omega_{N_2}$ values are similar to the Ω_{N_2} value (12.04 nm^2) reported for Ub_1^{5+} based on measurements using structures for lossless ion manipulations implemented with a constant electrostatic gradient,[58] Ω_{N_2} values for Ub_1^{6+} based on measurements using an electrostatic drift tube (12.00 nm^2)[28] and trapped ion mobility spectrometry (12.17 nm^2),[59] and projection superposition approximation calculations for the N-state of Ub_1 (12.09 nm^2).[28] However, the $^{TJ}\Omega_{N_2}$ values of energy-relaxed Ub_1^{6+} and Ub_1^{7+} are 12.6(3) and 12.8(3) nm^2, respectively, which are 4.6(2)% and 5.9(1.8)% larger than the corresponding value for Ub_1^{1+}. These values are similar to those reported for the compact conformer of Ub_1^{7+} determined using electrostatic drift tube (12.70 nm^2)[28] and trapped ion mobility spectrometry (12.83 nm^2)[59] experiments. These larger $^{TJ}\Omega_{N_2}$ values for modestly charged ions, *i.e.*, charge states that have been associated with native-like structures of Ub_1, illustrate that excess charges can have a significant impact on IM data in the context of structural biology and biophysics.

4.4.1. Origin of Differences between $^{TJ}\Omega_{He}$ and $^{TJ}\Omega_{N_2}$ Values. $^{TJ}\Omega_{N_2}$ values depend more strongly on z than $^{TJ}\Omega_{He}$ values, which can be attributed to stronger ion-induced dipole interactions with N_2 relative to those with He. To visualize the interactions between these gases and a protein ion, Figure 3 shows potentials that were calculated using Equation 1 for energy-relaxed Ub_1 with the charge distributed evenly. As charge is increased from 0 to 20+, the attractive wells of the potentials with each gases deepen and extend farther from the repulsive regions of the potential (colored black), but the changes are much more substantial with N_2. For the 20+ ion, the minimum of the potential with He is −4 kJ mol^{-1}, whereas that with N_2 is −16 kJ mol^{-1}. The attractive region of the potential with N_2 extends several Å farther from the protein

atoms, and as a consequence, more N_2 molecules in the buffer gas will interact with the ion. Glancing collisions will also become more significant, as individual N_2 molecules will experience the attractive potential over a larger portion of their trajectories. These effects all combine to result in greater average momentum transfer with N_2 compared to He.

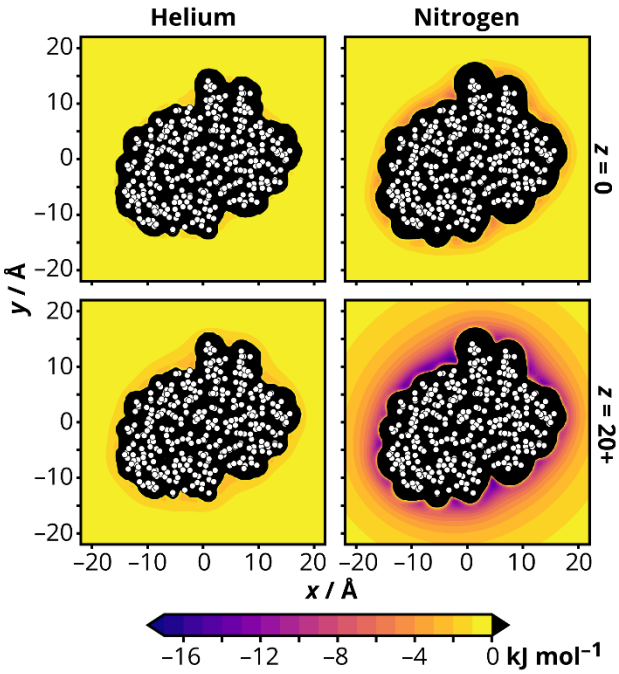

Figure 4.3. The interaction potentials between an energy-relaxed model of Ub_1 with evenly distributed charge and He (*left column*) or N_2 (*right column*). These potentials were calculated using Equation 1. Results for the 0 and 20+ charge states are shown on the *top* and *bottom row*, respectively. These images show the potential on an arbitrary plane whose origin is the center-of-mass of the model. For context, the repulsive regions of each potential are colored black and the *xy* position of each atom within 3 Å of the *xy* plane is plotted using a hollow circle.

The repulsive regions of the potentials with N_2 extend farther from the protein atoms than those for He, which is consistent with the larger size of an N_2 molecule. Furthermore, the edge of the repulsive regions with N_2 are smoother than those with He, which indicates that the larger size of N_2 prevents it from sampling the protein ion surface as finely as He does. This visualization is consistent with a previous report that structural details may be lost when using N_2 relative to using He.[28] Therefore, these results indicate that IM measurements of protein ions in different gasses are sensitive to slightly different properties of the protein ion.

4.4.2. Effects of Charge Distribution Method. In the discussion above, the charge state of the ion was distributed evenly on all atoms in the energy-relaxed model of Ub_1. In order to assess the effects of charge distribution, calculations were also performed using the three alternative charge distribution methods described in Table 1. The resulting absolute and relative $^{TJ}\Omega$ values are shown in Figure 2. Similar $^{TJ}\Omega_{N_2}$ values were found using each of these four methods for each charge state. For example, as charge state increases from 1+ to 22+ the $^{TJ}\Omega_{N_2}$ values increase from 12.0(1) to 23.1(5) nm^2, 12.1(1) to 23.0(3) nm^2, 12.2(1) to 23.5(5) nm^2, and 11.9(1) to 23.7(3) nm^2 using the Even, FF-Even, FF-Protonation, and CoD methods, respectively. The ratio of $^{TJ}\Omega$ at a given charge to the $^{TJ}\Omega$ for Ub_1^{1+} is shown in Figure 2B, which reveals that the $^{TJ}\Omega_{N_2}$ values of the 22+ ions increase by 91(2)%, 90(1)%, 92(2)%, and 99(1)% for the Even, FF-Even, FF-Protonation, and CoD methods, respectively. By comparison, these four charge distribution methods all yielded increases in $^{TJ}\Omega_{He}$ values of approximately 13% over this range of *z*.

The $^{TJ}\Omega_{N_2}$ values for energy-relaxed Ub_1 and energy-relaxed Ub_2 (Figure S1) depend weakly on the charge distribution method for the methods considered here, which is consistent with the previously reported $^{TJ}\Omega_{He}$ values for these protein models.[46] The Even method is easy to

implement because it does not require the calculation or assignment of partial charges. Furthermore, assignment of charges using the FF-Even and FF-Protonation methods may make the calculations more sensitive to small errors in the structure.[46] Therefore, we continue to recommend the Even method in cases where accurate structural models are elusive, *e.g.*, proteins,[46] and will use the Even method for the remaining calculations.

Table 4.1. Description of Charge Distribution Methods

Method	**Description**
Even	All atoms are assigned an equal fraction of the net charge.
FF-Even	Atoms are assigned a partial charge using Amber force field plus an equal fraction of remaining net charge.
FF-Protonation	Atoms are assigned a charge using Amber force field; the net charge is then increased by increasing the partial charge of the most negative atom by one *e*.
CoD	The net charge is then placed at the center of mass of each domain. The partial charge on each atom is zero.

4.4.3. Effects of Size. Although the relative contribution of long-range, ion-neutral interactions to Ω values becomes less significant with increasing molecular size,[9,21,23] the relative magnitude of these contributions to Ω_{N_2} as a function of z and ion size remains unclear. To probe the relationship between ion size, z, and Ω_{N_2}, we will directly compare the ${}^{TJ}\Omega_{N_2}$ values of Ub_1 and Ub_2. Over the full range of charge states used in the TJ method calculations, the ${}^{TJ}\Omega_{N_2}$ values for energy-relaxed Ub_1 ($1 \leq z \leq 22$) increased by 91(2)%, whereas those for Ub_2 ($1 \leq z \leq 30$) increased by 66(1)% (Figure S1). If only the charge states observed in IM experiments (including charge states observed from native, denaturing, and supercharging solution conditions) for Ub_1 ($6 \leq z \leq 18$) and Ub_2 ($9 \leq z \leq 26$) are considered, the increase in ${}^{TJ}\Omega_{N_2}$ for Ub_1, 54(2)%, is larger than that of Ub_2, 39(5)%. Comparing the results for Ub_1 and Ub_2 of a given z, the contributions of the ion-induced dipole interactions on Ω are diminished for Ub_2 as a direct result of the increased size of the ion, which is consistent with previous findings using carbon cluster ions.[28] For comparison, ${}^{TJ}\Omega_{N_2}$ values were also calculated for a native model of alcohol dehydrogenase (ADH), which is significantly larger than Ub_1 or Ub_2. As shown in Figure S2, the ${}^{TJ}\Omega_{N_2}$ values increase by 82(2)% from 90.3 to 164.7 nm^2 as the charge state increases from 1+ to 150+. To directly compare the results for Ub_1, Ub_2, and ADH, Figure S3 shows the ratio of the ${}^{TJ}\Omega_{N_2}$ and ${}^{TJ}\Omega_{He}$ values for each charge state to the value for the corresponding 1+ ion, as a function of m/z. These results show that although all three proteins yield ratios near one at high m/z and ratios that increase with decreasing m/z, the increase is protein dependent.

4.4.4. Effects of Structure. The structures of protein ions in the gas phase can depend on their charge state and the solutions from which they were generated.[19,60–62] For example, cation-to-anion proton transfer reactions of Ub_1[63] and lysozyme[64] ions generated from denaturing conditions yield products that can undergo significant compaction following charge reduction.

To probe the relationship between structure, z, and Ω_{N_2}, four structural models each for Ub_1 and Ub_2 were investigated. Figure 4 shows the $^{TJ}\Omega$ values of the native, energy-relaxed, α-helical, and linear models of Ub_1 with He and N_2. $^{TJ}\Omega$ values increase with z for each structural model and buffer gas, and each $^{TJ}\Omega_{N_2}$ value is systematically larger than the corresponding $^{TJ}\Omega_{He}$ value. For each structural model, the $^{TJ}\Omega_{N_2}$ value of Ub_1^{1+} is larger than the $^{TJ}\Omega_{He}$ value of Ub_1^{22+}. Ω is lowest for the energy-relaxed model, and increases concomitantly with surface area of the model in the order of the native, the α-helical, and then the linear model. The 91(1)% and 79(2)% increases in $^{TJ}\Omega_{N_2}$ values for the energy-relaxed and native models, respectively, are far larger than the 27(1)% and 9(1)% increases for the α-helical and linear models, respectively. This trend is attributed to greater charge density in the energy-relaxed and native models compared to the more extended α-helical and linear models. Similar to Ub_1, the magnitude of the increases in $^{TJ}\Omega_{N_2}$ with z for Ub_2 depend on the charge density of the model. The $^{TJ}\Omega_{N_2}$ values of the energy-relaxed and native models of Ub_2 increased by 66(1)% and 52(1)%, whereas those for the α-helical and linear models increased by only 11(2)% and 3(1)%, respectively (Figure S4).

Overall, the increases in Ω values with increasing z are much more substantial with N_2 than with He, which is attributed to differences in ion-induced dipole interactions as described above. An extreme example of this effect is that the $^{TJ}\Omega_{N_2}$ values of the native model of Ub_1 exceeds those of the corresponding α-helical models for $z > 30+$ (Figure S5), which was not observed with He. These findings demonstrate that increases in ion-induced dipole interactions at high charge states can significantly increase Ω_{N_2}, even without changes in ion structure. By comparison, ion-induced dipole interactions with He can account for only a comparatively small fraction of the observed increase in Ω without additional changes in structure.

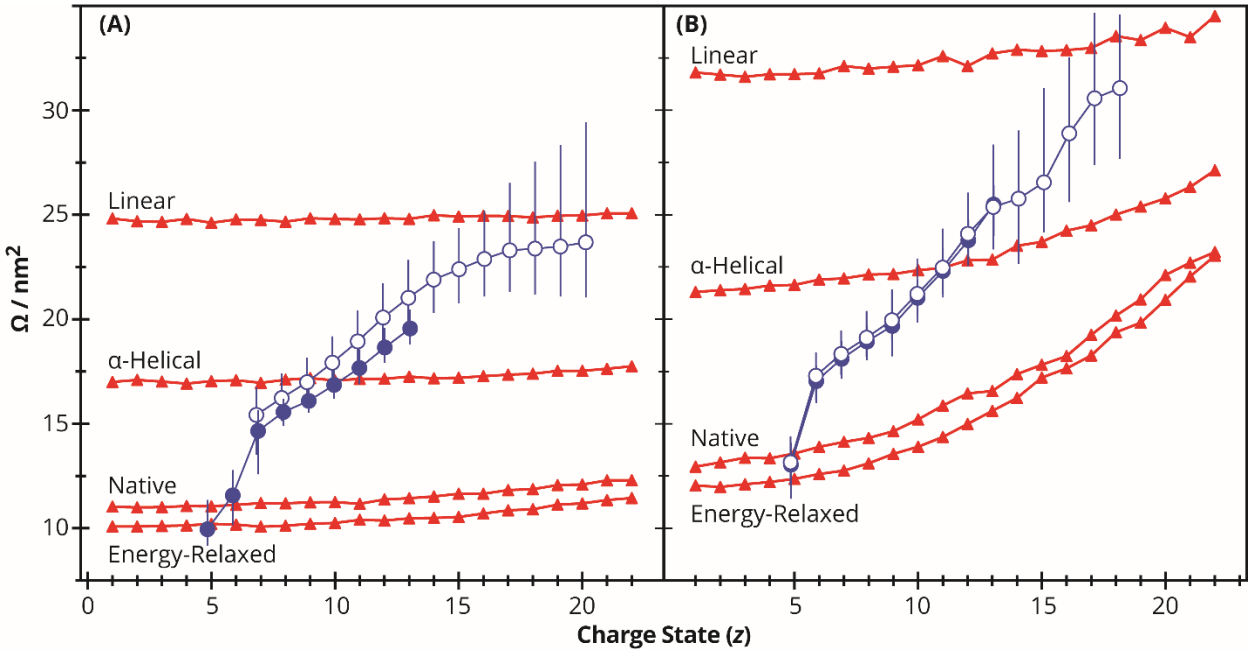

Figure 4.4. $^{TJ}\Omega_{He}$ (A) and $^{TJ}\Omega_{N_2}$ (B) values calculated using the Even method for all four structural models (*red triangles*). $^{DT}\Omega_{He}$ (A) and $^{DT}\Omega_{N_2}$ (B) values for Ub_1 ions generated from denaturing (*solid blue circles*) and denaturing-supercharging (*hollow blue circles*) conditions. The markers indicate the median of the apparent $^{DT}\Omega$ distribution and the bars span from 10% to 90% of the cumulative distribution function of the apparent $^{DT}\Omega$ distribution, as described previously.[69] All Ω_{He} values were reported previously.[46]

4.4.5. Structures of Protein Ions in the Gas Phase: Comparison of $^{TJ}\Omega$ and $^{DT}\Omega$. The mobilities of Ub_1 ions generated from denaturing and denaturing-supercharging conditions were measured using a radio-frequency confining drift cell. The resulting $^{DT}\Omega_{He}$ and $^{DT}\Omega_{N_2}$ values are plotted in Figures 4A and 4B, respectively. With increasing *z* (5+ to 13+) for Ub_1 from denaturing conditions, $^{DT}\Omega_{He}$ values increased from 9.9 to 19.6 nm^2 and $^{TJ}\Omega_{N_2}$ values increased from 13.5 to 25.4 nm^2. The respective increases for Ub_1 ions (7+ to 18+) from denaturing-

supercharging conditions were 15.3 to 23.3 nm^2 and 18.2 to 30.1 nm^2. Although experiments with both gases used identical solutions and electrospray conditions, different ranges of charge states and extents of contributions from compact structures were observed. These differences are attributed to inadvertent collisional activation of ions during transfer into the N_2-filled drift cell (collisions with N_2 molecules result in more efficient conversion of energy from translational to internal degrees of freedom) and persisted even after the pressure in the drift cell was reduced to 0.8 mbar for N_2 (relative to 2 mbar for He). Alternatively, proton transfer to the N_2 gas in the drift cell would also explain the absence of the highest-*z* ions in the N_2-based IM experiments. That explanation is consistent with the recent observation of proton transfer from highly charged cytochrome *c* and myoglobin cations to N_2 molecules, which was attributed to the exceptionally weak gas-phase basicities of those highly charged protein ions.[65]

The $^{DT}\Omega_{He}$ and $^{DT}\Omega_{N_2}$ values for Ub_1^{5+} and the $^{DT}\Omega_{He}$ value for Ub_1^{6+} are all similar to the corresponding $^{TJ}\Omega$ values calculated for the native and energy-relaxed models. In contrast, the $^{DT}\Omega_{N_2}$ value for Ub_1^{6+} is significantly larger than the corresponding $^{TJ}\Omega$ values calculated for the native and energy-relaxed models and $^{DT}\Omega_{N_2}$ values reported previously for native-like Ub_1.[28,58] This difference is consistent with Ub_1^{6+} generated in these experiments undergoing collision-induced unfolding[66] during transfer into the N_2-filled drift cell. The $^{DT}\Omega_{He}$ and $^{DT}\Omega_{N_2}$ values both increase with increasing *z*. For $z \approx 10$, the $^{DT}\Omega$ values are similar to the corresponding $^{TJ}\Omega$ values for the α-helical model. At the highest *z*, the $^{DT}\Omega$ values are similar to but smaller than the $^{TJ}\Omega$ values for the linear model. Analogous data for Ub_2 are shown in Figure S4. Although there are several similarities between the results for Ub_1 and Ub_2, compact ions were not observed in the nitrogen-based experiments. The $^{DT}\Omega$ values are similar to the corresponding $^{TJ}\Omega$ values for the α-helical model for $z \approx 18$ and $z \approx 24$ with He and N_2, respectively. The $^{DT}\Omega$ values for the

highest-*z* ions are smaller than the $^{TJ}\Omega$ values for the linear model, which suggests that interactions between neighboring residues persist even for the highest-*z* ions.

4.4.6. Relationship between Ω_{He} and Ω_{N2}. As discussed in the *Introduction*, the results of many experiments that probe mobilities in N_2 are interpreted via calibration with standards of known Ω_{He}. This process yields effective Ω_{He} values and assumes a functional relationship between the mobilities in each gas, and that the relationship is the same for all analytes and standards in the analysis. We will evaluate this relationship based on the experiments and calculations reported here.

One method for visualizing this relationship is to plot K_{0,N_2}^{-1} values as a function of the corresponding $K_{0,He}^{-1}$ values. This approach has been used for a large set of peptide, protein, and protein complex ions,[25] as well as sets of peptide[26] and carbohydrate ions.[67] The correlation between experimentally determined $K_{0,He}^{-1}$ and K_{0,N_2}^{-1} values for Ub_1 and Ub_2 from denaturing and denaturing-supercharging conditions (Figure 5) is reasonably linear, but the coefficient of determination ($r^2 = 0.797$) is significantly smaller than that reported previously for the large set of peptide, protein, and protein complex ions (0.991).[25] The origin of this discrepancy appears to be systematic differences between the results for Ub_1 and Ub_2 as well as between the results for ions from denaturing and denaturing-supercharging conditions, which are all included in the present study. Figure 5 shows that the relationship between $K_{0,He}^{-1}$ and K_{0,N_2}^{-1} for Ub_1 from denaturing conditions in this study is similar to that for Ub_1, cytochrome *c* (12.6 kDa), and myoglobin (16.7 kDa) from denaturing conditions in the previous study.[26] However, those values are shifted from those for Ub_1 from denaturing-supercharging conditions in this study, as well as from Ub_2 under both solution conditions. For comparison, the linear regression of the $K_{0,He}^{-1}$ and K_{0,N_2}^{-1} values that were reported previously[25] is also shown in Figure 5. That linear function,

which was determined using the larger set of ions that included peptides, native-like proteins, and native-like protein complexes, deviates from the present experimental data with decreasing mobility. Although calibration errors can be minimized by careful selection of standards,[25,26,68,27] these results highlight the challenge of estimating the errors of interconverting between He- and N_2-based mobilities *a priori.*

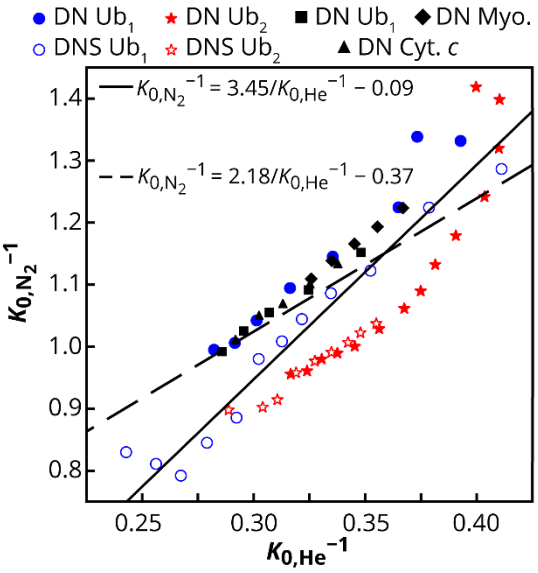

Figure 4.5. Experimentally determined reciprocal reduced mobilities of Ub_1 (*blue circles*) and Ub_2 (*red stars*) in He and N_2 under denaturing (DN, *solid markers*) and denaturing-supercharging (DNS, *hollow markers*) conditions. Previously reported data for Ub_1, myoglobin, and cytochrome *c* are shown using *black markers* as indicated.[25] The *solid line* corresponds to line of best fit for all experiments from this work ($r^2 = 0.797$). The *dashed line* is a line of best fit reported previously ($r^2 = 0.991$ for the data in that work).[25]

Figure 6A shows K_{0,N_2}^{-1} values plotted as a function of $K_{0,He}^{-1}$ values for the four structural models of Ub_1 and Ub_2, based on these TJ method calculations. Note that this plot

contains data for the full range of charge states observed experimentally in both gases for Ub_1 ($6 \leq z \leq 18$) and Ub_2 ($9 \leq z \leq 26$); data for a wider range of charge states will be discussed below. This data is correlated linearly ($r^2 = 0.997$), but the relative differences between the data and the best-fit line (Figure 6B) span from approximately −6% to +15% and the root-mean-square of the relative differences (%RMSD) is 5.6%. The %RMSD values were greatest for the native (9.1%) and energy-relaxed (5.4%) models, and smallest for the α-helical models (2.1%). An analogous plot that includes all calculated mobilities for Ub_1 ($1 \leq z \leq 22$) and Ub_2 ($1 \leq z \leq 30$) in both gases is shown in Figure S6A. Although the data in the larger set has a larger coefficient of determination ($r^2 = 0.9989$), the results for the highest-z ions exhibit significant positive deviation and relative differences of up to 30% (Figure S6B).

The results in Figures 5 and 6 show that many factors can affect the relationship between $K_{0,He}^{-1}$ and K_{0,N_2}^{-1}, even for closely related species (in this case, Ub_1 and Ub_2 ions). As a consequence, it is challenging to estimate the errors associated with interconverting between He- and N_2-based mobilities without extensive characterization in both gases (as done here), even with well-paired analytes and standards. Therefore, we recommend that this practice be avoided as this study indicates that there are not yet effective ways to integrate this additional uncertainty into the reporting of effective Ω values.

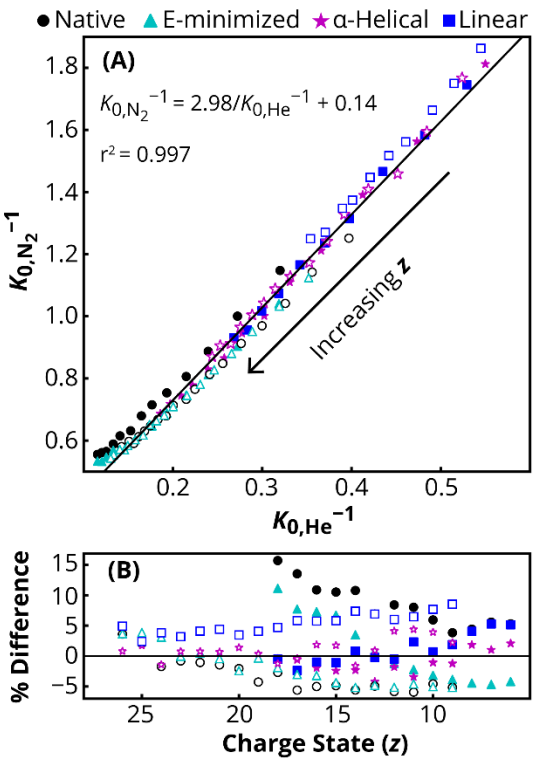

Figure 4.6. (A) Reciprocal reduced mobilities in He and in N_2 for the four structural models of Ub_1 (*solid markers*) and Ub_2 (*hollow markers*), with charge distributed using the Even method. The black line was determined by linear regression of all data plotted. (B) Relative percent difference between K_{0,N_2}^{-1} and the best-fit line in (A) as a function of z. This analysis was limited to ions whose charge states were observed experimentally, *i.e.*, from the lowest-z ions observed from native-like conditions to the highest-z ions observed from denaturing, supercharging conditions. Results for the analysis that considers data for all charge states considered computationally is shown in Figure S6.

4.5. Conclusions

The effects of charge state, charge distribution, ion size, and ion structure on Ω_{N_2} values were evaluated using TJ method calculations and IM experiments. Ω_{N_2} values of Ub_1 and Ub_2

were compared with Ω_{He} values reported previously in order to characterize the correlation between mobility information obtained in both buffer gasses. When excess charges are added to an energy-minimized model of Ub_1, $^{TJ}\Omega_{N_2}$ values increased significantly more than the corresponding $^{TJ}\Omega_{He}$ values. $^{TJ}\Omega_{N_2}$ values increased by ~90% as charge was added from 1+ to 22+, but over this same range of *z*, $^{TJ}\Omega_{He}$ values only increased by approximately 13% (Figure 2). The larger increase for the N_2-based values is attributed to stronger ion-induced dipole interactions with that gas. The potential of the ion-N_2 interaction, calculated as the sum of Lennard-Jones and ion-induced dipole interactions, is much more attractive and extends much farther from the protein atoms than the corresponding ion-He potential (Figure 3). These differences result in more collisions and efficient momentum transfer during collision with N_2. Most significantly to structural biology experiments, ion-induced dipole interactions between N_2 and energy-relaxed Ub_1 ions with native-like charge states, *i.e.*, 6+ and 7+, result in $^{TJ}\Omega_{N_2}$ values that are 4.6% and 5.9% larger than the corresponding 1+ ion. Therefore, not treating charge explicitly in calculations can result in large errors even for modestly charged ions. Additionally, these experiments investigated protein cations, but similar changes to Ω_{N_2} values due to ion-induced dipole interactions are expected as charge increases on protein anions.

Three additional structural models of each protein were constructed and all yielded increasing $^{TJ}\Omega_{N_2}$ values with increasing *z*, but the magnitude of the increase depended on the model. Over the same range of evenly distributed *z* (1+ to 22+), $^{TJ}\Omega_{N_2}$ values increased by 91(1)% ,79(2)%, 27(1)%, and 9(1)% for the energy-relaxed, native, α-helical, and linear models of Ub_1, respectively (Figure 4). This further demonstrates that the contributions from ion-induced dipole interactions decrease with decreasing charge density. Experimental $^{DT}\Omega$ values of Ub_1 and Ub_2 also increased with increasing *z*. For Ub_1, $^{DT}\Omega_{N_2}$ values of the lowest-*z* ions are similar to the

$^{TJ}\Omega_{N_2}$ values for the native and energy-relaxed models. Concomitant with increasing z, $^{DT}\Omega_{N_2}$ values rapidly approach the $^{TJ}\Omega_{N_2}$ values for the α-helical structure and nearly converge with $^{TJ}\Omega_{N_2}$ values for the linear structure for the highest-z ions observed experimentally (Figure 4). These results all indicate that charge state affects Ω_{N_2}, even in the absence of structural changes. Therefore, we continue to recommend the TJ method when interpreting results from IM for applications in structural biology and biophysics, particularly when N_2 is used as the drift gas.

The relationship between $K_{0,He}^{-1}$ and K_{0,N_2}^{-1} values were assessed using experimental results and TJ method calculations. The strength of the correlation between $^{DT}K_{0,He}^{-1}$ and $^{DT}K_{0,N_2}^{-1}$ values of Ub_1 and Ub_2 was weakened as a result of deviations for ions with the highest and lowest charge states and for ions generated from different solution conditions (Figure 5). Similarly, deviations from the correlation between $^{TJ}K_{0,He}^{-1}$ and $^{TJ}K_{0,N_2}^{-1}$ values were observed for lowest and highest charge states of Ub_1 and Ub_2 , with relative errors as high as 30% from the line of best fit (Figure S6). For the charge states of Ub_1 and Ub_2 observed experimentally (Figure 6), the %RMSD between $^{TJ}K_{0,He}^{-1}$ and $^{TJ}K_{0,N_2}^{-1}$ values are as high as 9.1%. Together, these findings indicate that the relationship between N_2- and He-based mobilities depend on charge state, ion shape, and size, and that it is difficult to estimate the errors associated with interconverting between mobilities in different gases. Therefore, we recommend reporting Ω and K values that are based on the predominant gas present in the separation and applying additional caution when comparing results from mobility experiments performed using different gases.

4.6. Acknowledgments. Acknowledgment is made to Eli Lilly and Company (Young Investigator Award in Analytical Chemistry to M. F. B.) and the Donors of the American Chemical Society Petroleum Research Fund for support of this research.

4.7. Supporting Information. Table S1, Figures S1 to S6, and a representative parameters file for IMoS.

4.8. References

(1) May, J. C.; McLean, J. A. Ion Mobility-Mass Spectrometry: Time-Dispersive Instrumentation. *Anal. Chem.* **2015**, *87*, 1422–1436.

(2) Campuzano, I. D.; Lippens, J. L. Ion Mobility in the Pharmaceutical Industry: An Established Biophysical Technique or Still Niche? *Curr. Opin. Chem. Biol.* **2018**, *42*, 147–159.

(3) Trimpin, S.; Plasencia, M.; Isailovic, D.; Clemmer, D. E. Resolving Oligomers from Fully Grown Polymers with IMS−MS. *Anal. Chem.* **2007**, *79*, 7965–7974.

(4) Katzenmeyer, B. C.; Hague, S. F.; Wesdemiotis, C. Multidimensional Mass Spectrometry Coupled with Separation by Polarity or Shape for the Characterization of Sugar-Based Nonionic Surfactants. *Anal. Chem.* **2016**, *88*, 851–857.

(5) Harkness, K. M.; Fenn, L. S.; Cliffel, D. E.; McLean, J. A. Surface Fragmentation of Complexes from Thiolate Protected Gold Nanoparticles by Ion Mobility-Mass Spectrometry. *Anal. Chem.* **2010**, *82*, 3061–3066.

(6) Gaye, M. M.; Nagy, G.; Clemmer, D. E.; Pohl, N. L. B. Multidimensional Analysis of 16 Glucose Isomers by Ion Mobility Spectrometry. *Anal. Chem.* **2016**, *88*, 2335–2344.

(7) Huang, Y.; Dodds, E. D. Ion Mobility Studies of Carbohydrates as Group I Adducts: Isomer Specific Collisional Cross Section Dependence on Metal Ion Radius. *Anal. Chem.* **2013**, *85*, 9728–9735.

(8) D'Atri, V.; Porrini, M.; Rosu, F.; Gabelica, V. Linking Molecular Models with Ion Mobility Experiments. Illustration with a Rigid Nucleic Acid Structure. *J. Mass Spectrom.* **2015**, *50*, 711–726.

(9) Lee, J. W.; Davidson, K. L.; Bush, M. F.; Kim, H. I. Collision Cross Sections and Ion Structures: Development of a General Calculation Method via High-Quality Ion Mobility Measurements and Theoretical Modeling. *Analyst* **2017**, *142*, 4289–4298.

(10) Uetrecht, C.; Barbu, I. M.; Shoemaker, G. K.; van Duijn, E.; Heck, A. J. R. Interrogating Viral Capsid Assembly with Ion Mobility–Mass Spectrometry. *Nat. Chem.* **2011**, *3*, 126–132.

(11) Pacholarz, K. J.; Porrini, M.; Garlish, R. A.; Burnley, R. J.; Taylor, R. J.; Henry, A. J.; Barran, P. E. Dynamics of Intact Immunoglobulin G Explored by Drift-Tube Ion-Mobility Mass Spectrometry and Molecular Modeling. *Angew. Chem. Int. Ed.* **2014**, *53*, 7765–7769.

(12) Song, Y.; Nelp, M. T.; Bandarian, V.; Wysocki, V. H. Refining the Structural Model of a Heterohexameric Protein Complex: Surface Induced Dissociation and Ion Mobility Provide Key Connectivity and Topology Information. *ACS Cent. Sci.* **2015**, *1*, 477–487.

(13) Zhao, Y.; Singh, A.; Xu, Y.; Zong, C.; Zhang, F.; Boons, G.-J.; Liu, J.; Linhardt, R. J.; Woods, R. J.; Amster, I. J. Gas-Phase Analysis of the Complex of Fibroblast GrowthFactor 1 with Heparan Sulfate: A Traveling Wave Ion Mobility Spectrometry (TWIMS) and Molecular Modeling Study. *J. Am. Soc. Mass Spectrom.* **2017**, *28*, 96–109.

(14) Giles, K.; Williams, J. P.; Campuzano, I. Enhancements in Travelling Wave Ion Mobility Resolution. *Rapid Commun. Mass Spectrom.* **2011**, *25*, 1559–1566.

(15) May, J. C.; Goodwin, C. R.; Lareau, N. M.; Leaptrot, K. L.; Morris, C. B.; Kurulugama, R. T.; Mordehai, A.; Klein, C.; Barry, W.; Darland, E.; et al. Conformational Ordering of Biomolecules in the Gas Phase: Nitrogen Collision Cross Sections Measured on a Prototype High Resolution Drift Tube Ion Mobility-Mass Spectrometer. *Anal. Chem.* **2014**, *86*, 2107–2116.

(16) Silveira, J. A.; Ridgeway, M. E.; Laukien, F. H.; Mann, M.; Park, M. A. Parallel Accumulation for 100% Duty Cycle Trapped Ion Mobility-Mass Spectrometry. *Int. J. Mass Spectrom.* **2017**, *413*, 168–175.

(17) Davidson, K. L.; Bush, M. F. Effects of Drift Gas Selection on the Ambient-Temperature, Ion Mobility Mass Spectrometry Analysis of Amino Acids. *Anal. Chem.* **2017**, *89*, 2017–2023.

(18) Jarrold, M. F. Unfolding, Refolding, and Hydration of Proteins in the Gas Phase. *Acc. Chem. Res.* **1999**, *32*, 360–367.

(19) Bohrer, B. C.; Merenbloom, S. I.; Koeniger, S. L.; Hilderbrand, A. E.; Clemmer, D. E. Biomolecule Analysis by Ion Mobility Spectrometry. *Annu. Rev. Anal. Chem.* **2008**, *1*, 293–327.

(20) Bowers, M. T. Ion Mobility Spectrometry: A Personal View of Its Development at UCSB. *Int. J. Mass Spectrom.* **2014**, *370*, 75–95.

(21) Mesleh, M. F.; Hunter, J. M.; Shvartsburg, A. A.; Schatz, G. C.; Jarrold, M. F. Structural Information from Ion Mobility Measurements: Effects of the Long-Range Potential. *J. Phys. Chem.* **1996**, *100*, 16082–16086.

(22) Shvartsburg, A. A.; Jarrold, M. F. An Exact Hard-Spheres Scattering Model for the Mobilities of Polyatomic Ions. *Chem. Phys. Lett.* **1996**, *261*, 86–91.

(23) Wyttenbach, T.; Bleiholder, C.; Bowers, M. T. Factors Contributing to the Collision Cross Section of Polyatomic Ions in the Kilodalton to Gigadalton Range: Application to Ion Mobility Measurements. *Anal. Chem.* **2013**, *85*, 2191–2199.

(24) Ruotolo, B. T.; Benesch, J. L. P.; Sandercock, A. M.; Hyung, S.-J.; Robinson, C. V. Ion Mobility–Mass Spectrometry Analysis of Large Protein Complexes. *Nat. Protoc.* **2008**, *3*, 1139–1152.

(25) Bush, M. F.; Hall, Z.; Giles, K.; Hoyes, J.; Robinson, C. V.; Ruotolo, B. T. Collision Cross Sections of Proteins and Their Complexes: A Calibration Framework and Database for Gas-Phase Structural Biology. *Anal. Chem.* **2010**, *82*, 9557–9565.

(26) Bush, M. F.; Campuzano, I. D. G.; Robinson, C. V. Ion Mobility Mass Spectrometry of Peptide Ions: Effects of Drift Gas and Calibration Strategies. *Anal. Chem.* **2012**, *84*, 7124–7130.

(27) Gelb, A. S.; Jarratt, R. E.; Huang, Y.; Dodds, E. D. A Study of Calibrant Selection in Measurement of Carbohydrate and Peptide Ion-Neutral Collision Cross Sections by Traveling Wave Ion Mobility Spectrometry. *Anal. Chem.* **2014**, *86*, 11396–11402.

(28) Bleiholder, C.; Johnson, N. R.; Contreras, S.; Wyttenbach, T.; Bowers, M. T. Molecular Structures and Ion Mobility Cross Sections: Analysis of the Effects of He and N_2 Buffer Gas. *Anal. Chem.* **2015**, *87*, 7196–7203.

(29) Hofmann, J.; Struwe, W. B.; Scarff, C. A.; Scrivens, J. H.; Harvey, D. J.; Pagel, K. Estimating Collision Cross Sections of Negatively Charged N-Glycans Using Traveling Wave Ion Mobility-Mass Spectrometry. *Anal. Chem.* **2014**, *86*, 10789–10795.

(30) Young, M. N.; Bleiholder, C. Molecular Structures and Momentum Transfer Cross Sections: The Influence of the Analyte Charge Distribution. *J. Am. Soc. Mass Spectrom.* **2017**, *28*, 619–627.

(31) Hines, K. M.; May, J. C.; McLean, J. A.; Xu, L. Evaluation of Collision Cross Section Calibrants for Structural Analysis of Lipids by Traveling Wave Ion Mobility-Mass Spectrometry. *Anal. Chem.* **2016**, *88*, 7329–7336.

(32) Richardson, K.; Langridge, D.; Giles, K. Fundamentals of Travelling Wave Ion Mobility Revisited: I. Smoothly Moving Waves. *Int. J. Mass Spectrom.* **2018**, *428*, 71–80.

(33) Haynes, W.; Lide, D.; Bruno, T. J. *CRC Handbook of Chemistry and Physics 2012-2013*, 93 ed.; CRC Press: Boca Raton, FL, 2012.

(34) Mason, E.; McDaniel. *Transport Properties of Ions in Gases*; John Wiley and Sons, Inc.: New York, New York, 1988.

(35) Beegle, L. W.; Kanik, I.; Matz, L.; Hill, H. H. Effects of Drift-Gas Polarizability on Glycine Peptides in Ion Mobility Spectrometry. *Int. J. Mass Spectrom.* **2002**, *216*, 257–268.

(36) Larriba-Andaluz, C.; Fernández-García, J.; Ewing, M. A.; Hogan, C. J.; Clemmer, D. E. Gas Molecule Scattering & Ion Mobility Measurements for Organic Macro-Ions in He versus N_2 Environments. *Phys. Chem. Chem. Phys.* **2015**, *17*, 15019–15029.

(37) Kim, H.; Kim, H. I.; Johnson, P. V.; Beegle, L. W.; Beauchamp, J. L.; Goddard, W. A.; Kanik, I. Experimental and Theoretical Investigation into the Correlation between Mass and Ion Mobility for Choline and Other Ammonium Cations in N_2. *Anal. Chem.* **2008**, *80*, 1928–1936.

(38) Shrivastav, V.; Nahin, M.; Hogan, C. J.; Larriba-Andaluz, C. Benchmark Comparison for a Multi-Processing Ion Mobility Calculator in the Free Molecular Regime. *J. Am. Soc. Mass Spectrom.* **2017**, *28*, 1540–1551.

(39) Larriba, C.; Hogan, C. J. Ion Mobilities in Diatomic Gases: Measurement versus Prediction with Non-Specular Scattering Models. *J. Phys. Chem. A* **2013**, *117*, 3887–3901.

(40) Larriba, C.; Hogan, C. J. Free Molecular Collision Cross Section Calculation Methods for Nanoparticles and Complex Ions with Energy Accommodation. *J. Comput. Phys.* **2013**, *251*, 344–363.

(41) Ouyang, H.; Larriba-Andaluz, C.; Oberreit, D. R.; Hogan, C. J. The Collision Cross Sections of Iodide Salt Cluster Ions in Air via Differential Mobility Analysis-Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **2013**, *24*, 1833–1847.

(42) Bleiholder, C.; Wyttenbach, T.; Bowers, M. T. A Novel Projection Approximation Algorithm for the Fast and Accurate Computation of Molecular Collision Cross Sections (I). Method. *Int. J. Mass Spectrom.* **2011**, *308*, 1–10.

(43) Lee, J. W.; Lee, H. H. L.; Davidson, K. L.; Bush, M. F.; Kim, H. I. Structural Characterization of Small Molecular Ions by Ion Mobility Mass Spectrometry in Nitrogen Drift Gas: Improving the Accuracy of Trajectory Method Calculations. *Analyst* **2018**, *143*, 1786–1796.

(44) Ewing, S. A.; Donor, M. T.; Wilson, J. W.; Prell, J. S. Collidoscope: An Improved Tool for Computing Collisional Cross-Sections with the Trajectory Method. *J. Am. Soc. Mass Spectrom.* **2017**, *28*, 587–596.

(45) Campuzano, I.; Bush, M. F.; Robinson, C. V.; Beaumont, C.; Richardson, K.; Kim, H.; Kim, H. I. Structural Characterization of Drug-like Compounds by Ion Mobility Mass

Spectrometry: Comparison of Theoretical and Experimentally Derived Nitrogen Collision Cross Sections. *Anal. Chem.* **2012**, *84*, 1026–1033.

(46) Laszlo, K. J.; Bush, M. F. Effects of Charge State, Charge Distribution, and Structure on the Ion Mobility of Protein Ions in Helium Gas: Results from Trajectory Method Calculations. *J. Phys. Chem. A* **2017**, *121*, 7768–7777.

(47) 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.* **2015**, *87*, 3973–3980.

(48) Davidson, K. L.; Oberreit, D. R.; Hogan, C. J.; Bush, M. F. Nonspecific Aggregation in Native Electrokinetic Nanoelectrospray Ionization. *Int. J. Mass Spectrom.* **2017**, *420*, 35–42.

(49) Allen, S. J.; Bush, M. F. Radio-Frequency (rf) Confinement in Ion Mobility Spectrometry: Apparent Mobilities and Effective Temperatures. *J. Am. Soc. Mass Spectrom.* **2016**, *27*, 2054–2063.

(50) Allen, S. J.; Giles, K.; Gilbert, T.; Bush, M. F. Ion Mobility Mass Spectrometry of Peptide, Protein, and Protein Complex Ions Using a Radio-Frequency Confining Drift Cell. *Analyst* **2016**, *141*, 884–891.

(51) Vijay-Kumar, S.; Bugg, C. E.; Cook, W. J. Structure of Ubiquitin Refined at 1.8 Å Resolution. *J. Mol. Biol.* **1987**, *194*, 531–544.

(52) Rohaim, A.; Kawasaki, M.; Kato, R.; Dikic, I.; Wakatsuki, S. Structure of a Compact Conformation of Linear Diubiquitin. *Acta Crystallogr. D Biol. Crystallogr.* **2012**, *68*, 102–108.

(53) Plapp, B. V.; Charlier, H. A.; Ramaswamy, S. Mechanistic Implications from Structures of Yeast Alcohol Dehydrogenase Complexed with Coenzyme and an Alcohol. *Arch. Biochem. Biophys.* **2016**, *591*, 35–42.

(54) Larriba-Andaluz, C.; Hogan, C. J. Collision Cross Section Calculations for Polyatomic Ions Considering Rotating Diatomic/Linear Gas Molecules. *J. Chem. Phys.* **2014**, *141*, 194107.

(55) Ridenour, W. B.; Kliman, M.; McLean, J. A.; Caprioli, R. M. Structural Characterization of Phospholipids and Peptides Directly from Tissue Sections by MALDI Traveling-Wave Ion Mobility-Mass Spectrometry. *Anal. Chem.* **2010**, *82*, 1881–1889.

(56) Knapman, T. W.; Berryman, J. T.; Campuzano, I.; Harris, S. A.; Ashcroft, A. E. Considerations in Experimental and Theoretical Collision Cross-Section Measurements of Small Molecules Using Travelling Wave Ion Mobility Spectrometry-Mass Spectrometry. *Int. J. Mass Spectrom.* **2010**, *298*, 17–23.

(57) Lalli, P. M.; Yuri E. Corilo; Maíra Fasciotti; Riccio Maria Francesca; Gilberto F. Sa; Romeu J. Daroda; Gustavo H. M. F. Souza; Michael McCullagh; Michael D. Bartberger; Marcos N. Eberlin; et al. Baseline Resolution of Isomers by Traveling Wave Ion Mobility Mass Spectrometry: Investigating the Effects of Polarizable Drift Gases and Ionic Charge Distribution. *J. Mass Spectrom.* **2013**, *48*, 989–997.

(58) Allen, S. J.; Eaton, R. M.; Bush, M. F. Analysis of Native-Like Ions Using Structures for Lossless Ion Manipulations. *Anal. Chem.* **2016**, *88*, 9118–9126.

(59) Liu, F. C.; Kirk, S. R.; Bleiholder, C. On the Structural Denaturation of Biological Analytes in Trapped Ion Mobility Spectrometry - Mass Spectrometry. *Analyst* **2016**, *141*, 3722–3730.

(60) 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* **2011**, *115*, 12266–12275.

(61) González Flórez, A. I.; Mucha, E.; Ahn, D.-S.; Gewinner, S.; Schöllkopf, W.; Pagel, K.; von Helden, G. Charge-Induced Unzipping of Isolated Proteins to a Defined Secondary Structure. *Angew. Chem. Int. Ed.* **2016**, *55*, 3295–3299.

(62) Going, C. C.; Xia, Z.; Williams, E. R. Real-Time HD Exchange Kinetics of Proteins from Buffered Aqueous Solution with Electrothermal Supercharging and Top-Down Tandem Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **2016**, *27*, 1019–1027.

(63) Laszlo, K. J.; Munger, E. B.; Bush, M. F. Folding of Protein Ions in the Gas Phase after Cation-to-Anion Proton-Transfer Reactions. *J. Am. Chem. Soc.* **2016**, *138*, 9581–9588.

(64) Laszlo, K. J.; Munger, E. B.; Bush, M. F. Effects of Solution Structure on the Folding of Lysozyme Ions in the Gas Phase. *J. Phys. Chem. B* **2017**, *121*, 2759–2766.

(65) Zenaidee, M. A.; Leeming, M. G.; Zhang, F.; Funston, T. T.; Donald, W. A. Highly Charged Protein Ions: The Strongest Organic Acids to Date. *Angew. Chem. Int. Ed.* **2017**, *56*, 8522–8526.

(66) Dixit, S. M.; Polasky, D. A.; Ruotolo, B. T. Collision Induced Unfolding of Isolated Proteins in the Gas Phase: Past, Present, and Future. *Curr. Opin. Chem. Biol.* **2018**, *42*, 93–100.

(67) Pagel, K.; Harvey, D. J. Ion Mobility–Mass Spectrometry of Complex Carbohydrates: Collision Cross Sections of Sodiated N-Linked Glycans. *Anal. Chem.* **2013**, *85*, 5138–5145.

(68) Salbo, R.; Bush, M. F.; Naver, H.; Campuzano, I.; Robinson, C.; Pettersson, I.; Jørgensen, T.; F Haselmann, K. Traveling-Wave Ion Mobility Mass Spectrometry of Protein Complexes: Accurate Calibrated Collision Cross-Sections of Human Insulin Oligomers. *Rapid Commun. Mass Spectrom.* **2012**, *26*, 1181–1193.

(69) Laszlo, K. J.; Buckner, J. H.; Munger, E. B.; Bush, M. F. Native-Like and Denatured Cytochrome *c* Ions Yield Cation-to-Anion Proton Transfer Reaction Products with Similar Collision Cross-Sections. *J. Am. Soc. Mass Spectrom.* **2017**, *28*, 1382–1391.

Appendix A. Native Anion-Exchange Chromatography: Rapid Sample Clean-Up and Fractionation for Native Mass Spectrometry

Daniele Canzani, Casey Chen, Claire E. Kozemchak, Pradipsinh K. Rathod, Matthew F. Bush

TABLE OF CONTENTS

Molecular Biology and Protein Purification.

Superfolder green fluorescent protein with a C-terminal hexa histidine tag (sfGFP-His6, Addgene plasmid #85492)[1] was transformed into BL21(DE3)pLysS Chemically Competent *Escherichia coli* (Thermo Fisher Scientific). *E. coli* cells were grown at 37 °C in Luria Broth (LB, Research Products International). Cells in the logarithmic growth phase (OD_{600} = 0.6) were collected by centrifugation and were resuspended in lysis buffer, which consisted of 50 mM Tris HCl pH 7, 150 mM NaCl, 0.1% TritonX-100, 10% glycerol, 5 mM imidazole, 2 mM β-mercaptoethanol, 4 cOmplete ULTRA Mini EDTA-Free Protease Inhibitor Cocktail tablets (MiliporeSigma), and 1 mM of phenylmethyl sulfonyl fluoride. The cells were sonicated on ice with a Branson Sonifier Cell Distributor 200 at pulse output control 3.5 and 35% duty cycle. Cell fragments were removed by centrifugation at 16,000 *g* for 30 minutes. The soluble protein fraction was immediately applied to a HisTrap FF Ni Sepharose 6 Fast Flow column (GE Healthcare). sfGFP-His6 was eluted from the column using elution buffer (50 mM Tris HCl pH 7.4, 150 mM NaCl, 250 mM imidazole, 2 mM β-mercaptoethanol, and 10% glycerol) then was dialyzed into storage buffer (20 mM Tris HCl pH 7.4, 150 mM NaCl, 2 mM β-mercaptoethanol, and 10% glycerol) and stored at -80°C. All steps were performed at 4 °C unless otherwise noted.

sfGFP-His6 Sequence

MVSKGEELFTGVVPILVELDGDVNGHKFSVRGEGEGDATNGKLTLKFICTTGKLPVPWPTLVTT
LTYGVQCFSRYPDHMKRHDFFKSAMPEGYVQERTISFKDDGTYKTRAEVKFEGDTLVNRIELKG
IDFKEDGNILGHKLEYNFNSHNVYITADKQKNGIKANFKIRHNVEDGSVQLADHYQQNTPIGDG
PVLLPDNHYLSTQSVLSKDPNEKRDHMVLLEFVTAAGITHGMDELYKGSHHHHHH

Reference

(1) Peeler, J. C.; Mehl, R. A. Site-Specific Incorporation of Unnatural Amino Acids as Probes for Protein Conformational Changes. In *Unnatural Amino Acids: Methods and Protocols*; Pollegioni, L., Servi, S., Eds.; Humana Press: Totowa, NJ, 2012; pp 125–134. https://doi.org/10.1007/978-1-61779-331-8_8.

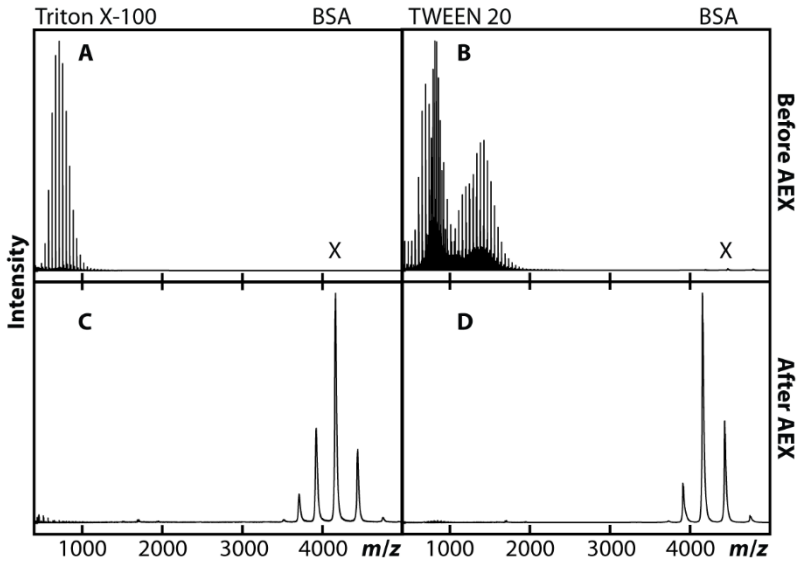

Figure A1. Native mass spectra of solutions that contained bovine serum albumin (BSA) and either (A) 0.05% Triton X-100 or (B) 0.05% Tween 20. BSA, which would be expected to appear at the position labeled "X" on each spectrum, was not readily identified, which is attributed to interference from the detergents. Those samples were loaded onto a native AEX column and eluted with an ionic strength gradient from 50 to 950 mM ammonium acetate (Figure 1B), resulting in the spectra shown in C and D in which BSA is easily identifiable and there are no significant interferences from detergents.

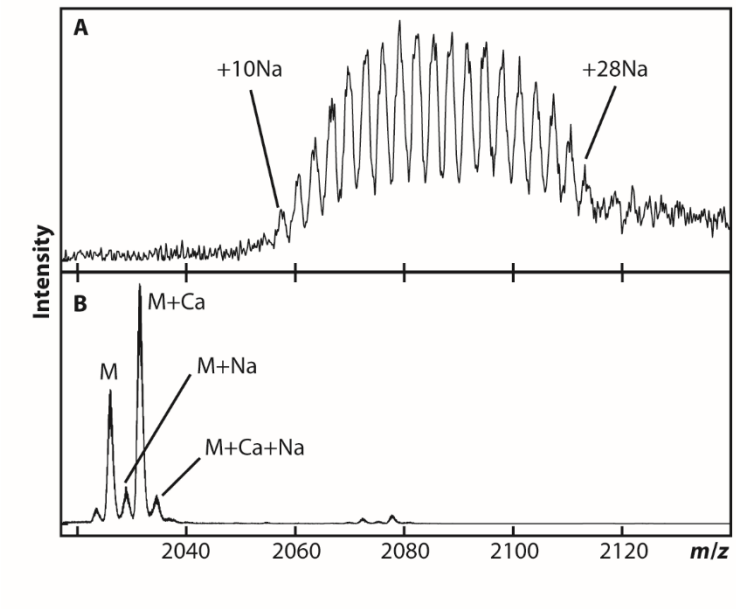

Figure A2. (A) α-lactalbumin was electrosprayed from aqueous 50 mM NaCl, which resulted in a high extent of sodium adduction. The 7+ charge state of α-lactalbumin is shown, which has between 10 and 28 sodium adducts. (B) Native mass spectrum a native AEX fraction collected during the separation of the sample used in panel A. Contributions from sodium ions are minimal after native AEX after native AEX. Calcium-bound α-lactalbumin could be identified, which was not the case before native AEX.

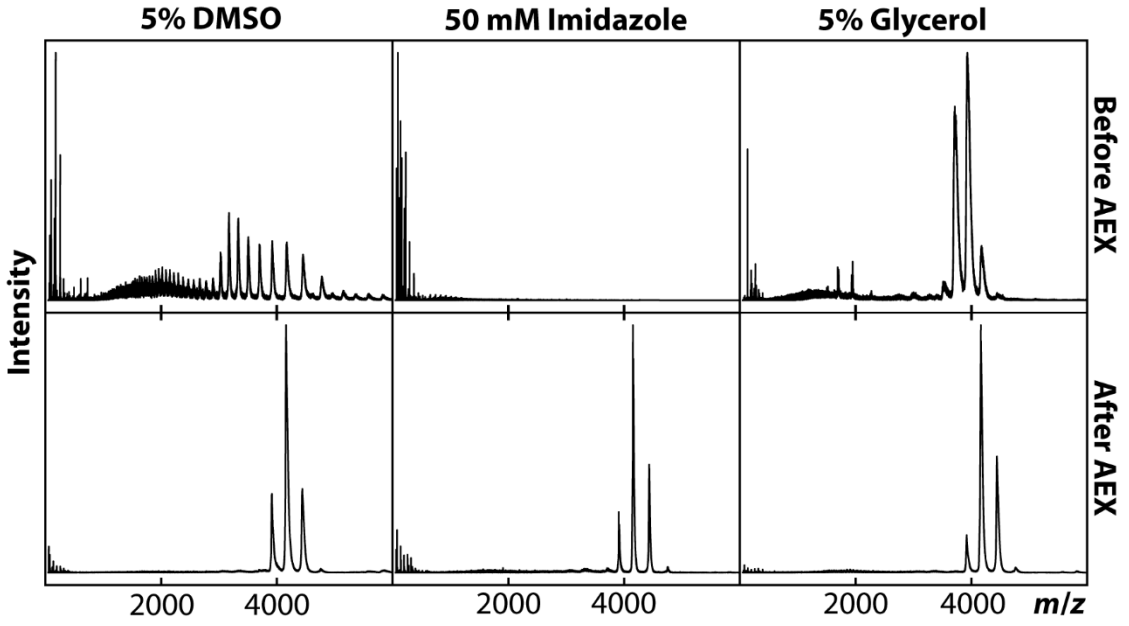

Figure A3. Non-volatile molecules (DMSO, imidazole, and glycerol) interfere with native MS of BSA (top row). These compounds were cleared from BSA samples using native AEX with an ionic-strength gradient from 50 to 950 mM ammonium acetate (bottom row).

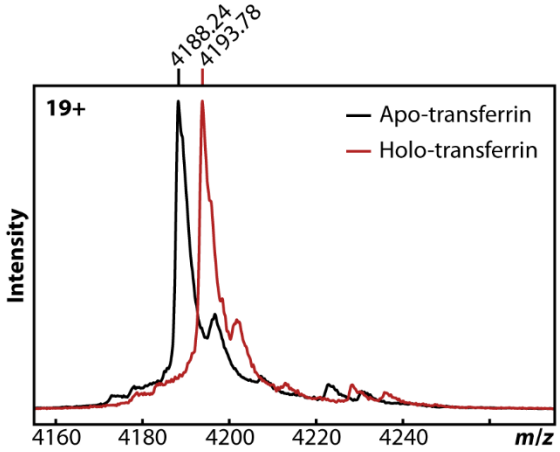

Figure A4. The black trace is a native mass spectrum of apo-transferrin ($[M+19H]^{19+}$) that was generated from a solution of aqueous 200 mM ammonium acetate. Holo-transferrin (iron-bound) was observed after native AEX (Figure 3B), but holo-transferrin was not detected using native MS alone (black trace). To confirm the assignment of holo-transferrin, the original sample was exchanged into an aqueous solution containing 150 mM ammonium acetate, 50 mM ammonium bicarbonate, and 0.5 mM $FeCl_3$, equilibrated for 20 minutes, and then analyzed with native MS. The resulting spectrum is shown in the rust trace, which contains predominantly iron-bound transferrin ($[M+2Fe+13H]^{19+}$) and corroborates the assignment of holo-transferrin in the native AEX experiment reported in Figure 3.

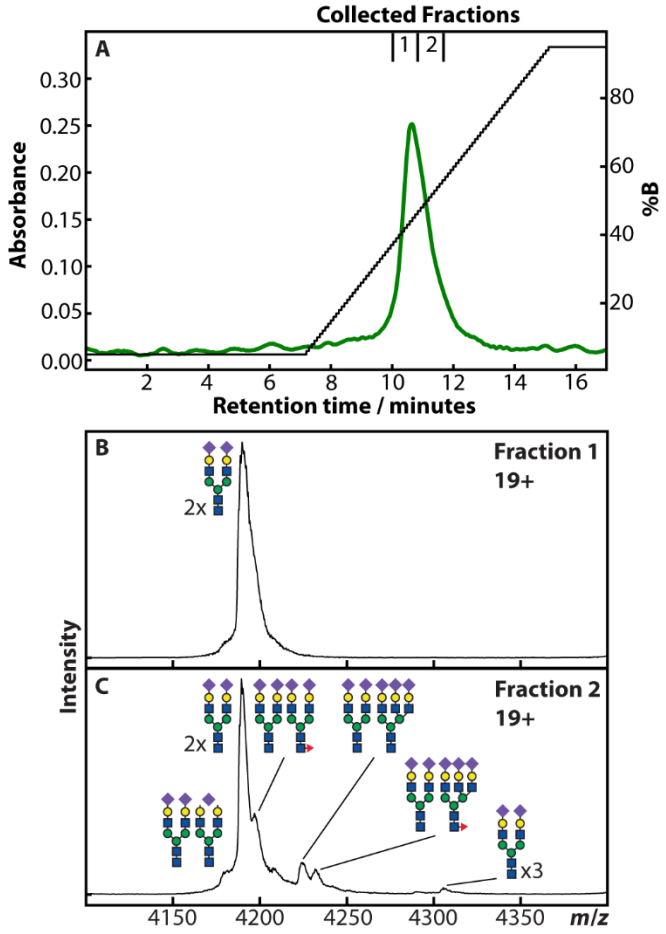

Figure A5. (A) Native AEX chromatogram of transferrin using a linear, ionic-strength gradient from 50 (solution A) to 950 (solution B) mM ammonium acetate. Two fractions were collected and analyzed using native mass spectrometry and are shown in panels B and C. Fraction 1 (panel B) contained one predominant glycoform, the 4-sialo variant (sialic acids are represented with purple diamonds). In fraction 2 (panel C), 3,4,5, and 6 sialo-transferrin were observed, as well as fucosylated (red triangle) variants of the 4 and 5 sialo-transferrin. This ionic-strength gradient did not separate the transferrin variants as well as the pH gradient (see chromatogram in Figure 2), but a longer gradient time may have improved the level of separation.

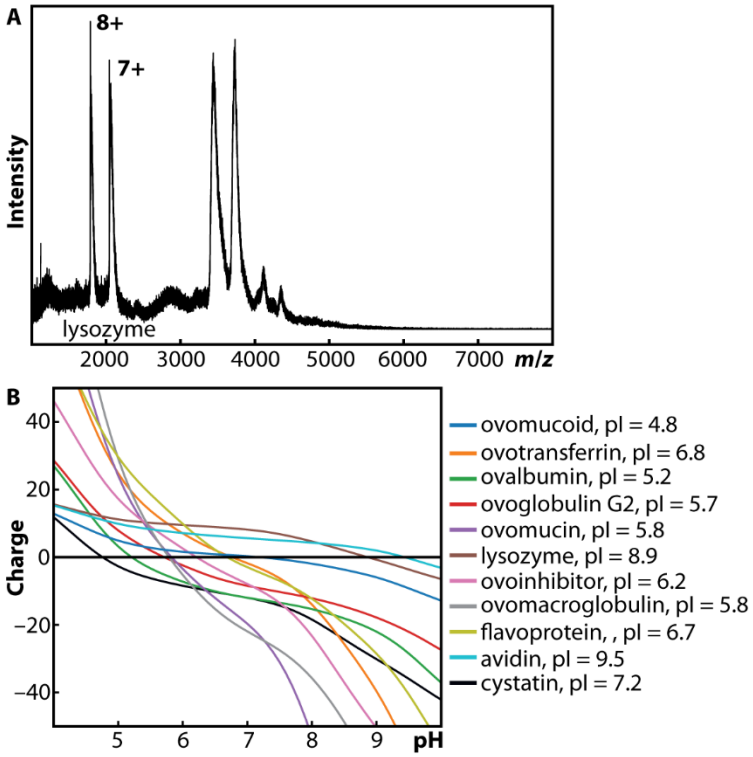

Figure A6. (A) Native mass spectrum of egg whites diluted 1:20 into aqueous 1 M ammonium acetate. Lysozyme was identified directly from this spectrum, but the remaining peaks could not be confidently assigned to a protein. (B) Theoretical titration curves and isoelectric points (pI) of egg white proteins, based on the sequences listed in UniProt.

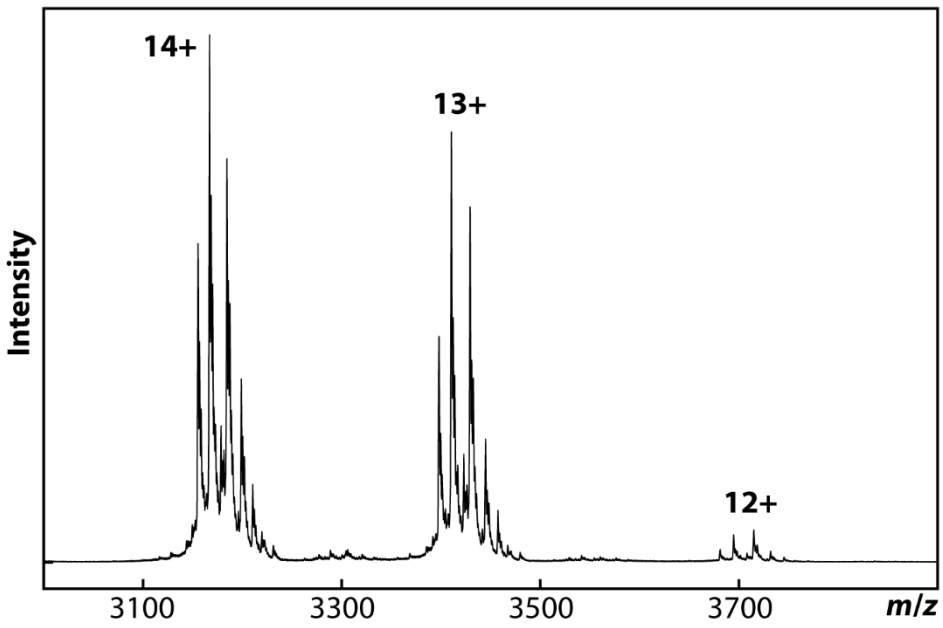

Figure A7. (A) Native mass spectrum of ovalbumin obtained from a fraction collected from native AEX of egg whites. A zero-charge version of this spectrum is plotted in Figure 3C and various glycoforms were identified.