Fourier Transform Based Analysis of Mass Spectra: Disentangling Mass Heterogeneity and Polydispersity

by

Andrew Kohl Swansiger

A dissertation accepted and approved in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

in Chemistry

Dissertation Committee:

Christopher Hendon, Chair

James Prell, Advisor

Julia Widom, Core Member

Benjamín Alemán, Institutional Representative

University of Oregon

Summer 2024

DISSERTATION ABSTRACT

Andrew Kohl Swansiger

Doctor of Philosophy in Chemistry

Title: Fourier Transform Based Analysis of Mass Spectra: Disentangling Mass Heterogeneity and Polydispersity

Understanding the interactions of small molecules with biomolecules and their complexes is fundamental to the clinical interpretation of biological functions and pharmaceutical development. Conversely, these delicate interactions present a multiplexed problem requiring highly specific and sensitive analytical techniques to capture their subtle variances. Advances in soft ionization mass spectrometry (MS) methods such as electrospray ionization (ESI) and desorption electrospray ionization (DESI) have brought together solution phase separation techniques and sensitive gas phase analysis, reducing both sample concentration and purification requirements and enabling fast multiplexed analysis of data-rich biological samples. As the limitations on analyte size and complexity continue to be pushed back by instrumental and experimental innovations, MS deconvolution tools need to continually advance to keep pace with the increased mass heterogeneity and polydispersity of what we can successfully spray.

Among current MS deconvolution algorithms, Fourier transform and Gábor transform (FT/GT) provide a consistent and invertible transform for quick recognition of several classes of periodic signal from polydisperse samples, requiring very few *a priori* assumptions about the sample while extracting the charge and mass information required by other algorithms for accurate modeling of congested mass spectra. The Prell group's iFAMS software represents the state-of-the-art in Fourier deconvolution of mass spectra, enabling flexible selection of analyte signals from a spectrogram of m/z and frequency to filter out interferent ions. However, assignment of aperiodic mass shifts in data-rich spectra still proves challenging, as they do not produce unique frequency signals, requiring an understanding of previously unutilized aspects of FT/GT deconvolution for mass spectrometry. Additionally, although iFAMS results are highly reproducible, applications of iFAMS data analysis have remained mostly exploratory, as GT lacks a sufficiently high-throughput implementation for analysis of large data sets.

In the first half of this dissertation, a new tool for mass spectrometry Fourier analysis is developed, utilizing the phase angle information from FT/GT for the characterization of small mass variants embedded in polydisperse mediums such as polymers and lipid membranes. The new method of FT/GT macromolecular mass defect (MMD) analysis achieved similar mass accuracy to mass-domain deconvolution methods and is robust to high instrument noise and low mass contaminants, enabling cross-validation of mass-domain deconvolution models. In a workflow complemented with liquid chromatography mass spectrometry, FT/GT MMD analysis enables characterization of polymer reaction intermediates. The second half of the dissertation extends the reproducibility of FT/GT analysis to protein quantitation of MS imaging data from biological tissue, developing a new workflow for batch deconvolution to process tens of thousands of spectra in a few hours. The distinct protein ion patterns generated by GT simplify characterization of brain tissue eluents, while expanding the range of isolatable proteoform signal available for imaging.

This dissertation includes previously published and unpublished co-authored material.

CURRICULUM VITAE

NAME OF AUTHOR: Andrew Kohl Swansiger

GRADUATE AND UNDERGRADUATE SCHOOLS ATTENDED:

University of Oregon, Eugene

University of Washington

DEGREES AWARDED:

Doctor of Philosophy, Chemistry, 2024, University of Oregon

Bachelor of Science, Chemistry, 2018, University of Washington

AREAS OF SPECIAL INTEREST:

Liquid Chromatography-Mass Spectrometry

Analytical Method Validation

Fourier Deconvolution

PROFESSIONAL EXPERIENCE:

Research and Early Development Intern, Synthetic Molecule Analytical Chemistry, Genentech, Inc., 2022

Graduate Teaching Fellow, University of Oregon Department of Chemistry and Biochemistry, 2018-2019 and 2020-2022

PUBLICATIONS:

Swansiger, A.K.; Crittenden, C.M.; Chan, S.A.; Yang, S.H.; Kou, D.; Prell, J.S.; Chen, B. Streamlining LC-MS Characterization of Pharmaceutical Polymers by Fourier-transform Based Deconvolution and Macromolecular Mass Defect Analysis. *Anal. Chem.* **2024**, *manuscript pending accepted with minor revisions*

Meldrum, K.L.; Swansiger, A.K.; Daniels, M.M.; Hale, W.A.; Kirmiz Cody, C.; Qiu, X.; Knierman, M.; Sausen, J.; Prell, J.S. Gábor Transform-Based Signal Isolation, Rapid

Deconvolution, and Quantitation of Intact Protein Ions with Mass Spectrometry. *Anal. Chem.*, **2024**, *96*, 9512-9523

Tang, S.; Pederson, Z.; Meany, E.; Yen, C.-W.; Swansiger, A.K.; Prell, J.S.; Chen. B.; Grosskopf, A.; Eckman, N.; Jiang, G.; Baillet, J.; Pellett, J.; Appel, E. Label-Free Composition Analysis of Supramolecular Polymer-Nanoparticle Hydrogels by Reversed-Phase Liquid Chromatography Coupled with a Charge Aerosol Detector. *Anal. Chem.* **2024**, *96*, 5860-5868

Wu, D.; Yan, R.; Song, S.; Swansiger, A.K.; Li, Y.; Prell, J.S.; Zhou, Q.; Robinson, C.V. The Complete Assembly of Human LAT1-4F2hc Complex Provides Insights into its Regulation, Function and Localisation. *Nat. Commun.* **2024**, *57*, 3711

Walker, D.R.; Jara, K.A.; Rolland, A.D.; Brooks, C.; Hare, W.; Swansiger, A.K.; Reardon, P.N.; Prell, J.S.; Barbar, E.J. Linker Length Drives Heterogeneity of LC8/ASCIZ Multivalent Complexes. *Biomolecules* **2023**, *13*, 404

Swansiger, A.K.; Marty, M.T.; Prell, J.S. Fourier-Transform Approach for Reconstructing Macromolecular Mass Defect Profiles. *J. Am. Soc. Mass Spectrom.* **2022**, *33*, 172-180

ACKNOWLEDGMENTS

I would like to thank the National Science Foundation for funding provided through their Chemical Measurement and Imaging Program, Industry-University Cooperative Research Centers Program, and Grant Opportunities for Academic Liaison with Industry Program, as well as the National Institute for General Medical Sciences, the National Institutes of Health Human BioMolecular Atlas Program, and Genentech, Inc. for funding this research.

TABLE OF CONTENTS

LIST OF FIGURES

Figure Page

LIST OF TABLES

CHAPTER I

INTRODUCTION

Soft-ionization mass spectrometry (MS) has seen rapid development over the past few decades as an analytical technique for characterizing non-covalent complexes and heterogeneous biological samples. Transfer of intact macromolecular ions to the gas phase for mass analysis is often achieved by electrospray ionization (ESI), originally adapted for mass spectrometry to study polymer chemistry by Dole and coworkers,[1] and later developed by Fenn and coworkers for the study of biopolymers (i.e., proteins and oligonucleotides).[2] Acknowledged by the 2002 Nobel prize in chemistry awarded to Fenn, Tanaka, and Wüthrich as a major advancement in the structural analysis of large biomolecules, ESI-MS can ionize analytes by deposition of charge carriers on their surface (e.g., protons, metal ions and other small charged species like ammonium) during their evaporative transition from aqueous droplets to the gas phase if they are not already ions in solution (Figure 1). For large analytes, ESI inherently produces multiply charged ions occupying several consecutive charge states, bringing the mass-to-charge ratios (m/z) of large macromolecular ions (100 kDa – 1 MDa) into detectable ranges for a broad array of mass analyzers.[3] Accurate masses can be determined for these ions in many ways, for example, by accelerating them to well-controlled kinetic energy and measuring their travel velocity in time-of-flight (TOF) instruments, or by their orbiting frequency within an electric field (Orbitrap-type instruments) or the high magnetic field of Fourier Transform Ion Cyclotron Resonance (FT-ICR) mass spectrometers. The high sensitivity of mass spectrometers, with lower limits of detection on the order of attomoles down to single ions, enables high-quality analyses using only a few microliters of sample at concentrations in the nano- to micromolar range. Due to this exquisite sensitivity, chemical species can be identified with mass accuracies below 0.1 ppm root-mean-squared (RMS) error for FT-ICRs and 1-5 ppm RMS error for commercially available TOF instruments. However, the charge state multiplicity inherent to ESI spreads analyte signal over several peaks in the mass spectrum, compounding the deleterious effects of increased mass heterogeneity from compositional mutations, post-translational modifications (PTMs), oligomerization, and adduction of salts and other small co-solutes. The resulting mass spectra can often be very broad and congested, with overlapped signals from different ion species. While chromatographic separation and other experimental tools, such as submicron

emitters, charge-reducing agents, and non-volatile salts, can help to address these sorts of problems, complementary experimental and data analysis approaches are sometimes needed that can handle heterogeneity and low-purity samples more directly.

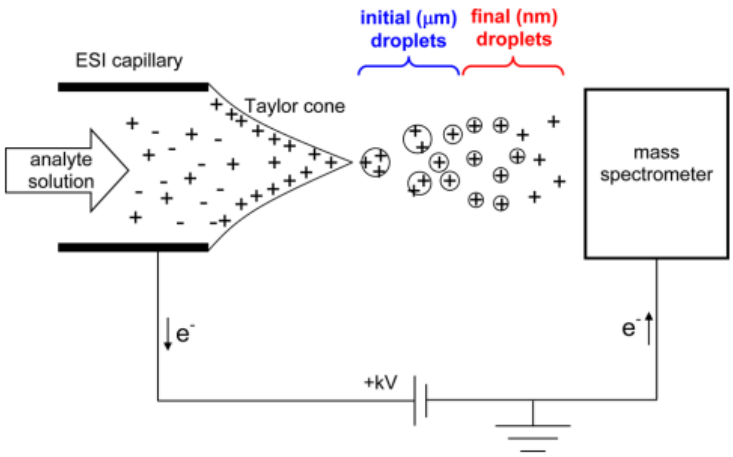


Figure 1. Schematic of a typical electrospray ionization source setup. Capillary tips are pulled to an internal diameter of a few micrometers and analyte solutions are sprayed with a capillary voltage potential of 0.5-2 kV, relative to ground. Reprinted with permission from Konermann, L.; Ahadi, E.; Rodriguez, A. D.; Vahidi, S. Unraveling the Mechanism of Electrospray Ionization. *Anal. Chem.* **2013**, *85* (1), 2-9. Copyright 2013, American Chemical Society.[4]

Advancements in the resolving power and sensitivity of ESI-MS instrumentation have drastically increased the limits on analyte size and structural complexity analyzable by this technique, allowing for more experimental freedom to tackle "dirty" multi-component samples. Much protein functionality stems from the multitude of ways they interact with their chemical environment, ligand compounds, and each other, and a large fraction of the human proteome is expected to exhibit two or more proteoforms with different masses.[5] Enzymatic digestion of biomolecular samples into smaller fragments followed by MS analysis (as in "bottom-up" proteomics and other "omics" techniques) can lead to cleavage of PTMs during gas-phase activation, loss of information about the proteoform distribution, and loss of native non-covalent interactions and ligand binding, so developing robust experimental methods which capitalize on the gentler nature of ESI help to preserve this information is highly desirable. Numerous groups have developed "top-down omics" methods,[6-8] in which intact biomolecules are analyzed by MS without prior digestion, and "native" MS methods, in which folded biomolecular ions and their complexes are transferred directly from buffered aqueous solution into the gas phase while

preserving much of their high-order structure.[9-15] For example, revolutionary methods developed by Robinson and coworkers to eject proteins from lipid membranes in the mass spectrometer as well as tailored detergent environments have been pivotal in understanding lipid-stabilized protein complexes.[16-22]

Ionizing analytes directly from solution phase also enables the coupling of common liquid chromatography techniques to MS and ion mobility-MS for improved analyte specificity, and ions can be unfolded or dissociated inside many mass spectrometers to provide additional structural and compositional information.[23-28] Recently, ESI has also been achieved directly from the surface of solid and liquid samples in Desorption ESI (DESI) and nanoelectrospray desorption ionization (nano-DESI) MS, enabling, for example, chemically specific spatial mapping of analyte concentrations across biological tissues with mass spectrometry. For over a century, the standard analytical imaging method in pathology has been light microscopy, with aid from histological stains to highlight various cellular structures within biological tissue.[29] However, these images provide very limited chemical information about the biomolecules present in the tissues (see Figure 2, Histology Image). Immunofluorescence staining enables high-resolution spatial imaging of multiple chemical distributions in tissue, but this technique still requires specific chemical tags or protein-targeting antibodies for each analyte (see Figure 3, IF Image), development of which is very time consuming and expensive. Thus, more universal chemically specific tissue imaging of a wide variety of biomolecules and their complexes, which is now possible with ESI- and nano-DESI-based MS imaging, is of great interest. The Laskin laboratory at Purdue University has achieved single-cell resolution with nano-DESI MS imaging (nano-DESI-MSI, see Figure 3), eluting lipids, proteins and other metabolites from biological tissue with an ESI-compatible aqueous solution through a narrow solvent bridge between two capillaries (10 μm in inner diameter). One capillary supplies the ESI solvent from a syringe pump, forming a small bead of solvent on the slide-mounted tissue surface, while the other capillary transfers the eluent from the surface of the tissue to the MS instrument inlet for ionization. The tissue sample is rastered underneath the nano-DESI setup on a motorized sample stage, maintaining consistent height of the capillary openings above the tissue surface with shear force microscopy.[30]

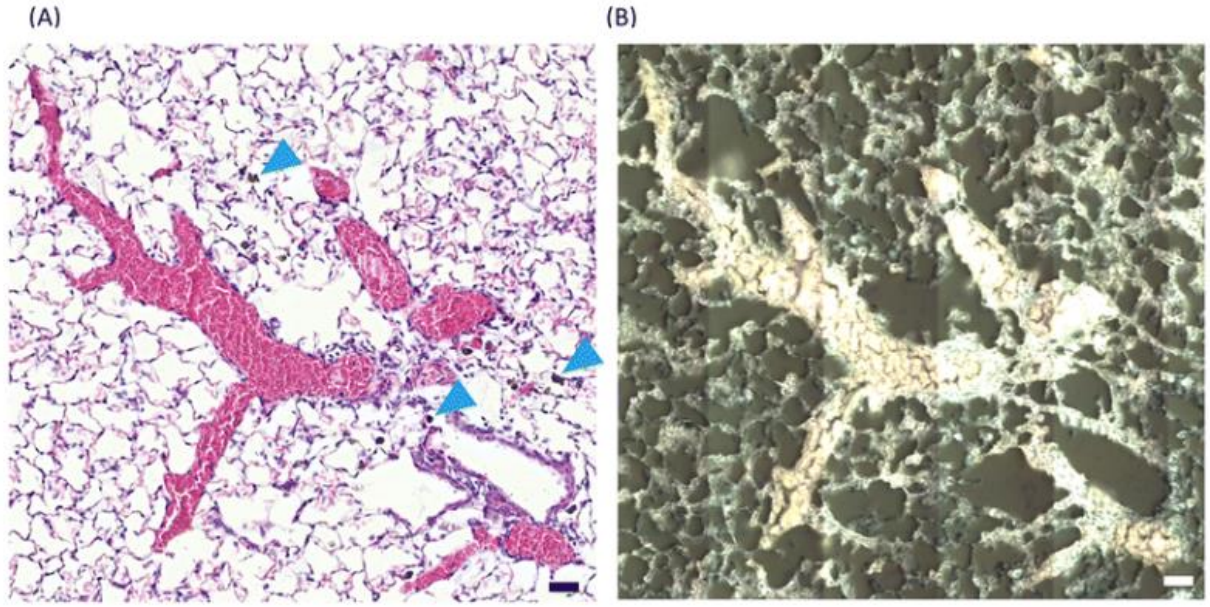


Figure 2. Hematoxylin and Eosin histology stain (a) and decolorized image (b) tracking accumulation of titanium oxide (blue arrows) in mouse lung tissue. Scale bars (bottom right) = 50 µm. Reprinted with permission from Akaji, S.; Sagawa, T.; Honda, A.; Miyasaka, N.; Sadakane, K.; Ichinose, T.; Takano, H. Post-staining Raman analysis of histological sections following decolorization. *Analyst* **2022**, *147* (20), 4473-4479. Copyright 2022, Royal Society of Chemistry;[31] permission conveyed through Copyright Clearance Center, Inc.

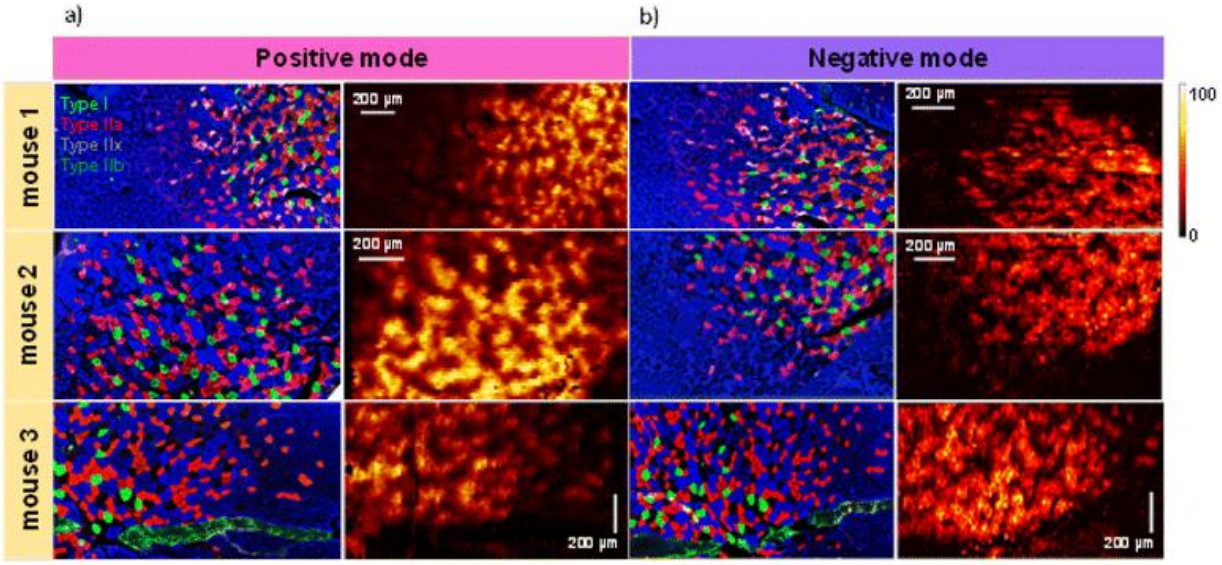


Figure 3. Immunofluorescence and nanoDESI images of mouse skeletal myotissue under positive (a) and negative ionization (b). Nano-DESI images (right) show tissue localizations independent of the cell types identified by immunofluorescence (left). Reprinted with permission from Unsihuay, D.; Hu, H.; Qiu, J.; Latorre-Palomino, A.; Yang, M.; Yue, F.; Yin, R.; Kuang, S.; Laskin, J. Multimodal high-resolution nano-DESI MSI and immunofluorescence imaging reveal molecular signatures of skeletal muscle fiber types. *Chem. Sci.* **2023**, *14* (15), 4070-4082. Copyright 2023, Royal Society of Chemistry;[32] permissions conveyed through Copyright Clearance Center, Inc.

Differentiating small mass variants in proteins is an active area of clinical study as biomarkers for disease states, with examples ranging from ribosomal protein mutations for identifying bacterial strains to characterization of cancers by the altered glycosylation and phosphorylation modifications they produce.[33-35] However, many of these studies still rely on analyte isolation, fragmentation, and comparison to reference databases. To complement the many experimental and instrumental advancements in soft-ionization mass spectrometry which have enabled the study of heterogeneous macromolecular complexes while maintaining their delicate intermolecular interactions, robust computational tools and deconvolution algorithms continue to be developed to disentangle the often extremely data-rich ESI mass spectra. Currently the most widely used techniques for handling large biomolecule mass spectra include maximum entropy algorithms (often called "MaxEnt") and Bayesian deconvolution modeling (implemented in "UniDec" from the Marty group at University of Arizona and PMI Intact from Protein Metrics, LLC).[36-38] Modified for use on ESI spectra and other signal processing applications by Skilling and co-workers in 1984,[39] MaxEnt algorithms have been developed by several instrument and software vendors, all fundamentally starting with a series of assumptions about the range of masses and charge states present in the original spectra. A "test" distribution of charge-normalized mass peaks (a "zero-charge" spectrum) based on the observed data is generated, adjusted for charge carrier mass, and divided by a range of charge states to synthesize a hypothetical mass (m/z) spectrum.[40] The difference between this hypothetical mass spectrum distribution and the observed data at each sampled point in m/z is computed, and the total Shannon entropy for this difference is calculated assuming a particular type of noise distribution, such as Gaussian white noise or Poisson noise. The original test spectrum is modified to maximize the total Shannon entropy, and this process is iterated until convergence is achieved to yield the final deconvolved mass spectrum. Bayesian mass spectral deconvolution works similarly. UniDec uses a Bayesian deconvolution method which models the mass spectrum as a rectangular matrix of m/z and charge state. For example, starting from a uniform matrix of m/z and charge state to represent the information in the mass spectrum, UniDec iteratively smooths the hypothetical charge state distribution to avoid non-consecutive charge assignments, sums and convolves the matrix along the charge state axis, then checks the resultant simulated spectrum against the experimental spectrum until mismatch is minimized.[36, 41, 42] A major improvement of UniDec over MaxEnt is that UniDec allows for different mass distributions for each charge state

belonging to an analyte, a situation that may occur for many realistic ion types, such as long-chain polymers and proteins with PTMs and/or bound ligands.

While both MaxEnt and Bayesian deconvolution excel at characterizing baseline-resolved mass spectra, both perform best when a significant amount of background knowledge of the sample being evaluated is known ahead of time. The more congested a mass spectrum is, the higher likelihood of ambiguous mass assignment, which can be sensitive to data pretreatment such as baseline subtraction and peak smoothing. UniDec is generally more robust to broader initial mass and charge state ranges, however both methods are required to match the observed spectrum as closely as possible, which can result in artifacts if given a mass or charge state range much larger or smaller than is actually present in the sample.[43, 44] Extracting mass and charge range information via other methods have been shown to improve the reliability of UniDec reconstructions of highly congested mass spectra,[45] but in general such spectrum modeling methods can be highly sensitive to initial information, and systematic optimization of initial deconvolution parameters is challenging with little prior knowledge of the sample.

Other mass spectrum deconvolution methods have been developed to circumvent these limitations, including use of linear transforms such as Fourier and Gábor transform (FT and GT). These methods can provide information about the mass spectrum without requiring a hypothetical model of the data or pre-smoothing of the spectrum, and they are invertible. Fourier analysis, originally developed by Joseph Fourier for the study of heat flow and greatly expanded over the past two centuries for an enormous range of applications including NMR spectroscopy, LiDAR, and speech recognition,[46-50] has already seen use in mass spectrometry to relate the time-domain cyclotron oscillations from resonance mass analyzers like FT-ICR to the m/z ratios of the cycling ions. Prebyl and Cook introduced the first direct application of Fourier transform for deconvolution of polymer ESI spectra in 2004,[51] noting how the chain structure of polymers generates a periodic distribution of peaks in the m/z-domain which, when Fourier transformed, often gives rise to intense, sharp peaks at a few frequencies (k) corresponding to the reciprocal of the peak spacing. The Prell group further developed the theoretical application of Fourier deconvolution for mass spectrometry data, producing the framework of the iFAMS deconvolution software in 2016.[52] In principle, the functional form of polymeric mass distributions (which can be denoted $s(m/z)$), or any other ion with resolved mass polydispersity

such as isotope distributions, can be thought of as an approximately Gaussian distribution, e(m/z), of periodic peaks (represented by a "comb function", c(m/z)) with some shared peak shape, p(m/z), spaced by the polydisperse unit mass over the charge state (ΔM/Z). Therefore, FT peaks occur at Z/ΔM with spacings between consecutive charge states of 1/ΔM, allowing the determination of polydisperse unit mass and occupied charge states of the polymeric ion and from frequency alone. Charge-state-specific information can also be isolated from the FT and inverse Fourier-transformed for reconstruction of the ion's zero-charge spectrum. However, the functional form of the Fourier transformed spectrum $S(k) = [C(k) * E(k)] \times P(k)$ still contains a periodic component C(k), making it clear that the functional shape of mass spectra does not result in a singular frequency signal (Figure 4). P(k) is inversely proportional to the width of p(m/z), thus higher peak resolution in m/z results in higher order frequency harmonics which can overlap between different charge state harmonics for the same polydisperse ion. In a growing number of common applications, this dual relationship between the mass spectrum and its Fourier transform can turn a highly congested mass spectrum that is exceptionally challenging to assign by conventional means into a relatively simple frequency spectrum with a much easier interpretation.

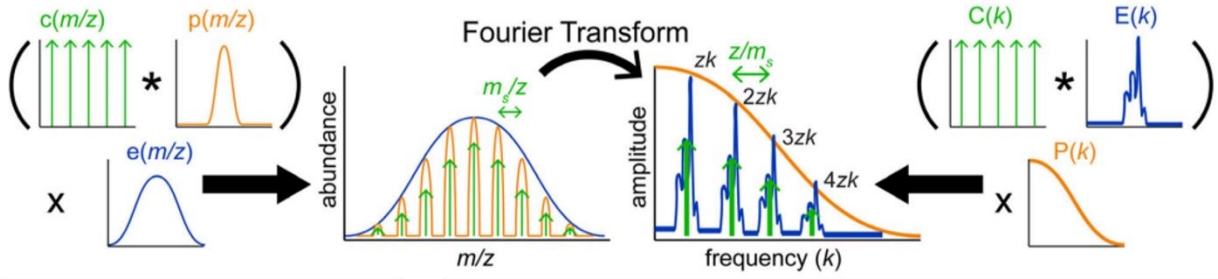


Figure 4. Cartoon of Fourier transform of a peak distribution from a mass spectrum. Within the equations for the mass spectrum (left) and its Fourier transform (right), "*" denotes convolution and "x" denotes multiplication. Adapted with permission from Cleary, S. P.; Li, H.; Bagal, D.; Loo, J. A.; Campuzano, I. D. G.; Prell, J. S. Extracting Charge and Mass Information from Highly Congested Mass Spectra Using Fourier-Domain Harmonics. *J. Am. Soc. Mass Spectrom.* **2018**, *29* (10), 2067-2080. Copyright 2018, American Chemical Society.[45]

Application of a windowed Fourier transform (also known as "Short-Time Fourier Transform", or STFT, from its common applications in other fields to time-domain data), plotted as a 2-dimensional spectrogram of m/z and frequency, localizes frequency information in m/z.

This introduces a major advantage over Fourier transform, in that charge states for higher harmonic data arising from high-resolution mass spectra can often be much more easily and uniquely assigned.[53] STFT bears many similarities to and shares many advantages with wavelet transform, which is used in many fields for data processing and compression. However, the choice of wavelet shape for wavelet transform is critical to intuitive interpretation of the resultant spectrogram and the ideal wavelet can in principle be very different depending on the ion of interest. An *m*/*z*-domain deconvolution algorithm somewhat analogous to wavelet transform is implemented in the sliding window adduct removal method (SWARM) developed by Klassen and coworkers, choosing an ion distribution shape or "waveform" observed in the input spectrum and manipulating it to probe for similar distributions, stretching the waveform for lower charge states and compressing it for higher charge states.[54] Peaks matching the probe's shape are recorded and removed from the mass spectrum via an onion-peeling algorithm and the process is repeated for remaining ion signals. While useful for well-resolved ion distributions and pre-determined target proteins, the use of more specialized waveforms makes wavelet transform less generalizable, requiring the user to build an intuition for each chemical species of interest. By contrast, using a Gaussian waveform for STFT, known as Gábor transform (GT), results in uniform resolution across all frequencies while requiring no *a priori* assumptions about the shape of the mass distribution. Additionally, much of the mathematical intuition developed across the sciences over the past two centuries for interpreting FT carries over to GT, thus many rules of thumb for interpreting GT spectrograms can be quickly learned with relative ease. For example, GT generates predictable and easily recognized charge state patterns in the spectrogram for mass spectral data of several classes of ion: proteins (and other ions for which the mass changes negligibly with charge state) occupy lower frequencies at higher mass-to-charge ratios (negatively chirped signal), while salt clusters (and other ions with a broad mass distribution for a single charge state) do not change in frequency with increased mass to charge (no chirp). FT/GT frequency patterns for multi-subunit polydisperse samples have also been explored in detail, culminating in 3 classes of heterogeneity that can be easily recognized from their associated FT spectra and GT spectrograms (Figure 5): independent mass distributions (i.e., simple mixtures), co-dependent or mean-proportional-variance polydispersity (e.g., block copolymers), and multinomial polydispersity (i.e., displacement of one subunit with another within a constrained oligomeric structure).

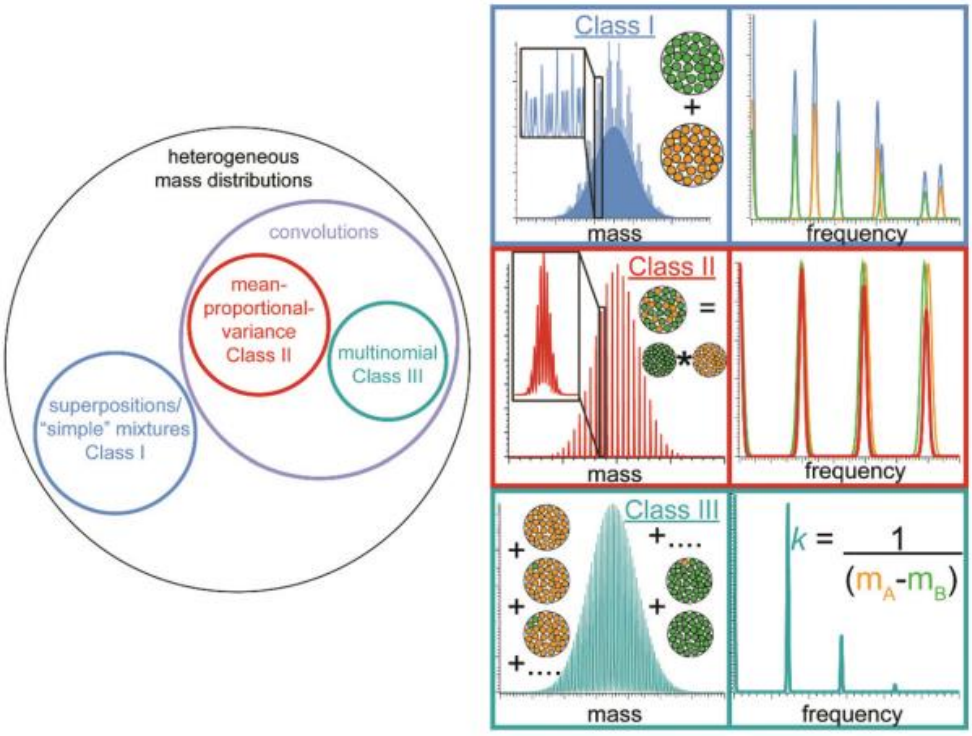


Figure 5. Cartoon describing the 3 classes of heterogeneity. Orange and green dots represent the two distinct polydisperse unit populations of different masses. Reprinted with permission from Cleary, S. P.; Prell, J. S. Distinct classes of multi-subunit heterogeneity: analysis using Fourier Transform methods and native mass spectrometry. *Analyst* **2020**, *145* (13), 4688-4697. Copyright 2020, Royal Society of Chemistry;[55] permission conveyed through Copyright Clearance Center, Inc.

FT/GT frequency analysis enables the extraction of an ion population's repeat unit mass, charge state distribution and mass distribution from overlapped charge state series, however, aperiodic mass shifts arising from, e.g., polymer end group functionalization, protein stoichiometries in lipid membranes, do not result in unique peaks in the frequency domain. Whereas a mass spectrum is by nature both real-valued and positive, its Fourier transform $S(k) = \int_{-\infty}^{\infty} s\left(\frac{m}{z}\right) e^{-2\pi i k} dk$ is inherently complex-valued, though it retains all compositional information represented in the mass spectrum. To fully take advantage of the compositional information preserved by ESI-MS, it is desirable to extend interpretation and application of FT/GT frequency-domain data beyond peak frequencies and magnitudes, i.e., to phase information, as well. Thus, the first goal of this dissertation research was to explore unutilized aspects in Fourier deconvolution of mass spectrometry data for the characterization of mass heterogeneities within polydisperse samples. The second goal of this dissertation was to develop

software tools to extend the applicability of the Prell lab's open-source FT deconvolution tool to novel problems in sample quantitation.

Chapters II and III focus on my work differentiating small mass heterogeneities from polydisperse mass environments, leveraging FT phase-angle information for mass defect analysis of multiply charged ions, yielding profiles analogous to Kendrick mass defect profiles commonly used in crude oil and polymer analysis. In Chapter II, I lay the groundwork of this noise-tolerant method of macromolecular mass defect (MMD) analysis and its capacity for cross-validation of UniDec's *m*/*z*-domain results. This chapter includes co-authored work with Michael T. Marty and James S. Prell. In Chapter III, I discuss the application of iFAMS MMD analysis to the characterization of polymer end groups, establishing a LC-MS analysis workflow for quality control and reaction monitoring of branched polymer synthesis. This chapter includes co-authored work from Christopher M. Crittenden, Simon A. Chan, Samuel H. Yang, Dawen Kou, James S. Prell, and Bifan Chen. In Chapter IV, I discuss the application and benefits of Gábor analysis for multi-protein spectra eluted from biological tissue samples and development of a parallelized batch Fourier deconvolution workflow and proteoform intensity mapping tool for mass spectral imaging. This chapter includes co-authored work from Lily M. Miller, Manxi Yang, Hang Hu, Julia Laskin and James S. Prell.

CHAPTER II

FOURIER-TRANSFORM APPROACH FOR RECONSTRUCTING MACROMOLECULAR MASS DEFECT PROFILES

First author paper co-authored with Michael T. Marty, who provided both nano-ESI data of melittin-embedded Nanodiscs and manuscript feedback during the revision process, and James S. Prell, who provided substantial support in developing the theoretical background for the FT phase analysis. Reprinted with permission from Swansiger, A. K.; Marty, M. T.; Prell, J. S. Fourier-Transform Approach for Reconstructing Macromolecular Mass Defect Profiles. *J. Am. Soc. Mass Spectrom.* **2021**, *33* (1), 172-180.

Introduction

Structural and compositional heterogeneity is important to the function of many types of natural and therapeutic biomolecules and biomolecular complexes, including protein complexes, protein-nucleic acid complexes, membrane protein-lipid assemblies, and antibodies.[56-62] With the growing importance of biotherapeutic proteins such as monoclonal antibodies, the need to accurately characterize intact protein complexes and their constituent isoforms for biomolecule characterization and biopharmaceutical design is rapidly growing. Native electrospray ionization mass spectrometry (ESI-MS) and ion mobility-mass spectrometry (IM-MS) have established themselves as highly sensitive tools for analysis of intact non-covalent complexes,[13, 63, 64] from antibody-drug conjugates and glycosylation states,[56, 65-67] to studies of membrane protein assembly structure and lipid binding,[68-74] and have been used to both update results from and inform sample preparation for more traditional crystallographic and microscopic structural analysis methods.[14, 67, 75] The development of several MS deconvolution tools has even allowed for compositional studies of analytes with exceptionally high mass polydispersity (due to, e.g., variation in the number of a repeated subunits, such as glycosylation states of glycoproteins, lipids in a membrane or micelle, or polymer chain length).[13] These types of analytes often produce native-MS spectra with several overlapped charge states and many tens or even hundreds of overlapped peaks.[14, 43, 52, 76] Fourier Transform (FT)-based signal processing methods take advantage of sample polydispersity as a source of periodic signals that are used to

determine ion charge states, subunit masses, composition, and total masses from the frequency domain.[14, 45, 51-53, 55] Conversely, Bayesian statistical analysis methods as used in the software packages UniDec from the Marty group and PMI Intact from Protein Metrics, Inc., deconvolve mass spectra directly from the *m*/*z* domain by iteratively optimizing a model spectrum based on an initial set of user-input parameters.[36, 37, 77-79]

While deconvolution methods like those mentioned above can often deconvolve a mass spectrum to a "zero-charge" mass spectrum with many fewer peaks, it can still be very challenging to interpret the zero-charge mass spectrum for an ion population that is highly polydisperse in the stoichiometry of a particular adduct, especially when there is significant heterogeneity in the underlying "base" composition, i.e., the portion of each ion other than the polydisperse adducts. Recently, Marty and co-workers have reported on the preferred lipid bilayer-incorporation stoichiometries of several antimicrobial peptides by macromolecular mass defect (MMD) analysis of UniDec reconstructed mass spectra,[44, 80-82] using modular arithmetic to filter out polydisperse signal in a manner similar to how Kendrick mass defect analysis is used to characterize end groups of polymers.[83-85] Their studies of melittin incorporation into Nanodiscs, self-assembled membrane mimics consisting of a phospholipid bilayer encircled by two amphipathic membrane proteins (MSPs),[86-88] revealed no strong preferences for specific stoichiometries—contrasting with results from earlier studies which suggested a preference for melittin tetramers in lipid bilayers. These results provide an excellent basis from which to explore orthogonal MMD analysis methods both to validate the Bayesian results and to provide alternative means of generating MMD profiles, possibly with unique advantages.

Here, we extend upon the utility of the Prell group's FT-based deconvolution software, Interactive FT Analysis for Mass Spectrometry (iFAMS), to characterize the base composition of membrane-protein-embedded Nanodiscs by measuring the MMD profile using phase information from the Fourier Transform. Because MMD profiles are generated solely from Fourier-domain information extracted from the raw data, this phase analysis method is totally orthogonal to the mass-domain analysis used in UniDec. We show that FT phase analysis (both with and without Richardson-Lucy peak sharpening) confirms melittin-incorporation stoichiometries previously reported using the UniDec deconvolution software, with similar or better agreement to theoretical MMDs based on the known masses of melittin and Nanodisc components. We then

illustrate the utility of this new method to reconstruct MMD profiles from lower-resolution ESI-Q-IM-TOF data by studying the formation of Nanodiscs from a mixture of two different scaffold proteins of similar size.

Methods

Sample Preparation. All phospholipid Nanodiscs were prepared using a method adapted from that of Sligar and coworkers.[87-90] Nanodiscs containing only scaffold proteins and lipids were prepared at the University of Oregon, using water purified to 18 MΩ•cm resistivity. Briefly, dimyristoylphosphatidylcholine (DMPC) lipids purchased from Avanti Polar Lipids as 5 mg/mL suspensions in chloroform were dried with nitrogen gas until opaque, then re-suspended to a concentration of 50 mM in a pH 7.4 aqueous buffer composed of 100 mM sodium cholate (Sigma-Aldrich, St. Louis, MO), 20 mM Tris (Bio-Rad, Hercules, CA), 100 mM sodium chloride, and 0.5 mM ethylenediaminetetraacetic acid. Histidine-tagged (MSP1D1His) and non-histidine-tagged (MSP1D1(−)) membrane scaffold proteins were reconstituted separately in identical pH 7.4 aqueous buffer without sodium cholate. For single-MSP experiments, these solutions were used directly. For mixed-MSP Nanodisc samples, equimolar mixtures of the two MSP1D1 variants were prepared at room temperature (~25 °C) in microcentrifuge tubes (Sorenson, Salt Lake City, UT), with a final protein concentration of ~200 μM. Each of the above MSP1D1 solutions were then mixed with DMPC-detergent suspensions and additional buffer for an expected Nanodisc concentration of 50 μM and incubated at room temperature for 15 minutes. Nanodisc self-assembly was started by 1000:1 dialysis by volume into buffer in which Bio-Beads SM-2 (Bio-Rad, Hercules, CA) were suspended to assist in removing cholate. Nanodisc samples were removed from dialysis after 24 hours and buffer-swapped twice via Micro Bio-Spin 6 columns (Bio-Rad, Hercules, CA) into 200 mM aqueous ammonium acetate (pH 7.0).

Data for melittin-embedded Nanodiscs prepared at the University of Arizona were taken from a previous publication, where their preparation is described in detail.[81] Briefly, dimyristoylphosphatidylglycerol (DMPG) Nanodisc samples with MSP1D1(−) scaffold proteins were prepared at room temperature similarly to the DMPC Nanodiscs described above, to a final concentration of 2.5 μM in 200 mM ammonium acetate (pH 6.8). Various dilutions of melittin (GenScript, Piscataway, NJ) in methanol were directly added to stock solutions of preformed

Nanodiscs to final Nanodisc concentrations of 2 μM in 13% methanol by volume, with molar ratios of melittin to Nanodiscs ranging from 0 to 24:1. Finally, a stock solution of 400 mM imidazole was added to the melittin-Nanodisc solutions as a charge-reducing agent, to a final concentration of 25 mM.

Native Mass Spectrometry. Mixed-MSP1D1 DMPC Nanodisc native mass spectra were acquired at the University of Oregon with a Synapt G2-S*i* ion mobility-mass spectrometer (Waters Corp., Milford, MA) using a static nanoelectrospray ionization (nanoESI) source at a capillary voltage of ~0.8 kV relative to instrumental ground and Trap/Transfer collisional energies of 75-100/5 V. This relatively high Trap potential was used to dissociate enough salt adducts from the native Nanodisc ions to resolve different scaffold protein compositions in the resulting mass spectra. NanoESI capillaries were prepared from 0.78 mm i.d. borosilicate capillary tubes on a Flaming-Brown P-97 micropipette puller (Sutter Instrument, Novato, CA) to a final i.d. of ~1 μm. Native mass spectra of melittin-embedded DMPG Nanodiscs were acquired at the University of Arizona in positive ion mode on a Q-Exactive HF quadrupole-Orbitrap mass spectrometer with Ultra-High Mass Range (UHMR) research modifications (Thermo Fisher Scientific, Waltham, MA) and a nanoESI source. Further experimental details for these melittin-embedded Nanodiscs are described in the original paper.[81]

Deconvolution and MMD Analysis. FT deconvolutions were performed via the Prell group's home-built program iFAMS v. 5.3. Lipid masses of 666.9 Da and 677.9 Da were used for DMPG and DMPC Nanodisc data, respectively, to determine the charge states present from FT spectra. The lipid polydispersity inherent to Nanodiscs results in multiple overlapped charge states in the mass spectrum, each with a broad distribution of lipid stoichiometries. However, the corresponding FT spectrum contains well-resolved peaks corresponding to fundamentals and harmonics for each charge state present, allowing for charge state deconvolution by inverse-FT of charge state-specific data.[45, 52, 55] MMD profiles for simulated data made use of 14 FT harmonics, while profiles of Synapt G2-S*i* data made use of 4 harmonics due to their lower resolution as compared to the Orbitrap UHMR data. All profiles for melittin-embedded Nanodiscs used 14 harmonics except for samples with melittin:Nanodisc ratios of 0 (10 harmonics), 1.5 (4 harmonics), and 24 (12 harmonics) due to differences in the density of data points in their mass spectra and the corresponding frequency range of their FT spectra.

Bayesian deconvolution was performed via UniDec v. 4.2.1, which generates a delta-function matrix to model the spectral data and iteratively adjusts charge state probabilities and ion abundances by comparing *m*/*z* peak intensities in the model and the spectral data.[36, 81] Input charge-state ranges were determined using iFAMS and lipid masses of 666.9 Da and 677.9 Da for DMPG and DMPC Nanodiscs, respectively, then entered in UniDec with two additional charge states above and below the found range. Output mass ranges were determined from the lowest *m*/*z* in the mass spectra multiplied by the lowest charge state and the highest *m*/*z* multiplied by the highest charge state. Melittin-embedded Nanodisc data were batch processed by MetaUniDec.[41] More detailed UniDec deconvolution parameters can be found in Tables S1 and S2.

MMD analysis was performed via both UniDec's built-in method, which uses the UniDec-deconvolved mass spectrum as input,[80, 81] and a new FT-based method in iFAMS (see Theory section below) that extracts spectral information from the complex phase of the Fourier Transform. In both cases, peaks in the reconstructed MMD profiles were centroided by integration between local minima to assign mass defects and calculate root-mean-squared-deviations (RMSDs) from the expected base masses (based on the theoretical exact masses of the scaffold proteins, lipids, and melittin). Scaffold protein mass contributions were expected to be 24661.6 Da for MSP1D1His and 22043.9 Da for MSP1D1(−), as reported by the manufacturer. MMD profile mass accuracy was calculated only for peaks present in both the FT-based and Bayesian MMD profiles. Richardson-Lucy peak sharpening was used to explore melittin-Nanodisc stoichiometry conclusions made in the original publication of the data,[81] using UniDec's built-in "Raw/Centroid mode" for the Bayesian profiles and an analogous Richardson-Lucy treatment (50 iterations using a gaussian point-spread function with a standard deviation of 30 *m*/*z*) for the FT-based MMD profiles.

Theory

Fourier Phase Analysis for Macromolecular Mass Defect Profile Reconstruction. Fourier MMD treatment assumes that the population of ions for a given charge state consist of one or more non-isobaric "base compositions" (e.g., protein isoforms, different polymer end groups, or heterogeneous protein oligomer masses) to which a polydisperse distribution of repeated subunits (non-covalent adducts like lipids/detergents, or covalent adducts like glycans in a

glycoprotein or monomers within a polymer) are adducted, and we call the masses of these base compositions "base masses". Let the *n* different base compositions be labeled B_1, B_2, …, B_n, and the repeated subunits have mass L, and finally assume that the net charge on each ion is supplied by charge carriers of mass X. Then, for charge state Z, we have a population of ions with stoichiometries $B_1L_jX_Z$, $B_2L_jX_Z$, …, $B_nL_jX_Z$, where j ranges over the (polydisperse) number of repeated subunits incorporated. It is assumed that, for each value of Z, the polydispersity in the number of lipids is sufficient to result in well-resolved peaks in the corresponding Fourier spectrum.[45] The goal of the present analysis is to determine the masses of B_1, B_2, and B_n *modulo* the mass of L as well as their relative abundances in the ion population. This procedure is referred to as Macromolecular Mass Defect (MMD) analysis,[80-82] and the goal of this paper is to demonstrate how it can be achieved using Fourier/Gábor Transform (FT/GT)-based methods as well as compare its efficacy to mass-domain MMD analysis such as those implemented in UniDec.[36, 41, 77, 78, 80, 81, 91]

In the following, we use similar notation to that in our previous publications on Fourier and Gábor Transform analysis of mass spectra.[45, 52, 53, 55] The mass spectrum for a particular charge state is represented as $s\left(\frac{m}{z}\right) = [c\left(\frac{m}{z}\right) * p\left(\frac{m}{z}\right)] \cdot e(\frac{m}{z})$, where $c\left(\frac{m}{z}\right)$ is a comb function with peaks spaced by $\frac{L}{z}$; $p\left(\frac{m}{z}\right)$ is the "peak shape" representing the shape of each peak in the comb; and $e(\frac{m}{z})$ is the envelope function that describes the abundance distribution of each adduct stoichiometry, L_j. $[c\left(\frac{m}{z}\right) * p\left(\frac{m}{z}\right)]$ itself is a periodic function with period $\frac{L}{z}$ which inherently encodes the *m*/*z* spacing as well as the relative abundance of B_1, B_2, …, B_n and mass spectral peak widths associated with each of them; and $e(\frac{m}{z})$ is typically much broader in *m*/*z* than $\frac{L}{z}$ for Nanodiscs, long-chain polymers, and many similar highly polydisperse ion populations.[45] The Fourier Transform of $s\left(\frac{m}{z}\right)$ is $S(k) = [C(k) \cdot P(k)] * E(k)$, with $[C(k) \cdot P(k)]$ a periodic function of period $\frac{2\pi Z}{L}$, and $E(k)$ now the shape of each peak in the frequency spectrum $S(k)$ (typically much narrower than the frequency spacing, $\frac{2\pi Z}{L}$). Our goal is to extract $[C(k) \cdot P(k)]$ from $S(k)$ and invert the FT to recover the *m*/*z* spacing as well as the relative abundance of B_1, B_2, …, B_n (*mod* L) and mass spectral peak widths associated with each of them.

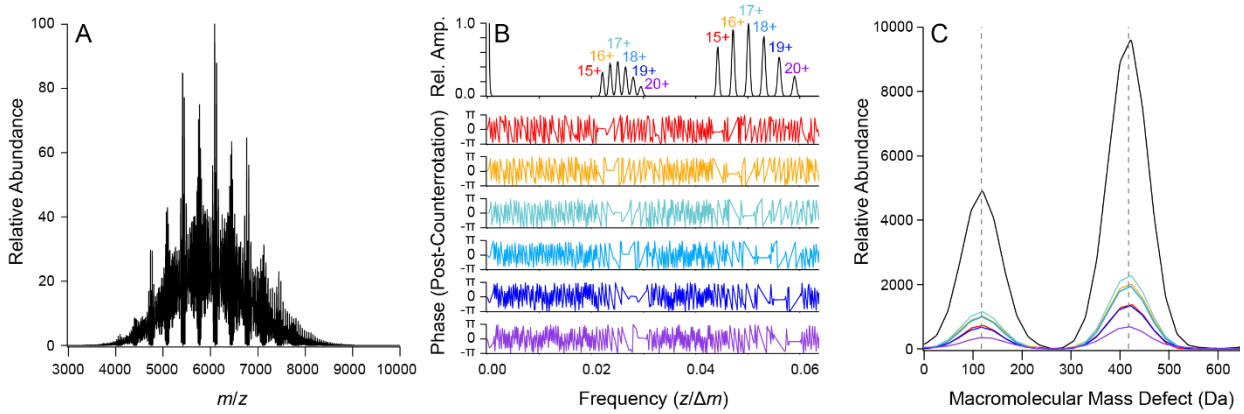


Figure 6. Illustration of MMD analysis on a simulated mass spectrum (A), containing charge states 15+ through 20+ of a polydisperse ion population containing multiple lipids with a mass of 678 Da and base masses with mass defects of 417 and 117 Da in a ratio of 2:1. (B) FT of the mass spectrum (top) with well separated charge-state-specific peaks, whose phase information (colored plots, corresponding to the charge states labeled in the FT) can be extracted after nullifying each charge state's specific global mass defect (flat areas in colored plots). These down-sampled, counterrotated Fourier-domain data for each charge state then undergo inverse-FT and are plotted modulo the mass of the polydisperse adduct (C).

At first glance, this might appear challenging, as the phase of $S(k)$ can oscillate very rapidly in the frequency domain, depending on where the maximum of $e(\frac{m}{z})$ is in the mass spectrum. (In fact, in the common case that the mass spectrum of the polydisperse ion population looks relatively "well-centered" within the *m*/*z* window being viewed, $S(k)$ oscillates in phase as rapidly as possible, i.e., by approximately π radians at each successive point in the frequency domain!) However, as long as the curvature of $e(\frac{m}{z})$ near its maximum is sufficiently low (i.e., the polydisperse ion population is in fact not very monodisperse), the Stationary Phase Condition[92] makes it possible to counter-rotate the phase of $S(k)$ by multiplying it pointwise by $\exp\left(-ik\left(\frac{m}{z}\right)_0\right)$, where $\left(\frac{m}{z}\right)_0$ is an *m*/*z* value at or near the maximum of $e(\frac{m}{z})$. This is mathematically equivalent to "sliding over" the original mass spectrum to the left by $\left(\frac{m}{z}\right)_0$, such that the top of $e(\frac{m}{z})$ is now very close to 0 *m*/*z* (an explicit proof of this mathematical transformation can be found in Appendix Chapter II). The phase of the resulting function $S'(k) = S(k) \cdot \exp\left(-ik\left(\frac{m}{z}\right)_0\right)$ will tend to be nearly flat at each peak in $[C(k) \cdot P(k)]$ as shown in the colored phase plots of Figure 6B, making it trivial to visually "read off" and/or numerically determine the phase and magnitude of $[C(k) \cdot P(k)]$ at each of its peaks, except possibly at $k = 0$, but it is a theorem that the phase there must be 0 because the mass spectrum is

real-valued. One needs to simply invert the recovered comb spectrum $[C(k) \cdot P(k)]$ to recover an image, averaged over all L_j states, of the *m*/*z* spacings and relative abundance of B_1, B_2, …, B_n (*mod* L) and their mass spectral peak widths. This spectrum is the MMD profile of the ion population. Unique advantages of this approach over "direct" approaches within in the *m*/*z* domain are that information from different charge states can in many cases be easily isolated and used to determine charge-state-specific MMD profiles, white noise at almost every frequency is eliminated, and a deconvolved "zero-charge" mass spectrum (in this case, *mod* L) can be readily reconstructed as desired from the individual charge-state-specific reconstructions.

Mathematically and empirically, two challenges that sometimes arise in this approach are identifying appropriate $(\frac{m}{z})_0$ values to use for each charge state and handling the ambiguity of frequency-domain charge state magnitudes at $k = 0$, where signal from all charge states overlap. In solving the first challenge, it is often useful to first analyze the FT spectrum of the initial mass spectral data (i.e., without counter-rotation) to reconstruct *m*/*z* envelopes for each charge state, as described in our previous publications.[45, 52, 55] These *m*/*z* envelopes can then be straightforwardly used to guess $(\frac{m}{z})_0$ needed for the phase reconstruction described above by computing the centroid of each envelope. This is generally easy to do, as long as $e(\frac{m}{z})$ has a clear maximum. The second challenge is circumvented by assuming both the phase and magnitude at k = 0 are zero. Because mass spectra are real valued by definition, the phase at $k = 0$ must be zero. We can also justify the zero-magnitude assumption for well-resolved ion populations with low baseline in the *m*/*z*-domain, where the integral over the mass spectrum and thus the FT magnitude at k = 0 for each charge state is relatively small and setting it equal to zero for each charge state before inverse FT merely adjusts the baseline of the reconstructed MMD profile by a small amount. For poorly resolved ion populations and/or those with a large noise baseline, the magnitude of the FT at k = 0 can be significant, and the apparent zero of abundance in the MMD profile will be systematically low by an indeterminate constant, although peak positions and shapes in the MMD are well-preserved. This baseline discrepancy also occurs in mass-domain MMD, where noise sums up across the spectra to yield a significant baseline, and in both cases can be compensated for by a linear baseline correction.

Results and Discussion

To explore the limits and robustness of the FT-based MMD analysis, ranging from signal-to-noise levels (Figure 7) to peak resolution (Table S3 and Figure S1), we start our discussion of the method with simulated spectra for which we know exactly the mass defects and their relative abundances. From there, we benchmark the method against the mass-domain MMD analysis integrated in UniDec using published, high-resolution Orbitrap spectra of melittin-embedded Nanodiscs (Figure 8). Finally, to illustrate the ability of the algorithm to handle lower-resolution data, as on currently available commercial Q-IM-MS instruments, we use it to investigate incorporation of Nanodisc MSPs with different masses using mass spectra acquired on a Synapt G2-S*i* Q-IM-TOF mass spectrometer (Figure 9).

Simulated Data. As a proof of concept, a model spectrum was constructed to contain charge states 15+ through 20+ of a polydisperse population of DMPC lipids (mass approximated to 678 Da) with a 2:1 ratio of base masses with MMDs (*mod* 678 Da) of 417 Da and 117 Da, respectively. The resultant total MMD profile peaks were of a similar full-width half-maximum to the peaks in the original model spectrum (~47.1 *m*/*z*) and also preserved the 2:1 ratio of mass defect relative abundances. The analysis was repeated with several levels of white noise relative to the model spectrum's global maximum, returning nearly identical results to those from a mass spectrum with a signal-to-noise ratio (SNR) of 5:1 (Figure 7, black), well below the commonly accepted limit of quantitation (SNR of 10:1).[93] The ability of the FT-based MMD algorithm to perform well at such a low SNR arises from elimination of white noise contributions at almost all frequencies other than at the harmonic peaks where the phase is sampled. However, the fewer harmonics that are sampled, the fewer total data points the MMD profile contains (reducing profile resolution) and the more likely it is, in principle, that artifacts may dominate the MMD profile. Due to the ambiguity of charge-state-specific Fourier-domain amplitude at 0 frequency, as described above, a baseline correction was applied to the reconstructed MMD profiles to account for relative abundance distortion from the summation of charge-state-specific MMD reconstructions. The difference between applying baseline correction to the individual charge state mass defect profiles and applying the correction to the summation of the charge state mass defect profiles was found to be negligible (see Figure S2). These results indicate that this FT-based MMD algorithm can be exceptionally useful for MMD analysis of low SNR data due to its high noise tolerance.

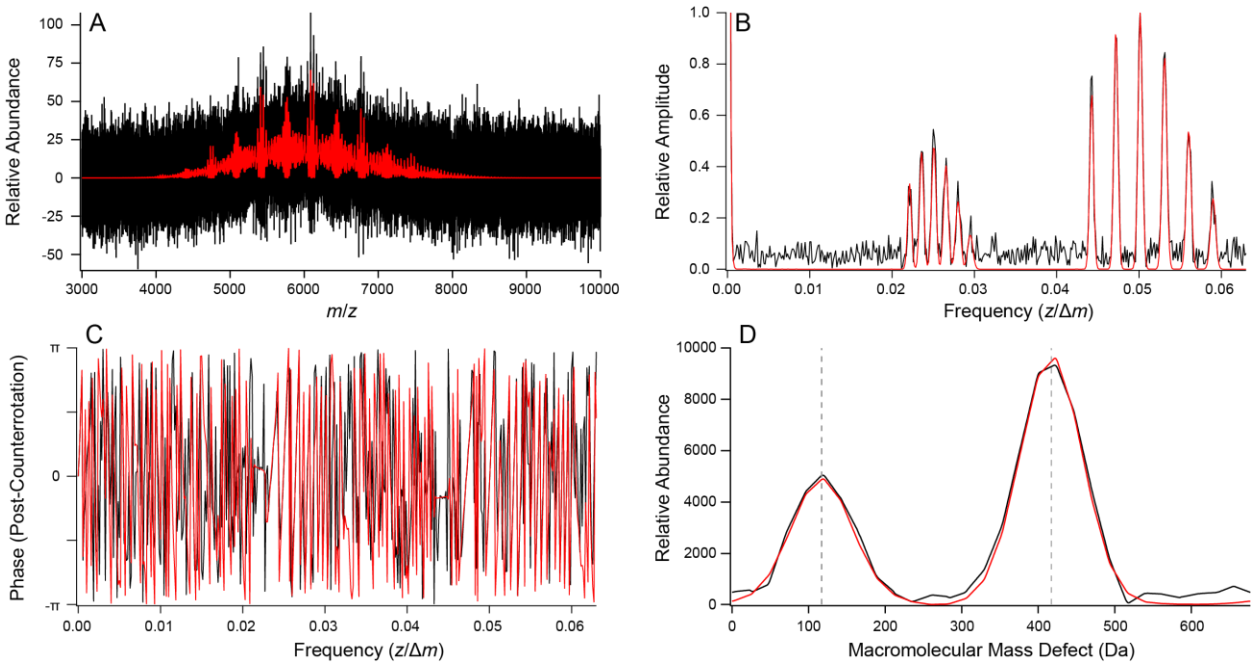


Figure 7. MMD analysis of simulated data containing charge states 15+ through 20+ of a polydisperse ion population containing multiple 678 Da lipids and base masses with mass defects of 417 Da and 117 Da in a 2:1 ratio. Red/black traces correspond to infinite/5:1 SNR. Simulated mass spectra (A), corresponding FT magnitude spectra (B), FT phase spectra (C), and reconstructed MMD profiles (D). SNRs are calculated as the ratio of the global maximum of the simulated *m*/*z* data to the root-mean-square white noise. Despite the low SNR of the mass spectrum , selection of frequency data only near FT peaks sufficiently filters out white noise to return MMD profiles comparable to those reconstructed from noiseless mass spectra.

Melittin-Embedded Nanodiscs. To compare the performance of the FT-based method with mass-domain MMD analysis, previously published experimental data collected by the Marty lab for melittin-embedded DMPG Nanodiscs were analyzed via both iFAMS and UniDec, using 666.9 Da as the molecular weight of DMPG. In the original publication,[81] the MMD profile information was distilled in plots tracking changes in the relative abundance of incorporated melittin stoichiometries as a function of bulk solution melittin:Nanodisc ratios, which we have recreated here for both UniDec and iFAMS MMD profile series. Features in each were assigned to the same predicted mass defects for the sake of comparison. In these data, most of the MMD information has a resolution (defined as peak separation divided by peak fwhm) of 1.5 or greater, indicating that good results are obtained at this resolution despite the overlap of many charge states and individual peaks in the corresponding polydisperse mass spectra, where adjacent peaks are often strongly overlapped. (An upper bound to the lowest tractable resolution in MMD profiles is identified in Figure S1, where a distinct shoulder peak can still be discerned at resolutions as low as 1.06 despite the 2:1 difference in peak height.) The RMSDs and melittin-

incorporation plots generated from MMD peak integration reflect the remarkable consistency with which the FT phase algorithm mirrors the results from UniDec for all 7 mass spectra, representing bulk melittin:Nanodisc molar ratios ranging from 0:1 to 24:1. Mass defects and relative abundances using both methods (see Table S4) were within the uncertainties of the previously published UniDec results.[81]

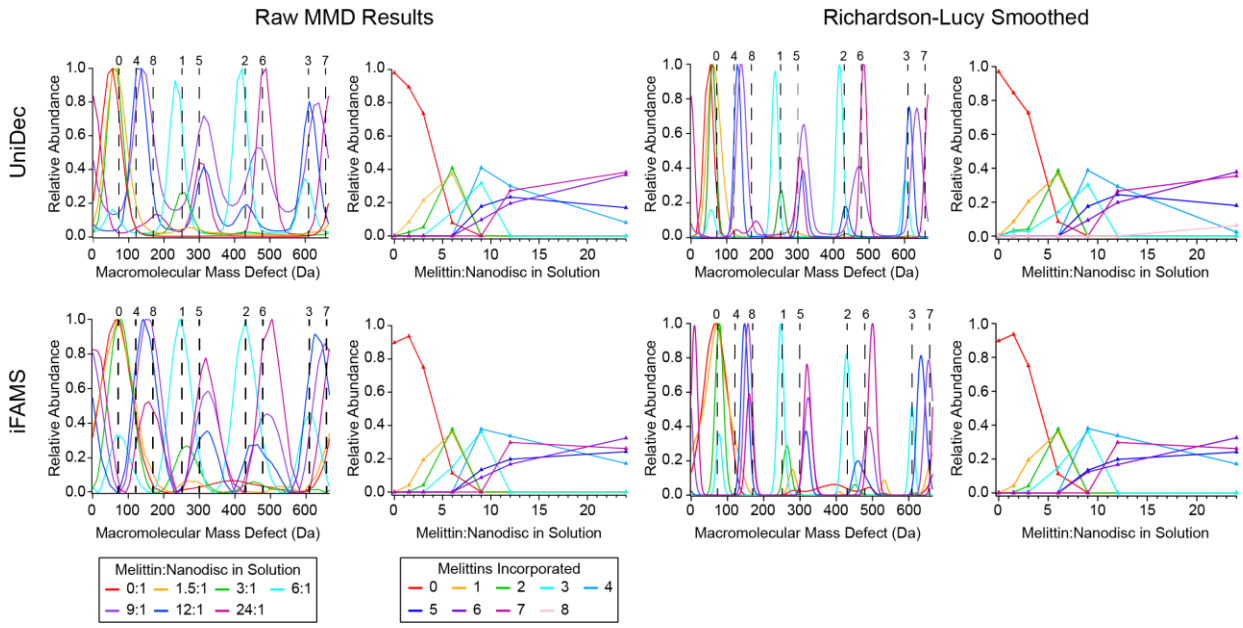


Figure 8. Comparison of melittin-embedded DMPG Nanodisc data deconvolved via UniDec (top) and iFAMS (bottom). MMD profiles are normalized to their global maximum. Dashed lines in MMD profiles indicate the expected mass defects for various numbers of incorporated melittin. MMD profile color legend (left) and x-axis of the melittin-incorporation plots indicate the bulk solution melittin-to-Nanodisc ratios prepared, while the melittin-incorporation plot color legend (right) signifies the melittin-incorporation ratio being tracked as a function of bulk concentrations

Minor differences between UniDec and iFAMS MMD profiles include the appearance of a broad, low-abundance peak around 400 Da in the iFAMS profiles for samples containing no melittin that is much larger than its counterpart in the UniDec MMD profile. This minor peak was not assigned in the original publication and is most likely contributed by a small population of cholate (expected mass of 407.56 Da) in the Nanodiscs which could be removed with further dialysis. A second minor difference between the UniDec and iFAMS MMD profiles shown in Figure 8 is the presence of a single peak at 180 Da in the iFAMS and UniDec MMD profiles for the solution with a bulk concentration ratio of 24 melittin monomers per Nanodisc, which

appears as two lower abundance peaks in the published UniDec MMD profile. These discrepancies are attributed to artifactual splitting of broad peaks that occurs when using a much narrower target peak shape, a well-known occurrence when using Richardson-Lucy sharpening algorithms.[91, 94]

Comparison of the MMD profiles presented in Figure 8 after Richardson-Lucy peak sharpening reveals another minor difference between results from the two MMD analysis methods, namely, that the peak at ~180 Da in the UniDec reconstruction assigned to incorporation of 8 melittins for the 24:1 bulk melittin:Nanodisc sample is absent in the iFAMS MMD reconstruction. This is attributed to slight differences in the breadth of the un-sharpened peak, which splits into two peaks after Richardson-Lucy sharpening for the UniDec data but remains a single, narrow feature for the iFAMS reconstruction. This result indicates that care must be exercised in interpreting results from Richardson-Lucy sharpening for initially broad peaks. From the results both with and without Richardson-Lucy sharpening, we find generally excellent agreement between both the MMD profiles and melittin incorporation profiles as a function of bulk melittin:Nanodisc concentration for both the FT phase reconstruction method in iFAMS and the mass-domain results from UniDec. These results confirm the previously published conclusion that melittin incorporation is essentially non-specific in these samples.

Nanodisc Scaffold Protein Incorporation. Having demonstrated the FT MMD analysis method's ability to replicate and verify mass-domain MMD results for high-resolution data, we applied the FT algorithm to identify and compare scaffold protein compositions of Nanodiscs formed in a mixture of two types of MSP1D1 in data acquired on a lower-resolution (Q-IM-TOF) instrument. These MSPs (with and without the histidine tag) are commercially available (Sigma-Aldrich, St. Louis, MO) for use in Nanodisc experiments, and the His-tagged variant can be used to immobilize the MSP on nickel resin for membrane protein incorporation or for purification purposes. One strategy for maximizing Nanodisc formation using nickel resin immobilization is to first immobilize the His-tagged variant and add additional non-His-tagged variant along with lipids and membrane protein during the Nanodisc synthesis steps. It is therefore of interest to determine to what extent His-tagged and non-His-tagged MSPs incorporate into mixed-MSP Nanodiscs. Given both MSP1D1 monomers and dimers are observed in native mass spectra of solutions containing this protein, indicating both states are present in solution (see Figures S3 and S4), we addressed the broader question of what Nanodisc

MSP compositions are obtained when His-tagged and non-His-tagged MSPs are used to synthesize Nanodiscs in solution.

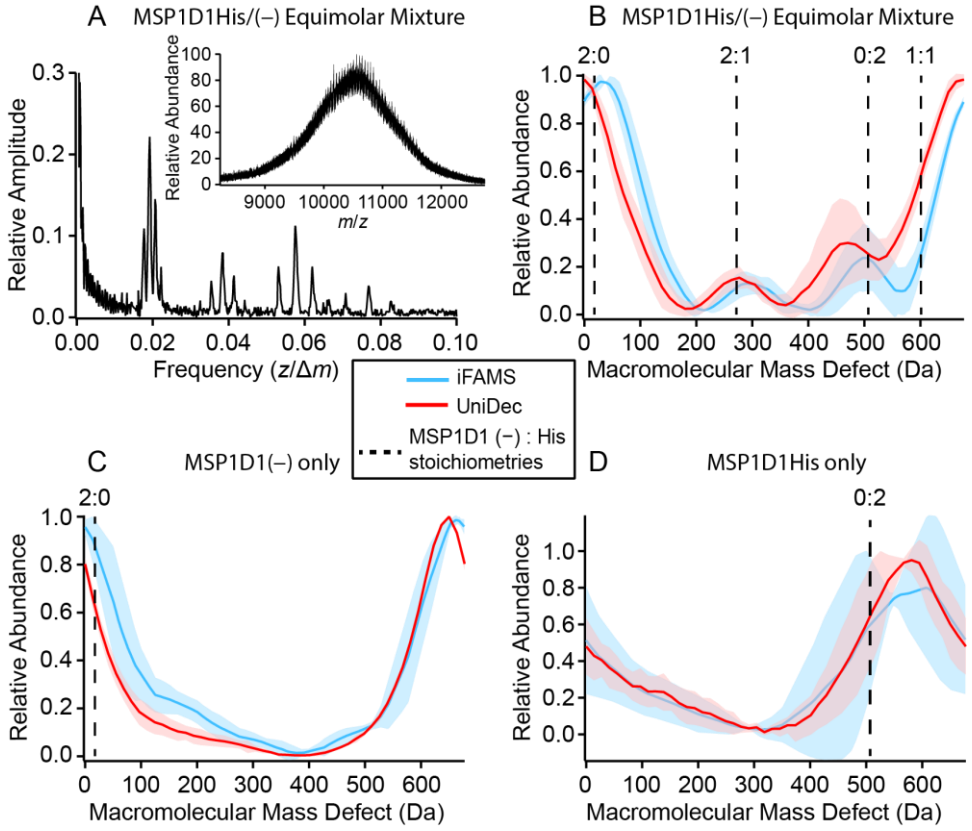


Figure 9. MMD analysis of "empty" MSP1D1 Nanodiscs with a variety of MSP incorporation states. (A)FT of Nanodiscs samples formed from a mixture of MSP1D1(−) and MSP1D1His. The inset shows the original mass spectrum truncated to just the Nanodisc distribution, serving as a reminder of the degree of charge state overlap from which the well-resolved charge state information in the FT is produced. Average MMD profiles of triplicate data deconvolved via both UniDec (red) and iFAMS (blue) are shown for MSP1D1 Nanodiscs prepared with both variants of MSP1D1 (B) and with only MSP1D1(−) (C) or MSP1D1His (D). Shaded regions denote 1 standard deviation in the abundance of the triplicate data and dashed lines indicate expected mass defects from various stoichiometries of MSP1D1(−) and MSP1D1His.

DMPC Nanodisc solutions prepared from a mixture of MSP1D1His and MSP1D1(−) were analyzed to determine whether the resulting Nanodiscs demonstrate a preference in MSP incorporation (see average MMD profiles in Figure 9). Despite the degree of charge state overlap observed in the mixed MSP Nanodisc mass distribution (inset of Figure 9A) the charge states are well-resolved in the frequency domain, allowing for identification of multiple base compositions from the MMD profiles. The dominant peaks for all samples correspond to MSP homodimers (expected *mod* 677.9 Da mass defects of 24.3 Da and 514.4 Da for MSP1D1(−) homodimer and

MSP1D1His homodimer, respectively), and there is little to no evidence heterodimers at the expected mass defect of 601.5 Da. The peak around 281.5 Da in the mixed MSP profiles indicates the presence of Nanodiscs containing three MSP monomers, a phenomenon which has been reported previously[95] and is believed to have physiological analogs (i.e., in mature lipoprotein particles containing three monomers of apolipoprotein A-1 rather than MSP1D1).[96] Although relative abundances for the triplicate experiments exhibit enough variability to make quantitation of the different protein stoichiometries difficult, peaks corresponding to at least three and possibly four different protein stoichiometries are clearly visible among the MMD profiles. Overall, the agreement of the measured MMD peak positions with their theoretical values is somewhat worse for these lower-resolution data than for the high-resolution melittin-Nanodisc data described above, presumably due to salt or other co-solute adduction in combination with the poorer resolving power of the Q-IM-TOF instrument. However, agreement of the FT-based algorithm results with theory is similar to or, in one case (the 514.4 Da peak), slightly better than the mass-domain algorithm (see Figure 9B).

Conclusions

Here, we have introduced a FT phase-based approach to MMD analysis in the program iFAMS which is complementary to existing mass-domain deconvolution tools (such as that in UniDec), allowing for orthogonal verification of MMD profile results. The spectral down-sampling of the FT method filters out nearly all background random noise to produce reliable MMD profiles from spectra with SNRs as low as 5:1, expanding the analysis capabilities of lower-resolution instruments, such as Q-TOF instruments, or for tandem experiments such as ion mobility spectrometry and surface-induced dissociation, which are not as widely available at present on high-resolution (Orbitrap and FT/ion cyclotron resonance) instruments. Comparison of results from the iFAMS and UniDec algorithms confirms previously published results for melittin incorporation into Nanodiscs and provides strong evidence for the accuracy of both deconvolution methods. Results presented here using both mass-domain and FT-based MMD profile reconstruction of Nanodiscs formed from an equimolar mixture of MSP1D1(−) and MSP1D1His indicate Nanodiscs containing two MSP1D1(−) more readily assemble into these Nanodiscs.

The FT phase-based method presented here has potential applications in extracting isoform profiles from a variety of heterogeneous samples exhibiting polydispersity, from branched polymers and polymer-conjugated protein therapeutics used to increase drug longevity *in vivo*[97] to membrane protein-lipid/detergent complexes, as well as reconstructing isotope profiles for highly complex mass spectra which still exhibit isotope resolution. The new algorithm also provides a straightforward means of confirming analyses obtained from other orthogonal approaches, such as the mass-domain methods in UniDec, PMI Intact, and other deconvolution tools.[13, 98] The FT MMD reconstruction algorithm described above has been incorporated into iFAMS v. 6.1 with both FT and GT-based implementations, which is publicly available as an open-source Python program at https://github.com/prellgroup/iFAMS/releases. Recent versions of UniDec are publicly available as an open-source Python program at https://github.com/michaelmarty/UniDec/releases.

Bridge

iFAMS MMD analysis introduces a noise-tolerant method of mass defect analysis with comparable mass accuracy to state-of-the-art *m*/*z*-domain approaches, enabling cross-validation of results retrieved from model optimization. Most applications of macromolecular mass defect deconvolution to date have been on titrations of single analytes into polydisperse transporters with little study of reaction mixtures with multiple components or the complications of non-specific small ion adduction states. In continuing this work with Genentech Inc. to study more multiplexed mass heterogeneities in branched polymers, we developed a workflow for the analysis of polymer functionalization intermediates during active synthesis, in tandem with the separative capabilities of online liquid-chromatography mass spectrometry (LC-MS).

CHAPTER III

STREAMLINING LC-MS CHARACTERIZATION OF PHARMACEUTICAL POLYMERS BY FOURIER-TRANSFORM BASED DECONVOLUTION AND MACROMOLECULAR MASS DEFECT ANALYSIS

First author paper co-authored with Christopher M. Crittenden, Simon A. Chan, Samuel H. Yang, Dawen Kou, James S. Prell and Bifan Chen. Christopher M. Crittenden and Bifan Chen produced the LC-MS data and provided feedback on the workflow. James S. Prell supported the development of the algorithm.

Introduction

The field of conjugating therapeutic agents such as peptides, proteins, and antibodies to synthetic polymers has grown remarkably over the years, seeing many of these research efforts successfully develop into clinical explorations.[99] The benefits of incorporating polymeric scaffolds include enhancing stability and solubility, controlling delivery, increasing efficacy, improving pharmacokinetics, and reducing toxicity. Advances in synthetic and conjugation chemistry continue to expand the repertoire of pharmaceutical polymers with different architectures and functionalities. However, keeping pace with the increasing complexity of these polymer conjugates remains a daunting task for thorough analytical characterization, particularly with highly branched and functionalized polymers of high molecular weight.

To characterize pharmaceutical polymers, liquid chromatography (LC) is arguably the most widely used technique, which can separate polymer samples based on chemical composition, molecular weight, end-groups, and branching.[100] LC can also be hyphenated with various detection methods with different pros and cons,[100] among which mass spectrometry (MS) including charge detection mass spectrometry (CDMS)[101] is playing an emerging role, offering absolute mass determination with unrivaled insight on molecular features.[102] While LC-MS hyphenation via electrospray ionization (ESI) is powerful, the biggest hurdle to unleashing its full capability is dissecting the complicated mass spectra generated, especially for the high-molecular-weight polymers. The intrinsic heterogeneity from polymer samples together with the multiple charge states caused by the ESI processes can create extremely congested and

overlapping spectra. To make matters worse, numerous non-covalent adducts often attach onto polymer ions during ESI, further complicating the interpretation. Since the mass spectra are extremely challenging to analyze by conventional approaches, previous efforts have been made to decrease spectra complexity by reducing the charge state distributions toward a narrower range of lower charge states, either by the use of a 210Po alpha particle source[103] or by the introduction of charge-reducing agents such as organic amines.[104, 105] However, these setups for charge reduction are often quite complicated and the method sensitivity can be compromised. We therefore focused on alternative solutions that directly address the data analysis challenge without requiring modification of common LC-MS setups.

Among different data processing approaches, Kendrick mass defect (KMD) analysis, which uses modular arithmetic to reduce complexity of polydisperse polymer mass spectra by focusing on the remainders from dividing the polymer's molecular weight by the mass of the repeated polymer subunit, has been widely used in the polymer field for identifying sample impurities and end-groups.[83, 84, 106] Recent theoretical advances incorporating fractions of a polymer subunit have extended the utility of KMD analysis to simple multicharged spectra,[85] however, the utility of KMD analysis directly from ESI spectra of heterogeneous, high-molecular-weight polymers remains limited. Conversely, the broad mass polydispersity of polymeric samples makes their corresponding mass spectra well-suited for Fourier analysis,[51] which can be used to convert their highly congested mass spectra to much simplified peak features in the frequency domain by Fourier-transform based deconvolution algorithms such as iFAMS.[45, 55] Here, we showcase how the iFAMS deconvolution algorithm and the associated Macromolecular Mass Defect (MMD) analysis can greatly streamline LC-MS analysis of pharmaceutical polymers, providing key information on size, composition, branching, end-groups, and potential impurities.

Experimental

All polyethylene glycol (PEG) samples were analyzed via reversed-phase LC-MS using a 4.6 x 150 mm, 3.5 um Waters Xselect CSH Phenyl-Hexyl column heated at 40°C. 0.1% formic acid in water and 0.1% formic acid in acetonitrile were used as mobile phase A (MPA) and mobile phase B (MPB), respectively. The following gradient with 1 mL/min flow rate was used: 0 min at 10% MPB; 2 min at 10% MPB; 5 min at 45% MPB, 10 min at 46% MPB, 12 min at

90% MPB; 13 min at 90% MPB, 13.1 min at 10% MPB; and 15 min at 10% MPB. 6k PEG samples and reaction mixture were analyzed on an Agilent 6545/XT Q-ToF mass spectrometer and a Thermo Orbitrap Fusion Lumos Tribrid mass spectrometer with standard parameters. 20 kDa, 30 kDa, and 40 kDa PEG samples were analyzed on the Q-ToF. Sample sourcing and reaction mixture preparation details can be found in Appendix Chapter III.

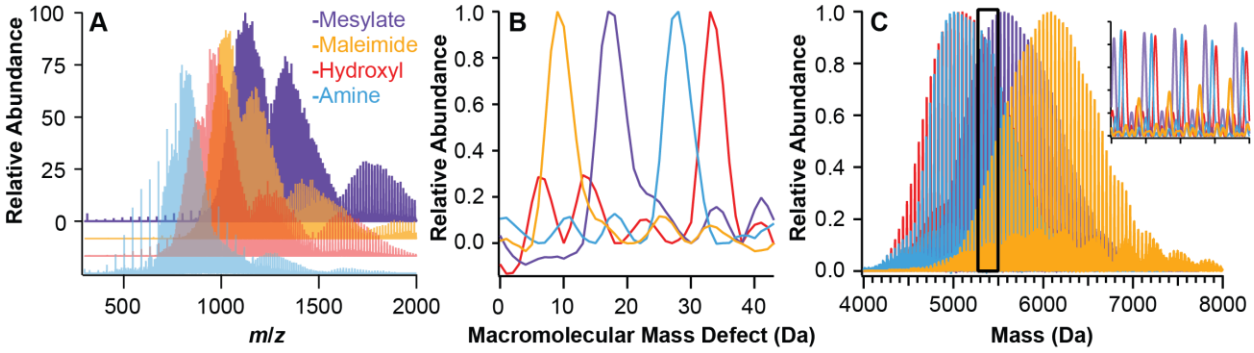
Data Analysis

Structural characterization of polymeric samples from complex LC-MS data requires several sequential steps, the theory of which has been detailed elsewhere,[45, 52, 53, 107] so brief explanations of each step will be included as they are discussed and the overall workflow and strategy are shown in Figure S5. First, mass spectra are collected by tandem reversed-phase liquid chromatography (RPLC) and conventional ESI-MS, then extracted to comma-delimited files (CSVs) from retention time peaks in the total ion chromatogram (TIC) for each chromatographically separable species (Figure S5A). iFAMS deconvolution software (open-source, publicly available software developed by the Prell group) automatically applies a Fourier transform (FT) or *m*/*z*-localizing Gábor transform (GT) to imported mass spectra, separating mass domain signal into frequency channels based on the peak spacings.[14, 52, 53] The periodic mass distributions inherent to polymeric samples lend themselves well to FT/GT analysis, with peak spacings dependent on the mass shift (ΔM) and charge state (Z) corresponding to sharp features in the frequency domain at fundamental frequencies Z/ΔM, enabling identification of the polymer composition and charge states (Figure S5B). GT features localized to the same *m*/*z* and occurring at integer multiples of Z/ΔM are higher harmonics of the same ion population; a greater number of harmonics visible in the spectrum corresponds to greater peak resolution in the original mass spectrum. Inverse FT/GT of charge state series of interest effectively filters out signal from interferents such as lighter polymer fragments and salt clusters as well as the majority of instrumental background noise, yielding deconvolved charge-state-specific mass reconstructions. Inclusion of more harmonic information generally increases feature resolution in these deconvolved mass spectra, and they can be summed as desired to identify polymer size from overall mass distributions (Figure S5C).

Modifications to the base polymer such as varying end-groups and the cores of multibranched polymers lead to more subtle shifts of the polymer mass distributions,

corresponding to measurable phase shifts in the frequency domain. With the polymer composition identified from FT/GT, the phase shifts can be measured as macromolecular mass defects (MMDs) *modulo* the average repeating unit mass.[107] The resultant charge-state-specific MMD profiles are compared to determine which features are supported by all charge states (Figure S6) and subsequently integrated between local minima to calculate peak centroids and peak widths (defined as 4 standard deviations, 4σ). Mass defects corroborated by the majority of the charge-state-specific MMD profiles are compared to predicted combinations of known sample components, calculated from average molecular weights (Table S5), to determine polymer core, branching, and end-group structures (Figure S5D). Statistically significant combinations of these aperiodic components can then be subtracted from the deconvolved mass reconstructions to determine polymer chain length distributions (i.e., how many repeating units make up the polymer).

Results and Discussion



D	Hydroxyl Functionalized		Mesylate Functionalized		Amine Functionalized		Maleimide Functionalized	
	Theor.	Exptl.	Theor.	Exptl.	Theor.	Exptl.	Theor.	Exptl.
Centroid	34.014	33.5	17.941	17.8	28.110	28.07	9.723	10.1
Width (4σ)		7.0		16.2		8.24		9.7
RMSD		0.2		0.9		0.04		0.4

Figure 10. Fourier transform based deconvolution and macromolecular mass defect analysis of 6-arm polymer standards functionalized with different end-groups. (A) Overlaid raw spectra of each polymer standard (red: hydroxyl, purple: mesylate, blue: amine, and yellow: maleimide). (B) Deconvolved mass spectra for each polymer standard. (C) Experimental MMD profiles for each polymer standard. (D) Comparison of average experimental (Exptl.) MMD values for each polymer standard to theoretical (Theor.) mass defects calculated from average molecular weights.

As a proof-of-concept of the workflow, we analyzed four 6-arm polymer standards with different end-group functionalizations but identical core structures. Synthesis of these standards starts with converting end-groups from hydroxyls to mesylates, then to amines, and finally to maleimides, which are used in protein conjugation applications.[108] The raw mass spectra of the four 6-arm polymer standards contain overlapping charge states from 3+ to 7+, with *m*/*z* domain localizations linked to differences in both mass and average polymer ion charge (Figure 10A). After GT analysis, the deconvolved mass spectra reveal that the overall polymer sizes are about 6 kDa and the polymer repeating unit is 44.04 ± 0.09 Da, consistent with polyethylene glycol (PEG; expected unit mass 44.053 Da) (Figure 10B inset). Even from the deconvolved spectra we can observe mass differences between the four standards, though notably the hydroxyl (red) and amine (blue) functionalized standards have nearly identical mass distributions. To identify these subtler mass variations, frequency domain phase shifts are extracted into MMD profiles *modulo* the average repeating unit mass determined from the deconvolved spectra. The experimental MMD profiles generated from these 6 kDa PEG standards demonstrate that each functionalization consists of a single major feature (Figure 10C) with root-mean-squared-deviations (RMSDs) relative to theoretical MMD values within 1 Da (Figure 10D).

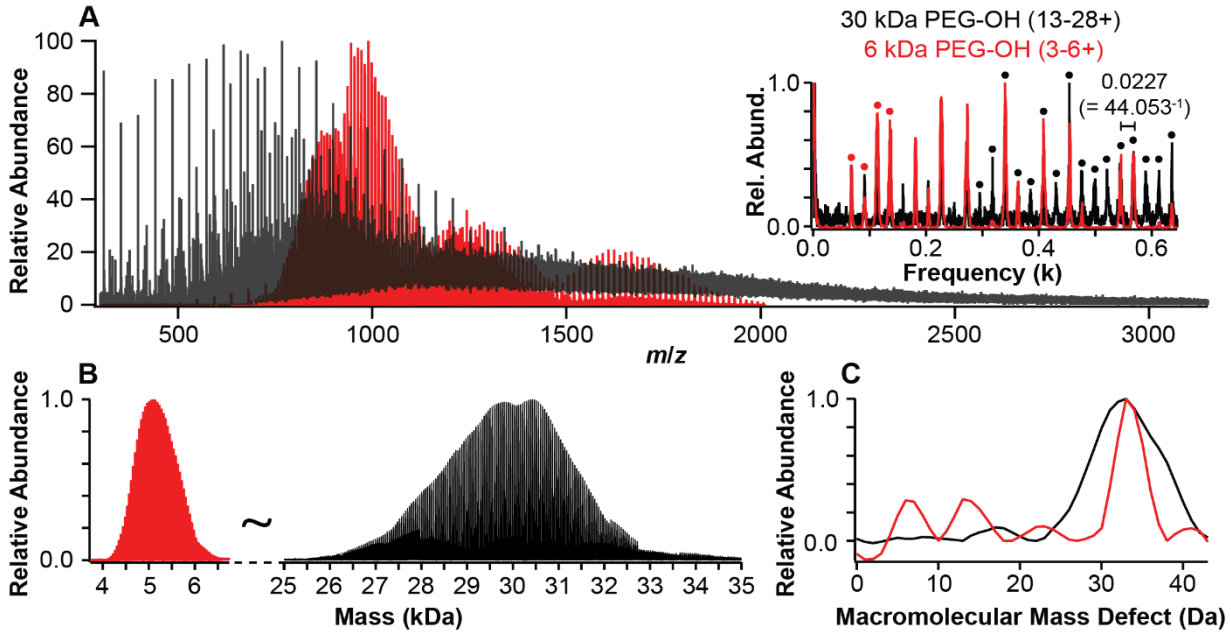


Figure 11. Fourier transform based deconvolution and macromolecular mass defect analysis of larger polymers. (A) Overlaid raw mass spectra, FT spectra, (B) deconvolved mass spectra, and (C) experimental MMD profiles of the 6 kDa and 30 kDa hydroxyl functionalized 6-arm PEG molecules. Red and black dots in the FT spectra (A, inset) label signals at fundamental charge state frequencies identified for the intact 6 kDa and 30 kDa polymers, respectively (unlabeled peaks at higher frequencies are harmonics of lower frequency signals).

While the charge states of the 6 kDa PEG scaffolds are still visually distinguishable, analysis of the mass spectra gets even more complex as polymer sizes increase. This is where Fourier-based deconvolution proves particularly useful, as exemplified by comparison of 6 kDa and 30 kDa PEG polymers prepared with the same end-groups and core structures (Figure 11). Although their sizes are very different, the average charge state of the electrosprayed polymer ions scales proportionally with polymer mass, causing their charge state distributions to populate similar *m*/*z* ranges (500-2000 *m*/*z*) (Figure 11A). Despite the strong overlap between all the PEG charge states, because polymer peak spacing is inversely proportional to charge state, FT/GT analysis in iFAMS readily separates polymer populations by composition and size across both *m*/*z* and frequency (Figure 11A, inset), effectively filtering out instrument noise and "polymer dust" signal (i.e., small interfering polymer fragment ions, which appear as spikes in the 30 kDa PEG spectrum). The deconvolved masses were consistent with the molecular weights specified from the vendors (Figure 11B). The mass defect profiles for both 6 kDa and 30 kDa PEG (Figure 11C) are very similar, with peak width differences stemming from larger polymers having greater isotopic variance (Figure S7), confirming that the two standards have the same end-groups (hydroxyls), core structure and branching (6-armed dipentaerythritol) despite very different mass distributions. Figure S8 further demonstrates iFAMS distinguishability of heavy, multibranched PEG standards of varying sizes, with mass defects corresponding to the degree of branching and differences in core structure. Different from 6-arm core of the 6k Da and 30kDa PEG-OH, MMD analysis reveals that the 20 kDa has a 4-arm core while the 40k Da PEG-OH has an 8-arm core with the potential presence of truncated species.

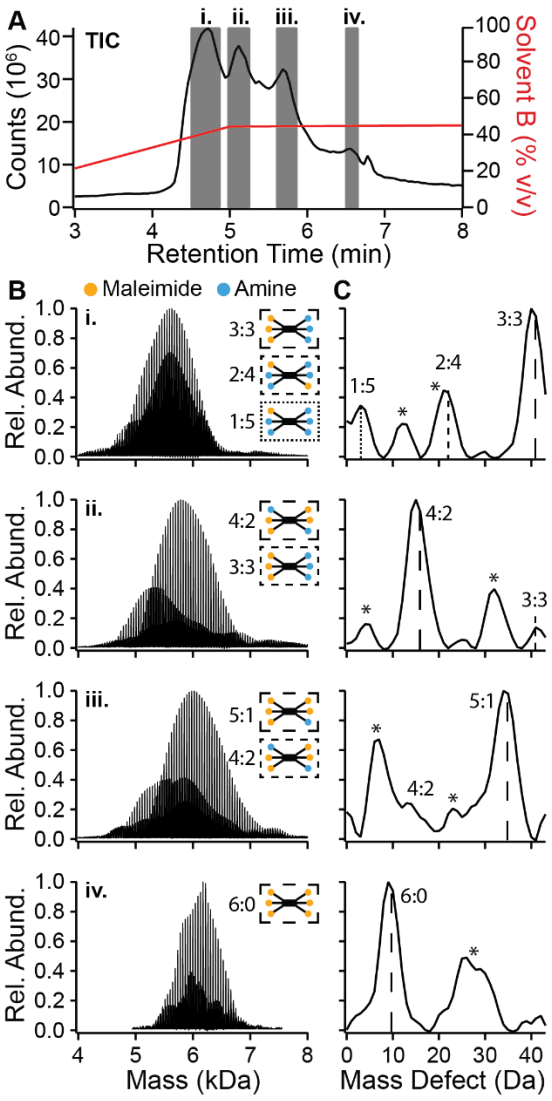


Figure 12. Fourier transform based deconvolution and macromolecular mass defect analysis for polymer reaction version monitoring. (A) Total ion chromatogram of the reaction mixture 60 minutes into the reaction (major peaks i, ii, iii, and iv). (B) Deconvolved mass spectra and (C) corresponding experimental MMD profiles of i, ii, iii, and iv. Dashed boxes and lines label end-group stoichiometries. Asterisks (*) denote low-level ammonium adduction states.

The iFAMS Fourier analysis workflow is also readily applicable to reaction monitoring experiments, enabling quality control before, during, and after each step of the polymer production process. We demonstrate this application with an in-progress maleimide substitution reaction (Figure 12), whose total ion chromatogram (TIC) from RPLC readily shows a few distinctive peaks (i, ii, iii, and iv), presumably representing the reaction intermediates. As the ratio of organic phase is increased (Solvent B, Figure 12A) the average mass of the eluted polymers gradually increases (Figure 12B), identified as increases in maleimide stoichiometry by

MMD analysis (Figure 12C). Maleimide:Amine end-group stoichiometries from 1:5 to 6:0 were observed in the MMD profiles, with additional lower abundance features resulting from co-elution of reaction intermediates and ammonium adductions (+17 Da) which occur readily with polyethers.[109] Although the level of ammonium adduction for the mass spectra in Figure 12 is relatively low and assignable (Figure S9, Table S6), it would be preferable to remove as much ammonium as possible experimentally prior to LC/MS, as desalting by gas-phase ion activation is not ideal when studying conjugated polymer scaffolds and is less effective for desalting high mass polymers. In the cases where several adducts are typically observed (Figure S10), MMD profiles can also get congested, severely obfuscating unique mass defect assignment.[110]

Conclusions

Thorough characterization of intrinsically heterogeneous polymer samples is a complex problem, but a crucial one to solve for the continued development of therapeutic polymer conjugates. As demonstrated, we provide a more generalizable solution for LC-MS characterization of polymer subunits, size, branching structure, and end-groups via Fourier-based deconvolution and MMD analysis. Our approach extracts critical attributes from polymer samples by taking advantage of mass polydispersity as frequency signal for Fourier-based deconvolution, with the potential to assess sample impurities directly from highly congested LC-MS data. This streamlined workflow is capable of identifying multiple polymer architectures existing simultaneously within reaction mixtures and disambiguating salt adduction states. A free public version of the iFAMS software incorporating the presented MMD analysis workflow (iFAMS Quant v.6.4) is available for download at https://github.com/prellgroup/iFAMS/releases.

Bridge

iFAMS MMD analysis greatly expands the repertoire for model-free assignment of chemical modifications to polydisperse samples of a single type, with a range of applications in the development and quality control of biopharmaceuticals from membrane mimetics to polymer and glycoprotein conjugates. However, in highly multiplexed samples where mass heterogeneity is distributed over several independent ion species, the risk of non-specific assignments is drastically increased, requiring careful consideration of which harmonics can be included. In collaboration with the Laskin group at Purdue University, the following chapter demonstrates benefits of iFAMS characterization for isobaric species eluted from biological tissue and rules of

thumb for deselection of GT signal, in the development of a batch deconvolution workflow for GT analysis of protein distributions from the hundreds of thousands of spectra generated in mass spectrometry imaging.

CHAPTER IV

IFAMS IMAGER: A DECONVOLUTION WORKFLOW FOR FLEXIBLE ANALYSIS AND COMPARISON OF MASS SPECTROMETRY IMAGING PROTEIN DATA

While the material included here is primarily my own work, Lily Miller assisted with parallelization of the batch deconvolution process and helped generate the iFAMS Imager protein intensity maps. Manxi Yang and Hang Hu prepared the tissue samples and ran the nano-DESI-MSI experiments. James S. Prell and Julia Laskin contributed to experimental design and interpretation. This work will form the basis of a manuscript to be submitted in the future, with the above named as co-authors.

Introduction

Mass spectrometry imaging (MSI) has become an important tool for highly chemically specific localization of a plethora of molecular complexes in biological tissues, ranging from cellular lipids and proteins to smaller pathological compounds and metabolites,[111-115] with or without chemical labeling. Matrix-assisted laser desorption ionization (MALDI) is currently the most widely used soft ionization technique for MSI by far,[116, 117] achieving spatial resolutions as high as 1 μm while maintaining the primary structure of biomolecules.[118] However, MALDI-MSI typically requires application of a matrix layer to samples after pretreatment as well as direct exposure of the tissue to the vacuum environment. Desorption electrospray ionization (DESI) and the related nano-desorption electrospray ionization (nano-DESI), on the other hand, are ambient ionization methods which use buffered solution to elute intact biomolecules directly from the tissue surface with minimal sample pretreatment.[119-121]

Nano-DESI allows for multiple scans of the same tissue sample with spatial resolutions as high as 10 μm.[30, 122, 123] During the electrospray process, protein-sized molecules typically ionize into a stochastic distribution of consecutive charge states, often giving rise to a highly congested mass spectrum containing numerous overlapped analyte signals. Potential strategies to facilitate assignment of these peaks and relate their abundance to local analyte density include mass spectral deconvolution, use of internal protein standards, and signal normalization, especially if the deconvolution step can be automated to batch-process the tens to hundreds of

thousands of pixels worth of mass spectra generated in many nano-DESI imaging experiments. Cooper and Marty recently introduced a high throughput tool to address this protein quantitation bottleneck within the state-of-the-art Bayesian deconvolution software UniDec,[124] though this algorithm does not currently incorporate internal standards or easily allow the user to examine and potentially omit signals from analyte charge states that may be overlapped with interferent signals. Prior publications have demonstrated the use of Fourier- and Gábor-transform (FT/GT) based deconvolution, e,g., as implemented in iFAMS software, to facilitate interpretation of highly congested mass spectra and separate *m*/*z*-overlapped ion signals from different species by the characteristic spacings between isotopes, adducts, ligand-binding states and other regularly spaced peaks.[45] This software readily allows the user to adjust the charge states used for deconvolution, and the recently released iFAMS-based quantitation package (iFAMS Quant) supports use of internal standards, such as in LC-MS quantitation workflows.[53, 125] Here, we introduce an iFAMS-based workflow for MSI analysis ("iFAMS Imager") that utilizes GT to identify protein ion series, deselect non-specific nano-DESI-MS features (i.e., peaks that overlap with interferents both in *m*/*z* and frequency), and rapidly produce and normalize proteome intensity maps for flexible development of quantitative MSI protocols. Unique advantages of this software for nano-DESI imaging are illustrated and discussed.

Methods

Optical microscopy images and nano-DESI-MS data for Sprague-Dawley rat brain tissue were acquired by the Laskin group at Purdue University. (Sample sourcing and preparatory methods are further described in Appendix Chapter IV.) Bright-field optical images were collected on a PathScan Enabler IV (Meyer Instruments, Inc.). Dark-field immunofluorescence (IF) images were collected on a Nikon A1Rsi confocal microscope (Nikon, Tokyo, Japan) over an emission range of 663-738 nm, using an excitation wavelength of 640 nm. Nano-DESI mass spectra were collected on an Agilent 6560C ion mobility-quadrupole-time-of-flight instrument, raster scanned across 76 lines with a pitch of 150 µm at a line scanning velocity of 40 µm/s, sufficient for resolving soft tissue structures with a total of 522 scans per line (Figure 13A). Each raster line is stored as a single Agilent ".d" file, with the individual scans corresponding to a single pixel in each tissue image (data acquisition rate of 1 Hz). These pixel-specific mass spectra were extracted from their respective ".d" files using the MHDAC software development

kit provided by Agilent Technologies and converted to individual comma-separated-value files (CSVs) for iFAMS deconvolution. All spectra were analyzed with version 6.3 of the iFAMS deconvolution software ("iFAMS Quant"). All raw intensity maps are normalized to their 99.9% quantile maximum to avoid signal spikes dominating the image, with normalized intensity scales reflecting that normalization scheme (i.e., an intensity of 0.014 represents a signal that is 1.4% of the 99.9% quantile intensity, and the top 0.1% quantile is represented with a single color).

Because the signal-to-noise ratio (SNR) for individual charge states of intact protein ions can be low for individual pixels and vary dramatically between different locations in the tissue, the data processing workflow used here starts by analyzing the full summed spectra from one ".d" file (~500 pixels) of a central raster line (Figure 13A, B). The iFAMS GT deconvolution algorithm, described in depth elsewhere,[53, 125] is a Gaussian-windowed method of Fourier transform which localizes periodic peak spacings ($\Delta M/Z$) to narrow sections of the mass spectrum and is represented as a two-dimensional spectrogram of signal magnitude vs. m/z and frequency. Here, "frequency" fundamentals are the reciprocal of peak spacing ($Z/\Delta M$), which for isotopically resolved data ($\Delta M \approx 1$ Da) appear at integer frequencies equal to their charge state, with harmonic peaks occurring at integer multiples of the fundamental frequency (Figure 13C). Analytes, such as proteins, with constant mass at each charge state (after accounting for the small charge carrier masses) have charge state distributions that appear as a conspicuous series of peaks following a downward-sloped hyperbolic trend in the GT spectrogram. This characteristic arrangement of signals belonging to each analyte's charge state distributions often enables fully or mostly automated detection and selection of the signals belonging to individual proteins from the GT spectrogram. Analyte signals that remain overlapped with interferent signals even in the GT spectrogram can be manually deselected before the deconvolution step. GT data for selected peaks are subsequently extracted, normalized for charge, and added together to produce a "zero-charge" mass spectrum for mass characterization and quantitation (Figure 13D).

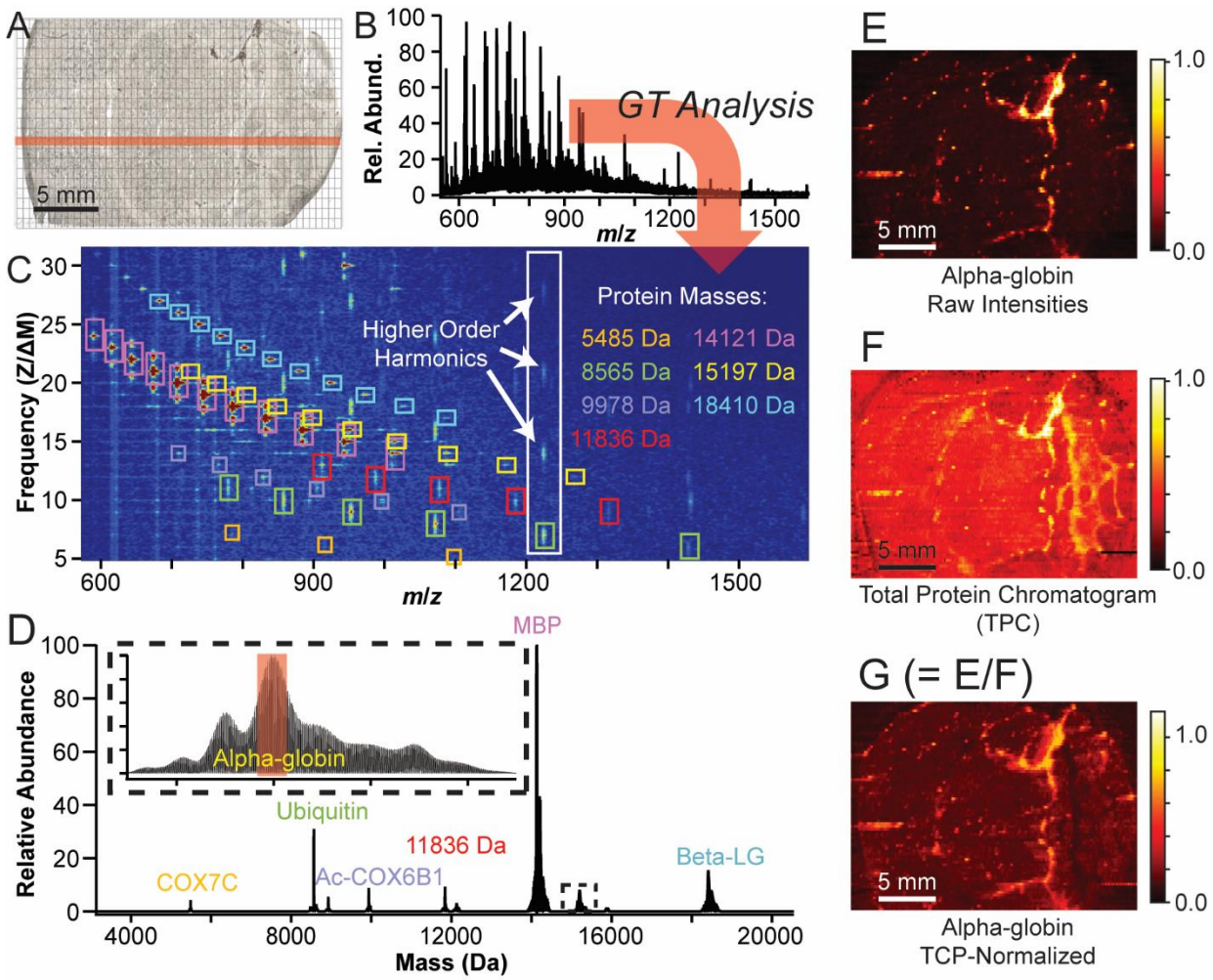


Figure 13. Workflow Schematic for analyzing MSI data with iFAMS Imager. All microscopy data and single-protein intensity maps are adapted with permission from Yang, M. et al. Signal suppression in imaging of intact proteins using nano-DESI MSI *manuscript in preparation*. (A) Light microscopy image of the rat brain tissue slice analyzed by nano-DESI, with 5 mm ruler for scale. (B) Mass spectrum of the summed intensities from a whole raster line of the tissue (A, redline) (C) Gábor transform (GT) spectrogram of the raster line mass spectrum (B). Fundamental frequencies (colored boxes) as well as any visible higher frequency harmonics (example harmonic series in white box) are selected for inverse GT. (D) Zero-charge spectrum of all protein series selected from the GT spectrogram (C), with colored labels for all identified proteins. The protein ion series of mass 11836 Da was not identified to a known proteoform but is clearly present in the GT spectrogram. Inset shows an example integration window (red shaded box) applied to the deconvolved distribution for Hemoglobin subunit alpha (alpha-globin). Protein intensity maps like E for alpha-globin are generated by batched deconvolution of pixel spectra with parameters identified for each protein from iFAMS analysis of the raster line spectrum (colored and shaded boxes in C and D), then summed into a total protein intensity map (F) for normalization of the color scale (G). All maps contain a 5 mm ruler for scale.

Initial iFAMS deconvolution of the raster line-integrated mass spectrum is used to identify proteins of interest and determine GT signal selection and peak integration parameters to be used in processing all of the individual MSI pixel spectra. The CSVs of individual pixel mass

spectra are first uploaded to the University of Oregon's high performance computing cluster ("Talapas") for batch processing using these batch parameters saved from the initial raster line analysis. Data processing on Talapas was parallelized across 20 nodes, each with 28 cores (see Appendix Chapter IV for a template SLURM command script), generating pixel-specific peak list files of ion mass and signal intensity for each protein. These peak list files are then exported and compiled into protein-specific arrays (Figure 13E) for intensity map cross-comparison via iFAMS Imager. Direct elution from tissue introduces a significant baseline of small metabolites, leading to matrix effects that can impact quantitation of larger complexes.[30] To mitigate the protein-specific biases in ionization efficiency, rather than normalizing to the total ion chromatogram signal for each pixel, a summed image from all iFAMS-deconvolved protein signals or "total protein chromatogram" image (TPC, Figure 13F) generated by iFAMS Imager is used for normalization of protein images (Figure 13G).

Results and Discussion

A major drawback to direct *m*/*z* analysis of congested mass spectra is the risk of ambiguous peak assignment. Use of GT helps disambiguate protein signal by separating overlapped *m*/*z* signal along the frequency axis, and inclusion of higher harmonics can further improve the resolution and signal stability of deconvolved mass distributions.[45] However, remaining overlapped signals in the GT spectrogram, if included in signal deconvolution, can result in analyte intensity maps "hybridized" with those of contaminant proteins. Figure 14A-C illustrates a striking example of this phenomenon resulting from the strong overlap of myelin basic protein (MBP^{18+}) and cytochrome c oxidase subunit 7c ($COX7C^{7+}$) signals at near-zero frequencies (Figure 14E). Because MBP is a much more abundant protein in the tissue (and in fact dominates the TPC image shown in Figure 14F), the corresponding contaminated COX7C image is very similar to the MBP image, despite COX7C and MBP being known to localize in complementary tissue types. Excluding this near-zero frequency data and instead using only fundamental and harmonic frequencies for both proteins results in a dramatically different intensity map for COX7C (Figure 14C) that much more closely matches the IF image (Figure 14D). Overlapped signals from different analytes can often be readily detected by visually examining the raster-line summed GT spectrogram. Such signals can then be easily deselected prior to generating protein intensity maps to avoid spurious image generation.

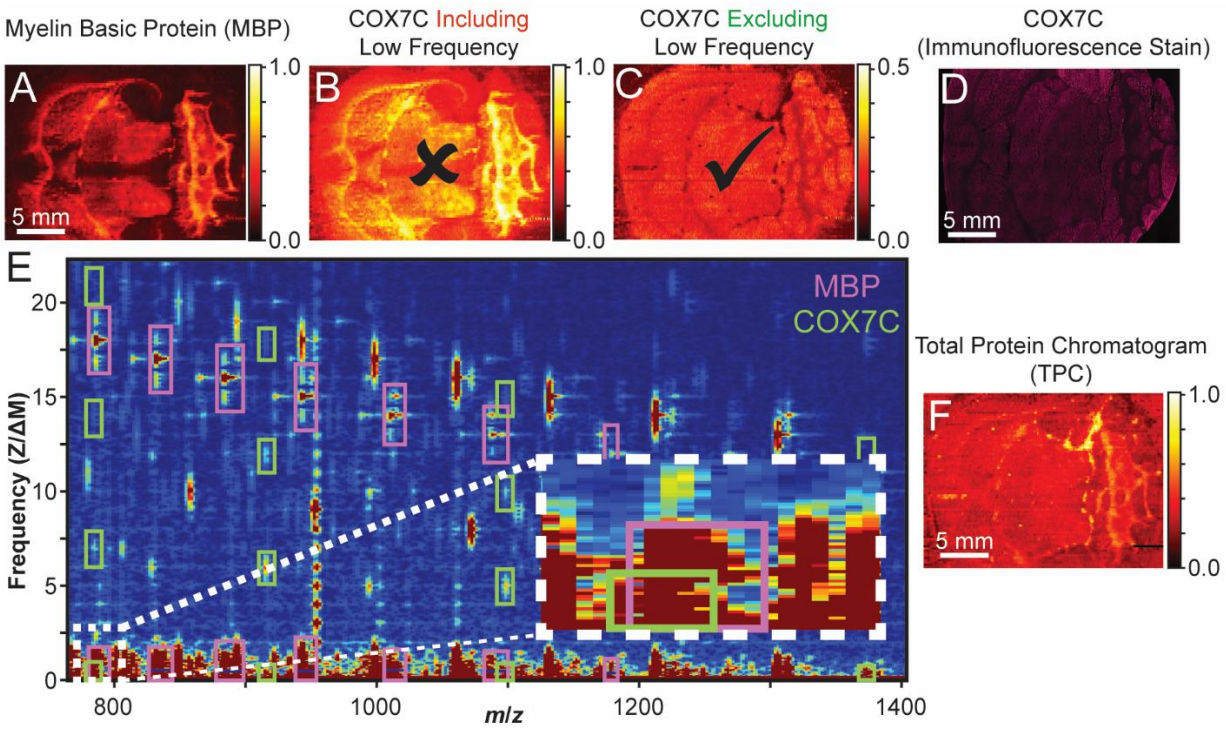


Figure 14. Intensity maps and Gábor spectrogram exhibiting overlapped protein signal. All microscopy data and single-protein intensity maps are adapted with permission from Yang, M. et al. Signal suppression in imaging of intact proteins using nano-DESI MSI *manuscript in preparation*. (A) TPC-normalized intensity map of MBP. (B) TPC-normalized intensity map containing both COX7C and MBP (not ideal). (C) TPC-normalized intensity map of COX7C alone (preferred) with immunofluorescence data for reference (D). (E) Gábor spectrogram of raster line scan of rat brain tissue, with overlapped signal of MBP (pink boxes) and COX7C (green boxes) at low frequencies (inset). (F) Total protein intensity map. The 5 mm rulers in images A, D and F apply to all intensity maps.

After deselection of overlapped protein signals in the GT spectrogram for raster line 39, 10 distinct protein charge series were identified and 8 were characterized, including hemoglobin subunit alpha (alpha-globin), MBP, COX7C, ubiquitin and 4 internal standard proteins added to the nano-DESI solvent (beta-lactoglobulin, myoglobin, bradykinin and lysozyme), which are discussed in more detail in a separate publication [Yang, M. et al. *manuscript in preparation*]. Because it has lower signal in the raster-line summed mass spectra than the much more dominant MBP and alpha-globin proteins, COX7C was chosen as an example for comparison to "ground truth" light microscopy imaging methods (Figure 15A, B) as well as "direct" intensity mapping based on mass spectral peak heights (Figure 15C). The iFAMS Imager-deconvolved intensity map for COX7C (Figure 15D) exhibits excellent agreement with the corresponding IF image, indicating this protein is localized rather uniformly in the grey matter and is present in much

lower quantities, if at all, in the vascular tissue and white matter. By contrast, the COX7C image generated directly from *m*/*z*-domain peak height mapping shown in Figure 15C exhibits much less uniform signal in the grey matter and has overall lower contrast for this protein between grey matter and other tissue types for this tissue slice. iFAMS Imager intensity maps for alpha-globin and MBP (Figures 3E, F), which are known to be strongly localized to the blood vessels and white matter, respectively, exhibit complementary tissue localizations to that of COX7C. These results highlight the utility of iFAMS Imager in identifying and excluding overlapped protein signals to generate more accurate protein-specific images.

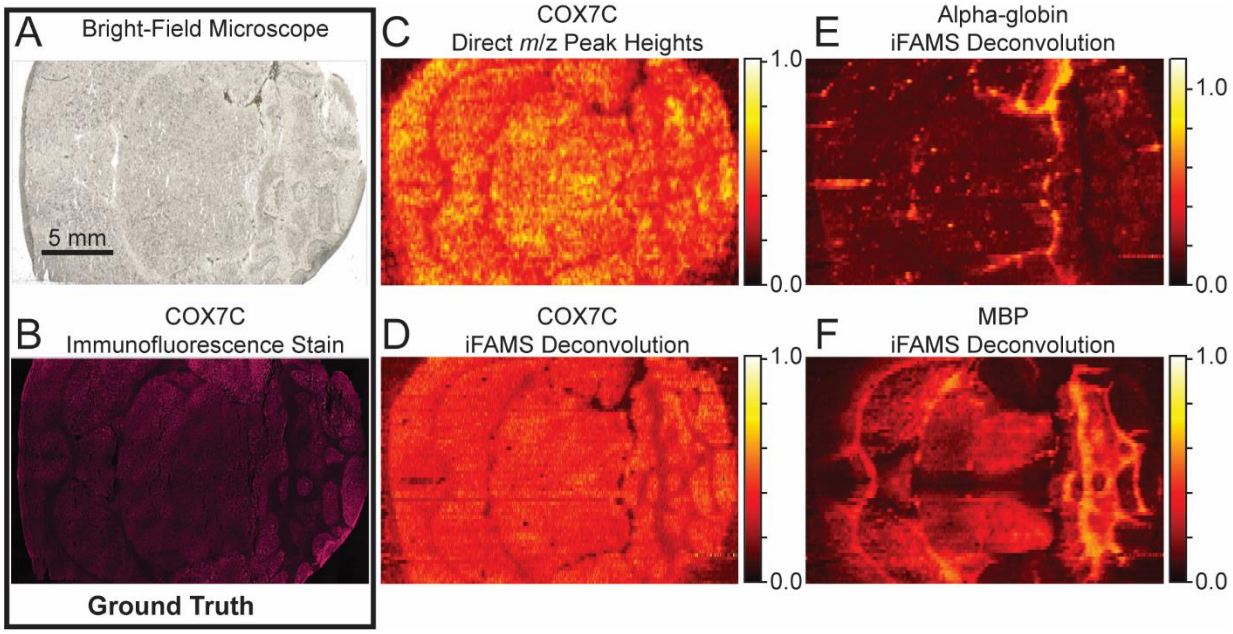


Figure 15. Comparison of iFAMS Imager and peak height protein intensity maps to "ground truth" light microscopy and immunofluorescence images. All microscopy data and single-protein intensity maps are adapted with permission from Yang, M. et al. Signal suppression in imaging of intact proteins using nano-DESI MSI *manuscript in preparation*. (A) Bright field optical image of rat brain tissue sample. (B) Immunofluorescence (IF) dark field image of a comparable tissue cross-section to that of the bright-field image, stained with COX7C rabbit Polyclonal primary antibody. Direct peak intensity image for COX7C (C) and the TPC-normalized intensity maps for COX7C (D), alpha-globin (E), and MBP (F) were all generated from the same nano-DESI data from the bright field-imaged tissue sample A. The black 5 mm ruler in the light microscopy image (A) applies to all subplots.

For both direct *m*/*z* peak assignment and GT-aided protein characterization from nano-DESI experiments, it was necessary to sum the mass spectra across a raster line to achieve sufficiently high SNR for analysis, a situation that is likely to be common in nano-DESI protein MSI in general. However, accumulating a representative sample of the tissue proteome in this

manner increases the chances of charge state series overlap from different analyte proteins and restricts the range of *m*/*z* peaks available for direct unambiguous protein quantitation from nano-DESI mass spectra. By separating the ion signals across both *m*/*z* and frequency, iFAMS provides the signal isolation needed to integrate broader ranges of protein charge states and mass variants in protein intensity maps for quantitation and localization by MSI. Figure 15 compares images generated with iFAMS Imager using different peak integration ranges for alpha-globin^{9-21+}, ranging from a single isotopolog (15197.57-15198.05 Da, Figure 16A, D) to a single proteoform (15182.25-15212.07 Da, Figure 16B, E) and a summation of multiple proteoforms (15134.22-15257.78 Da, Figure 16C, F). Brief loss of ESI current during collection of line 67 corresponds to the bright row of pixels in the TCP-normalized images (see Figure S11). Aside from minor reduction in the modulation of background noise with broader integration windows, the intensity maps for each integration window are very similarly localized to blood vessels in the rat brain tissue, indicating the deconvolved signal is exclusively from alpha-globin-related proteoforms for all three integration ranges. Robustness of these images down to a single isotopolog is attributed to the inclusion of the full range of observed charge states, as single charge state images have significantly lower SNR compares to those summed across multiple charge states (see Figure S12). With isotope-resolved data, multiple isotope peaks can be selected in iFAMS parametrically by user-defined minimum abundance thresholds and noise tolerance and compiled into a single peak list array, enabling flexible isotope or proteoform-specific intensity mapping in iFAMS Imager.

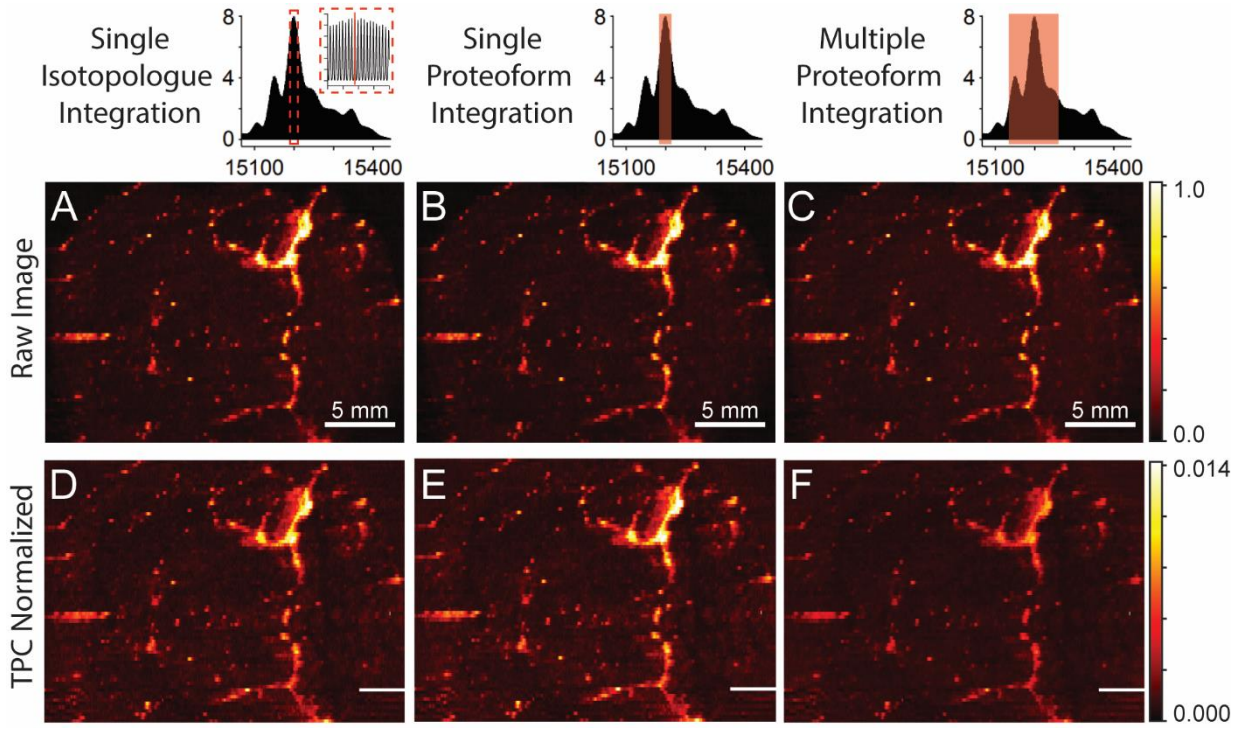


Figure 16. Protein intensity maps comparing the effects of narrow (A, D), medium (B, E), and wide (C, F) integration windows for deconvolved alpha-globin. 5 mm rulers in images A, B, and C apply to all images. Color bars on the far right correspond to all protein intensity maps within their respective row, raw protein intensity maps (A, B, C) and total protein chromatogram images (D, E, F). White lines in images D, E, and F are pixels where nano-DESI spray failed (see Figure S11) and define the normalized image maxima at 1.0 on the color bar.

Whereas experimental data collection for nano-DESI imaging often requires a few hours, protein characterization directly from data-rich biological tissue spectra can be a much more time-consuming process, taking on the order of days to weeks, depending on spatial resolution and the number of analytes to be separately imaged. While iFAMS deconvolution takes around 10 seconds to process each image pixel (hundreds of hours of computational time per image), this processing time is compensated for in several key ways. Firstly, parallelizing the deconvolution tasks using a Message Passing Interface script can drastically reduce the processing time, in this case down to 12 minutes for ~40,000 pixels processed across 20 nodes with 28 cores per node. Although the initial data upload to the computing cluster takes 1-2 hours from the MHDAC file conversion to file transfer completion, this step only needs to be done once per full raw data set, with batch parameter uploads and peak list downloads taking only 5 minutes per protein. Most critically, iFAMS raster-line analysis reduces the multi-protein characterization step down to 10-30 minutes of deconvolution and integration per tissue sample,

including visual confirmation that signal selections are protein-specific. While several computational tools exist to batch process MSI spectra for known compounds in a matter of minutes,[124, 126, 127] they may have limited flexibility for analyzing new compounds or isolating signals overlapped in *m*/*z*. These tools thus represent useful advantages of iFAMS Imager in handling interferent signals in highly congested mass spectra.

Conclusions

Overlapped ion signals and broad proteoform profiles often pose challenges for *m*/*z*-based deconvolution algorithms, making accurate quantitation of individual analyte signals a major hurdle for MSI of biological tissues. Here, we have illustrated how iFAMS Imager, a Gábor Transform-based MSI deconvolution workflow, provides excellent protein signal and selectivity for rapid identification and rejection of interferent signals that can otherwise dominate images for low-abundance analytes. iFAMS Imager provides additional back-end tools needed for TPC generation and intensity map normalization, with flexible deconvolved peak integration to generate images ranging from single isotopologs to combinations of proteoforms. Workflow implementation is aided greatly by multi-core processing using Message Passing Interface-based parallelization, allowing generation of protein intensity maps from complex tissue spectra that can closely agree with ground-truth immunofluorescence images within a matter of hours after MSI data collection. Further improvements may be possible when paired with sparse data sampling methods.[128] The version of iFAMS used for the initial raster line spectrum analysis (iFAMS Quant) is publicly available for download at https://github.com/prellgroup/iFAMS/releases. A modified version of iFAMS Quant for implementation on a SLURM-based computing cluster, along with iFAMS Imager and other workflow support scripts are publicly available for download at https://github.com/prellgroup/iFAMS_Imager/releases.

CHAPTER V

OUTLOOK

Soft ionization mass spectrometry offers many advantages for the study of intact pharmaceuticals and biological complexes, from its low sample requirements to its high flexibility in coupling solution phase and gas phase separation techniques with extreme ion sensitivity and high mass specificity. Developments in top-down "omics" and membrane protein analysis, along with innovations in instrumentation, have tested the limits of what information can be directly interpreted from mass spectral data in the *m*/*z* domain. Generalized deconvolution algorithms capable of handling both high- and low-resolution data can in many cases aid sample characterization far beyond what is manually assignable, but those developments have lagged behind the most challenging problems facing the field. The work described in this dissertation advances previously under-explored aspects of highly versatile Fourier-based deconvolution for mass spectrometry while also hinting at new areas of development for FT/GT analysis.

Fourier phase information, as demonstrated in Chapter II, enables frequency-based analysis of aperiodic mass heterogeneities that is complementary to state-of-the-art mass-domain modeling methods, confirming previously observed stochastic incorporation of melittin in lipid bilayers. Analysis of low-resolution Q-TOF data at low signal-to-noise ratios expands the applicability of macromolecular mass defect analysis to more readily available ESI-MS instrumentation. Although iFAMS MMD analysis does not require high-resolution data from FT-ICR and Orbitrap-type instruments, in principle it could also utilize FT-ICR-resolved isotope frequencies for characterization of elemental compositions from isotope fine structure. The iFAMS MMD workflow presented in Chapter III combines Fourier analysis with highly tunable LC-MS separation for multi-component analysis of functionalized polymers, enabling characterization of the repeat unit mass, end-groups, branching structure, and average polymer size from in-progress reactions. Characterizing the impurities and degradation products observed by LC often requires a prolonged and multimodal assessment, especially for the broad mass distributions that arise from the use of heavier polymer linkers. Fourier-based deconvolution simplifies polymer impurity assessment before, during, and after synthesis and could be coupled with other separation techniques such as ion mobility MS to improve characterization of large

protein complex conformers with similar collisional cross sections, differentiating them by mass variants.

Chapter IV continues the theme of overcoming sample heterogeneity, focused on extending the ion selectivity of Fourier-based deconvolution to more multiplexed high-throughput applications. iFAMS deconvolution allows deselection of isobaric interferents from GT-resolved isotope frequency harmonics, enabling summation over a larger number of scans without significant loss of signal specificity. The increased non-isobaric ion series coverage from the Gábor spectrogram results in more accurate protein intensity maps, which, when coupled with flexible option for signal normalization, moves the field toward more standardized and reliable protein quantification.

Mass polydispersity is a universal problem in characterizing dynamic and "dirty" real-world samples such as reaction mixtures and biological tissues, from oligomeric stoichiometries to isotopic variants. As the biological samples studied by ESI-MS grow more heterogeneous, subtle differences in sequence, conformation, stoichiometry, and small molecule binding are revealed to be essential for understanding biological function, with the number of variations discovered far outpacing protein-specific labeling methods. Coupling deconvolution algorithms to new instrumentation and separation and localization techniques continues to inform algorithm development, with Fourier-based analysis bridging many of the gaps left by the resolution restrictions of mass-domain analyses.

APPENDIX

SUPPORTING INFORMATION FOR CHAPTERS II-IV

CHAPTER II

Mathematical Proof

Elimination of Global (Envelope) Phase for Analysis of Local (Peak) Phase

Throughout the following proof, we denote a function $f(y)$ that depends parametrically on quantities $a, b, ...$, as $f(y; a, b, ...)$. We also denote the mass-to-charge ratio *m/z* by x for typographical clarity, and k is defined as the frequency variable conjugate to x in the frequency domain. M_L denotes the mass of the repeated subunit/ligand in the ion population that gives rise to periodic signals in the frequency domain.

Let T_α denote a translation operator that translates a function to the right by α, e.g., $\mathrm{T}_\alpha\{f(y)\} = f(y-\alpha)$. Let $c(x; \frac{M_L}{Z})$ denote a comb function of x with peak spacing $\frac{M_L}{Z}$ (with a peak at $x = 0$). Let $e(x; x_{max}(Z), Z)$ denote a slowly varying envelope function of x with maximum, $x_{max}(Z)$. Finally, let $p(x; Z)$ be a peak shape function, defined between $x = 0$ and $x = \frac{M_L}{Z}$. (It is trivial to account later for additional mass due to a charge carrier, such as a proton, in defining $p(x; Z)$; see below.)

The type of mass spectrum under consideration here is any such one that can be decomposed into the sum of charge-state-specific mass spectra $s(x; Z) = \left[c\left(x; \frac{M_L}{Z}\right) \times e(x; x_{max}(Z), Z)\right] * p(x; Z)$, with $Z \in \{Z_{min}, Z_{min}+1, \dots, Z_{max}\}$ the range of observed charge states. It is assumed that $e(x; x_{max}(Z), Z)$ and $x_{max}(Z)$ may be different for each Z. Without loss of generality, then, $e(x; x_{max}(Z), Z)$ can be defined such that $x_{max}(Z)$ coincides with a peak of $c\left(x; \frac{M_L}{Z}\right)$, i.e., $x_{max}(Z)$ is an integer multiple of $\frac{M_L}{Z}$. Emphatically, reconstructing the Macromolecular Mass Defect (MMD) profile for a particular charge state Z is equivalent to recovering $p(x; Z)$ from observation of $s(x; Z)$. (Optionally, if we furthermore assume that the mass distribution represented by $p(x; Z)$ is identical for all observed charge states, this amounts to recovering a

single peak shape function $p(x)$ common to all Z by summing over or otherwise averaging over the individual $p(x;Z)$ after normalizing the x-axis for charge state.)

To determine $p(x;Z)$ from a charge-state-specific mass spectrum, we first note that

$$s(x;Z) = \left[c(x;\frac{M_L}{Z}) \times \mathrm{T}_{x_{max}(Z)}\{e(x;0,Z)\}\right] * p(x;Z)$$

By the Fourier Convolution Theorem,

$$S(k;Z) = [C(k;\frac{Z}{M_L}) * e^{-2\pi i k x_{max}(Z)} E(k;0,Z)] \times P(k;Z)$$

and we note that $kx_{max}(Z)$ is an integer at each peak in the comb $C(k;\frac{Z}{M_L})$ by choice of $x_{max}(Z)$.

For sufficiently broad $e(x;0,Z)$, $E(k;0,Z)$ is narrow and decays to zero within $\frac{1}{M_L}$ either side of $k = 0$. In this case, we shall show that, by sampling this function at integer multiples of $\frac{Z}{M_L}$ (i.e., where the magnitude of $S(k;Z)$ achieves local maxima), one can reconstruct $p(x;Z)$. This can be difficult to do with the function in the form shown above, however, because $e^{-ikx_{max}(Z)}$ generally contributes rapid phase oscillations to $S(k;Z)$ that make it practically very difficult to determine phase at specific values of k for realistic data.

To improve the situation, we first sample $S(k;Z)$ by multiplying $S(k;Z)$ by another copy of $C(k;\frac{Z}{M_L})$:

$$S(k;Z) \times C\left(k;\frac{Z}{M_L}\right) = C(k;\frac{Z}{M_L}) \times [C(k;\frac{Z}{M_L}) * e^{-ikx_{max}(Z)} E(k;0,Z)] \times P(k;Z)$$

On the right hand side of this equation, multiplication by the comb has the effect of sampling $[C(k;\frac{Z}{M_L}) * e^{-ikx_{max}(Z)} E(k;0,Z)]$ at values of k that are integer multiples of $\frac{Z}{M_L}$, where $\left[C\left(k;\frac{Z}{M_L}\right) * e^{-ikx_{max}(Z)} E(k;0,Z)\right] = E(0;0,Z) \times e^{-ikx_{max}(Z)}$.

Thus, at the sampled values of k, $S(k;Z) = E(0;0,Z) \times e^{-ikx_{max}(Z)} \times P(k;Z)$, or, equivalently,

$$S(k;Z) \times \frac{e^{ikx_{max}(Z)}}{E(0;0,Z)} = P(k;Z)$$

Note that this function has stationary (i.e., very slowly varying) phase due to the nearly even symmetry of $e(x;0,Z)$. That is, the phase of this function at the sampled values of k depends only on $P(k;Z)$, because we have eliminated the (generally, rapidly oscillating) contributions from the envelope function, $E(k;x_{max}(Z),Z)$. This is the key strategy that enables practical MMD profile reconstruction using the FT-based method, for the phase at peaks in the frequency domain is now relatively easy to see visually and measure accurately. In other words, sampling the Fourier Transform of the mass spectral signal, modulated by the scaled exponential on the left-hand side of this equation, is equivalent to sampling $P(k;Z)$.

Now, $E(0;0,Z)$ is equal to the integral over $e(x;x_{max}(Z))$, which we denote $\bar{e}(Z)$, thus it is proportional to the total ion population with charge state Z, divided by Z. Because the spacing of the frequency-domain samples is $\frac{Z}{M_L}$, the inverse Fourier Transform of the sampled data set spans the interval $x \in [0, \frac{M_L}{Z})$, the same as the interval on which $p(x;Z)$ is defined.

Taking the inverse Fourier Transform of both sides, we have, on the interval $x \in [0, \frac{M_L}{Z})$:

$$\mathrm{T}_{-x_{max}(Z)}\{s(x;Z)\}/\bar{e}(Z) = p(x;Z)$$

defined at a set of equally spaced points equal to the number of samples (i.e., harmonics used) in the frequency domain.

Equivalently,

$$\mathrm{T}_{-x_{max}(Z)}\{s(x;Z)\} = \bar{e}(Z) \times p(x;Z)$$

This is a reconstruction of $p(x;Z)$ weighted by the total ion population for charge state Z divided by Z, i.e., it is the MMD profile for charge state Z, the main target for reconstruction.

To combine $p(x;Z)$ from all observed values of Z into a single charge-state-averaged MMD profile $p(x)$, one can simply translate each $p(x;Z)$ cyclically to the left by (mass of the charge carrier/1), normalize for charge by replacing x by Zx, and interpolate and sum the resulting "zero-charge" MMD profiles over the set of all observed Z. Note that the axis transformation

$x \mapsto Zx$ naturally accounts for the $1/Z$ scaling of $e(x; x_{max}(Z), Z)$ such that each charge state's contribution to the summed MMD profile ($p(x)$) is simply proportional to its total ion population.

In summary, the FT-based MMD reconstruction presented here is based on the following idea: we take advantage of the Stationary Phase Approximation to confidently determine the phase contribution of $P(k; Z)$, which is what is needed to reconstruct $p(x; Z)$.

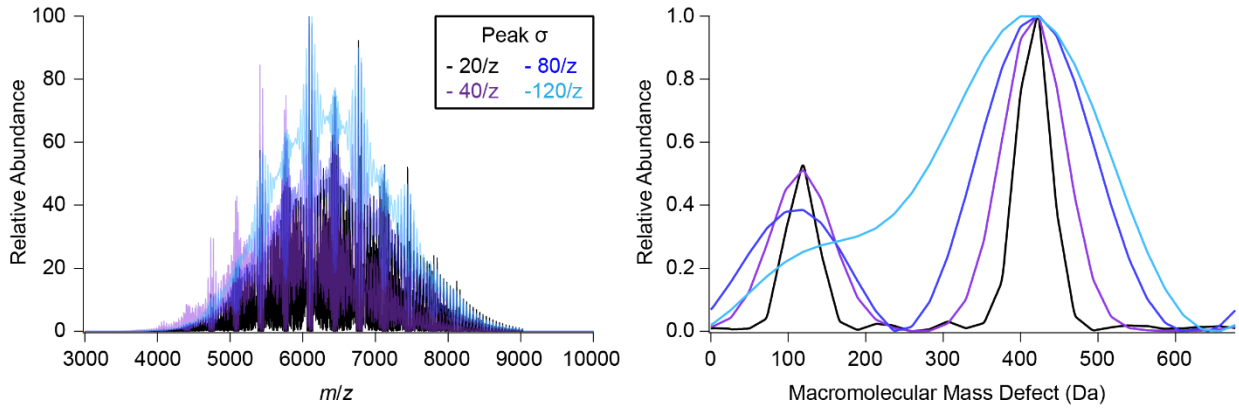


Figure S1 Comparison of simulated spectra and MMD profiles as peak width is varied. Wider peaks in the mass spectra lead to greater superposition, visually appearing as a higher curved baseline. The reduced resolution of the simulated mass spectra is mirrored in their MMD profiles, and inconsistent peak height ratios justifies our integral treatment of relative mass defect abundance (see Table S3)

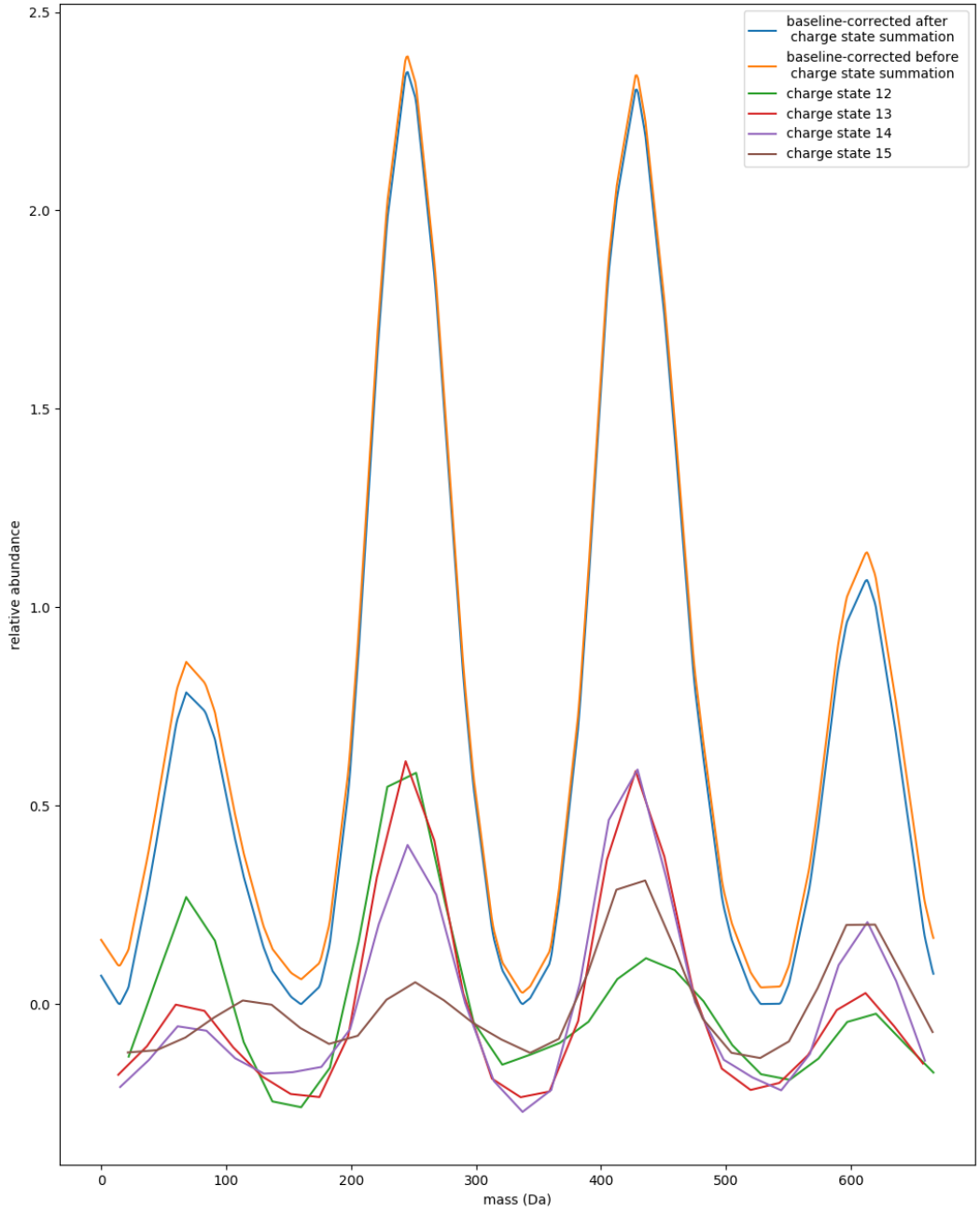


Figure S2 Charge-state-specific (green, red, purple and brown traces) and total MMD profiles (blue and orange traces) for melittin-embedded Nanodiscs prepared at a bulk melittin:Nanodisc ratio of 6:1. Baseline correction was applied to either the individual charge state profiles (orange) or their interpolated sum (blue) with negligible difference, so the MMD profile baseline correction implemented in iFAMS 6.1 is arbitrarily applied after summation of the charge-state-specific profiles.

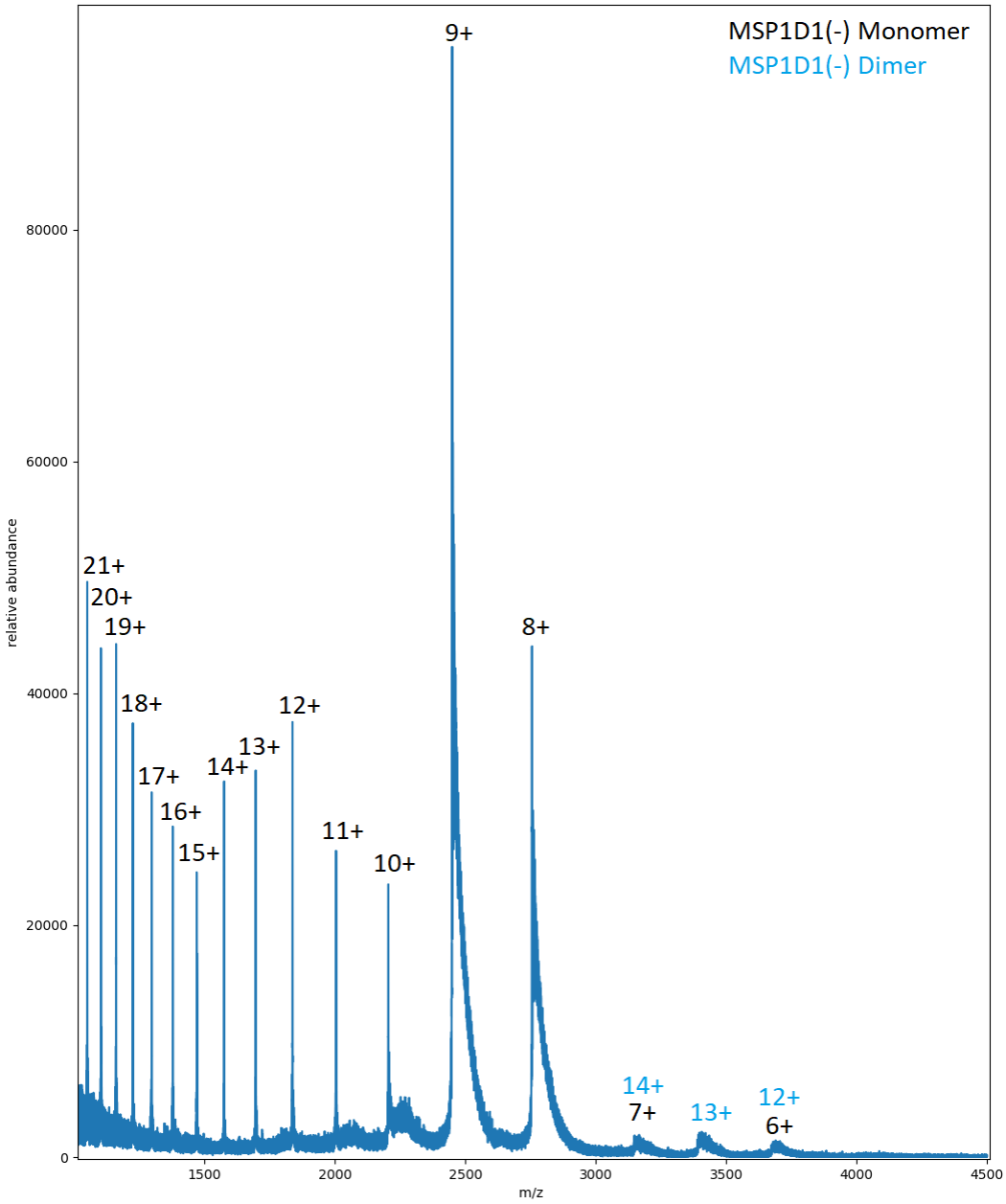


Figure S3 Example native mass spectrum of membrane scaffold protein MSP1D1(−), taken with a trap voltage of 5 V and trap gas flow rate of 3 mL/min, showing both native monomer (black charge states) and dimer (blue charge states) signals. The presence of higher charge states (~11-21+) indicates some partially unfolded monomer population.

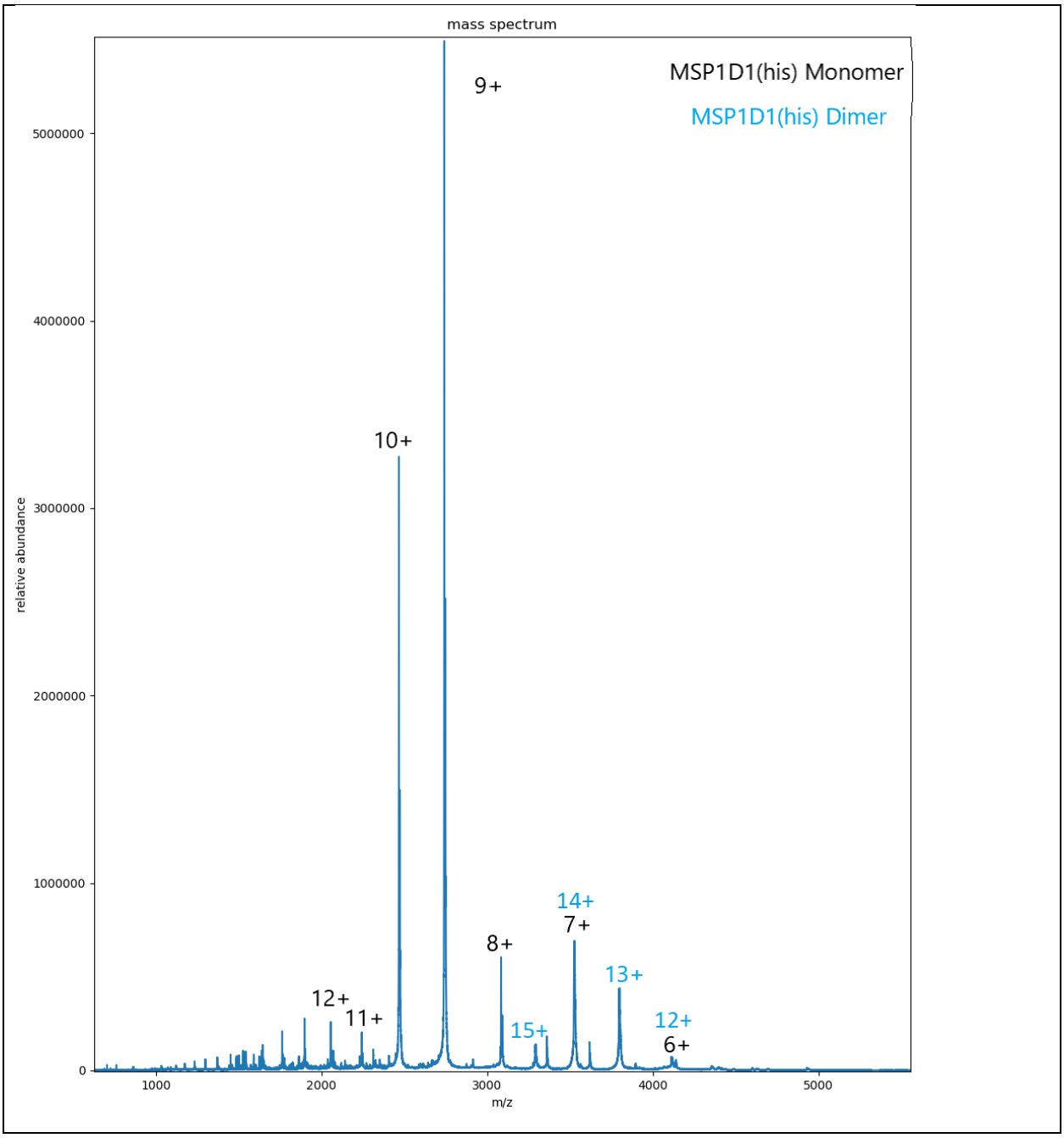


Figure S4 Example native mass spectrum of membrane scaffold protein MSP1D1His, taken with a trap voltage of 25 V and trap gas flow rate of 5 mL/min, showing both native monomer (black charge states) and dimer (blue charge states) signals. The presence of higher charge states (~11+ and above) indicates some partially unfolded monomer population.

UniDec Deconvolution Parameters for Melittin-Embedded Nanodiscs

Data Processing			**Peak selection/ extraction/plotting**		
m/z range	1979.925	1000000	Picking range	50000	
Background subtraction	0		Picking threshold	0.1	
Bin every	0		Peak normalization type	Max	
Normalize data?	Yes		Peak extraction method	Height	
UniDec Params			Extraction window	0	
Charge range	10	18	Extraction Threshold	10	
Mass range	100000	200000	Extraction normalization type	Max	
Sample mass every	10		**Add'tl Plotting Params**		
Quick Controls			2D color map type	nipy_spectral	
Smooth charge state distribution?	Yes		Peaks color map type	rainbow	
Auto *m/z* peak width?	Yes		Spectra color map type	rainbow	
Level of Nearby-Point Smoothing	Some		Discrete plot?	No	
Level of artifact suppression	None		Publication mode? (Richardson-Lucy peak sharpening)	No	
Mass defect	666.9		Reconvolve w/peakshape or raw	Raw	
Add'tl Deconv Params			Marker threshold	0.1	
Peak FWHM	10		Species separation	0.025	
Peak shape function	Gaussian		Integration range	-1000	1000
Beta	0		Limits on # of spectra	100	20
Charge smooth width	1		**Mass Defect Analysis**		
Point smooth width	1		Number of Defect Bins	50	
Mass smooth width	1		Subunit mass	666.9	
Native charge offset range	-100	100	X-axis	mass(Da)	
Isotopes (isotopic resolution only)	Off				
Manual mode	No				
Negative mode	No				
Charge scaling	No				
Mass list window	N/A				
m/z to mass transformation type	Interpolate				
Maximum number of iterations	100				
Adduct mass	1.007276				

Table S1 List of all UniDec parameters used for spectral deconvolution of melittin-embedded Nanodisc data, in order as they appear in MetaUniDec Version 4.2.1. Richardson-Lucy peak sharpening reconstructions to recreate published profiles consisted of selecting publication mode, changing “Nearby Point Smoothing” to “None” and setting “Point Smooth Width” to zero.

UniDec Deconvolution Parameters for Mixed Scaffold Protein Nanodiscs

Data Processing			**Peak selection/ extraction/plotting**		
m/z range	Full spectrum		Picking range	500	
Background subtraction	0		Picking threshold	0.1	
Bin every	0		Peak normalization type	Max	
Normalize data?	Yes		Peak extraction method	Height	
UniDec Params			Extraction window	0	
Charge range	9,10,11	15,17,18	Extraction Threshold	10	
Mass range	8000	300000	Extraction normalization type	Max	
Sample mass every	10		**Add'tl Plotting Params**		
Quick Controls			2D color map type	nipy_spectral	
Smooth charge state distribution?	Yes		Peaks color map type	rainbow	
Auto *m/z* peak width?	Yes		Spectra color map type	rainbow	
Level of Nearby-Point Smoothing	Other		Discrete plot?	No	
Level of artifact suppression	None		Publication mode? (Richardson-Lucy peak sharpening)	No	
Mass defect	677.9		Reconvolve w/peakshape or raw	Raw	
Add'tl Deconv Params			Marker threshold	0.1	
Peak FWHM	0.85		Species separation	0.025	
Peak shape function	Gaussian		Integration range	-1000	1000
Beta	0		Limits on # of spectra	100	20
Charge smooth width	1		**Mass Defect Analysis**		
Point smooth width	100		Number of Defect Bins	50	
Mass smooth width	1		Subunit mass	677.9	
Native charge offset range	-100	100	X-axis	mass(Da)	
Isotopes (isotopic resolution only)	Off				
Manual mode	No				
Negative mode	No				
Charge scaling	No				
Mass list window	N/A				
m/z to mass transformation type	Smart				
Maximum number of iterations	100				
Adduct mass	1.007276				

Table S2 List of all UniDec parameters used for spectral deconvolution of mixed scaffold protein Nanodisc data, in order as they appear in UniDec Version 4.2.1.

Mass Defect Accuracy by Integration for Simulated Melittin-Incorporated DMPG Nanodisc Data with Varying Degrees of Peak Resolution

Single-Peak Assumption	Relative Abundance	1 / 1	0.5 / 1	0.25 / 1	0.125 / 1
Peak Separation					
0σ		0.00 / 0.00	0.00 / 0.00	0.00 / 0.00	0.00 / 0.00
1σ		-50.00 / 50.00	-66.67 / 33.33	-80.00 / 20.00	-88.89 / 11.11
2σ		-100.00/100.00	-133.33 / 66.67	-160.00 / 40.00	-177.78 / 22.22
3σ		6.06 / -5.66	20.89 / 3.89	48.25 / 11.63	-266.67 / 33.33
4σ		1.81 / -1.59	4.87 / 0.27	9.29 / 1.25	15.65 / 1.68
5σ		0.44 / -0.36	1.17 / 0.06	2.13 / 0.22	3.68 / 0.32
6σ		0.09 / -0.06	0.21 / 0.00	0.41 / 0.04	0.70 / 0.05
Multi-Peak Assumption	Relative Abundance	1 / 1	0.5 / 1	0.25 / 1	0.125 / 1
Peak Separation					
0σ		80.11 / -79.47	80.11 / -79.47	80.11 / -79.47	80.11 / -79.47
1σ		39.88 / -39.24	33.90 / -43.91	27.68 / -46.94	22.53 / -48.70
2σ		16.95 / -16.38	8.22 / -21.95	-3.63 / -25.11	-16.66 / -26.80
3σ		6.06 / -5.66	20.89 / 3.89	48.25 / 11.63	-29.35 / -12.69
4σ		1.81 / -1.59	4.87 / 0.27	9.29 / 1.25	15.65 / 1.68
5σ		0.44 / -0.36	1.17 / 0.06	2.13 / 0.22	3.68 / 0.32
6σ		0.09 / -0.06	0.21 / 0.00	0.41 / 0.04	0.70 / 0.05

Table S3 Deviations of peak centroid calculations of simulated data in the assessment of the automated peak selection method used for calculation of mass defect peak areas (approximating peak boundaries by dropping a line to zero at local minima), means and full-width-half-max (FWHM). The simulated data consisted of two gaussian peaks with σ = 100, with deviations for both peaks represented in each cell, separated by " / ". The single and multi-peak assumptions respectively reflect whether the simulated data was assumed to contain a single peak or multiple peaks.

Melittin-Incorporated DMPG Nanodisc Data						
	# of Melittin	**Mass Defect (Expected)**	**Mass Defect (Peak Integration)**	**Standard Deviation**	**FWHM**	**RMSD**
UniDec						
0	0	72.4253	48.1208	33.4451	78.7571	**24.3045**
1.5	0	72.4253	62.9280	39.0801	92.0266	**35.8189**
	1	251.2879	289.6118	58.3700	137.4509	
	2	430.1505	478.0050	39.9752	94.1343	
3	0	72.4253	60.4752	27.3609	64.4301	**20.0798**
	1	251.2879	255.3657	33.6639	79.2724	
	2	430.1505	462.5566	61.9354	145.8468	
6	0	72.4253	79.1632	40.4490	95.2501	**11.5677**
	1	251.2879	234.0467	25.7910	60.7331	
	2	430.1505	418.3673	26.3267	61.9945	
	3	609.0131	601.6822	26.5076	62.4206	
9	4	120.9090	131.5617	28.7969	67.8115	**21.9732**
	5	299.7716	312.1929	30.2316	71.1899	
	6	478.6341	438.2619	34.9606	82.3258	
	3	609.0131	614.8098	29.0172	68.3304	
12	4	120.9090	141.2823	36.6193	86.2319	**16.8861**
	5	299.7716	312.1315	37.9574	89.3828	
	6	478.6341	467.4528	39.8328	93.7991	
	7	657.4967	636.3380	35.7912	84.2818	
24	4	120.9090	166.9398	38.3100	90.2131	**23.9999**
	5	299.7716	305.9560	24.8769	58.5807	
	6	478.6341	485.1766	24.0776	56.6984	
	7	657.4967	0.6997	21.9070	51.5870	
iFAMS						
0	0	72.4253	64.9467	42.9032	101.0292	**7.4786**
1.5	0	72.4253	66.7901	48.1732	113.4392	**19.8326**
	1	251.2879	281.1103	28.0318	66.0098	
	2	430.1505	414.0611	22.2513	52.3978	
3	0	72.4253	78.6316	30.3582	71.4882	**17.0410**
	1	251.2879	265.3121	29.5446	69.5723	
	2	430.1505	455.3694	28.9354	68.1377	
6	0	72.4253	78.4087	26.1854	61.6618	**3.5184**
	1	251.2879	247.6519	28.4922	67.0940	
	2	430.1505	430.3194	30.9833	72.9601	
	3	609.0131	609.6964	24.9687	58.7967	
9	4	120.9090	148.7955	29.0347	68.3716	**21.5303**
	5	299.7716	317.0059	28.5935	67.3325	
	6	478.6341	483.3089	25.1844	59.3048	
	3	609.0131	636.5390	30.7236	72.3486	
12	4	120.9090	157.6944	29.5105	69.4920	**22.4782**
	5	299.7716	322.9966	29.8685	70.3350	
	6	478.6341	489.9335	31.3901	73.9181	
	7	657.4967	656.5886	30.6896	72.2685	
24	4	120.9090	161.5878	28.4118	66.9046	**27.5792**
	5	299.7716	323.0740	30.4655	71.7408	
	6	478.6341	500.3669	29.5352	69.5501	

	7	657.4967	9.7932	27.5030	64.7646	
Mixed-MSP DMPC Nanodisc Data						
	Mass Defect (Expected)	**Mass Defect (Peak Integration)**		**Standard Deviation**	**FWHM**	**RMSD**
UniDec						
Histidine-tagged only	514.4000	617.3985		121.7994	286.8157	102.9985
His-tag cleaved only	24.3000	668.0286		94.19822	221.8199	34.17143
Mixture Ambient	281.5000	272.2444		40.82698	96.1402	38.61799
	514.4000	456.4018		40.52664	95.43293	
	608.3000	670.1908		66.94955	157.6541	
Mixture 90°C for 10 minutes	281.5000	278.2126		41.57944	97.91209	24.16901
	608.3000	642.3217		96.37299	226.9411	
Mixture 80°C for 60 minutes	514.4000	445.7393		41.72021	98.24359	62.39332
	608.3000	663.7217		83.5257	196.688	
iFAMS						
Histidine-tagged only	514.4000	611.1677		124.0128	292.0279	96.76774
His-tag cleaved only	24.3000	13.20665		110.7438	260.7817	11.09335
Mixture Ambient	281.5000	302.1674		41.05884	96.68617	16.36484
	514.4000	496.8346		36.28081	85.43478	
	24.3000	32.53034		62.16065	146.3771	
Mixture 90°C for 10 minutes	281.5000	249.8713		4.766691	11.2247	35.5234
	608.3000	647.3313		111.9027	263.5106	
Mixture 80°C for 60 minutes	514.4000	461.7888		44.15045	103.9664	59.98376
	608.3000	674.8444		81.62567	192.2138	

Table S4 Statistical information for assessment of MMD analysis mass defect accuracy as expected from empirically determined masses. Peak integration for calculating peak centroids and standard deviations was performed with the assumption each peak begins and ends abruptly at local minima in the mass defect profile. Data is presented only for peaks selected by an automated peak selection method and which are shared between the two MMD analysis methods.

CHAPTER III

Extended Methods

Sources for Chromatography Solvents and Polymer Standards

LC-MS grade formic acid, HPLC grade triethylamine (TEA), and HR-GC grade Dichloromethane (DCM) were purchased from Millipore (Burlington, MA, USA); HPLC grade acetonitrile (ACN) was purchased from J.T. Baker (Phillipsburg, NJ, USA).

Multi-arm 6 kDa pentaerythritol-core PEG standards (hydroxyl-functionalized, mesylate-functionalized, amine-functionalized, and maleimide-functionalized) were obtained from Biovectra (Charlottetown, Canada), while the heavier pentaerythritol-core PEG standards (20, 30 and 40 kDa) were obtained from JenKem Technologies (Plano, TX, USA). All PEG polymer samples were prepared in 50:50 water: acetonitrile (v:v) to a concentration of 2-10 mg/mL. The reaction mixture sample was prepared in DCM with the addition of TEA by mixing 3 equivalents of 3-maleimidopropionic acid N-hydroxysuccinimide ester with 1 equivalent of 6 kDa hydroxyl-functionalized PEG standard.

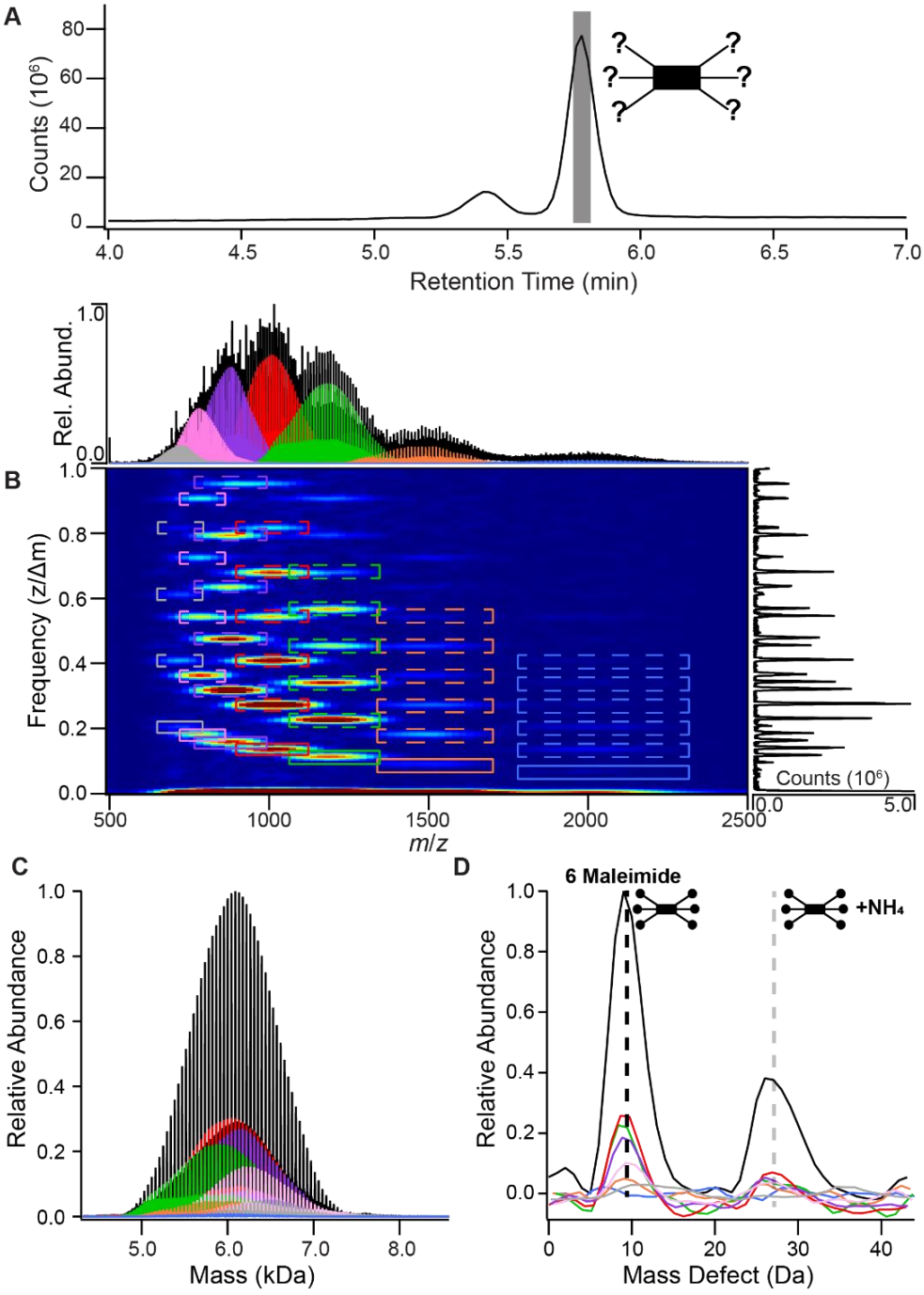


Figure S5 Demonstration of the iFAMS Macromolecular Mass Defect MMD analysis workflow on branched polymer data. (A) the polymer sample is separated on a reversed-phase LC column to separate components and contaminant populations by polarity, the chromatographic features of which are extracted to MS1 data. (B) Mass spectra extracted from the chromatogram are Gábor transformed to separate out charge state information across a 2-dimensional spectrogram in both *m/z* and frequency (colored boxes). Solid boxes signify the fundamental frequencies of polymer charge state over subunit mass (z/ΔM) and dashed boxes correspond to their higher harmonics. The polymer-specific selections are inverse Gábor transformed individually to be charge-normalized into a deconvolved mass spectrum (determining polymer size, C) or undergo macromolecular mass defect analysis (end-group and adduct identification, D)

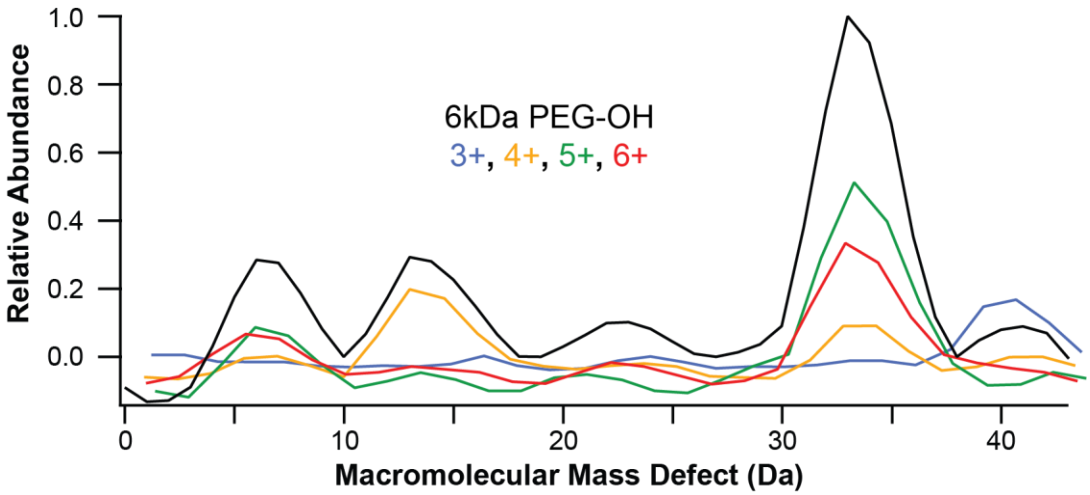


Figure S6 Charge-state-specific MMD profiles (colored) and their sum (black). Though there are other features 20-30% the size of the main features, they are not supported by all charge states and are more likely artifactual contributions from noise.

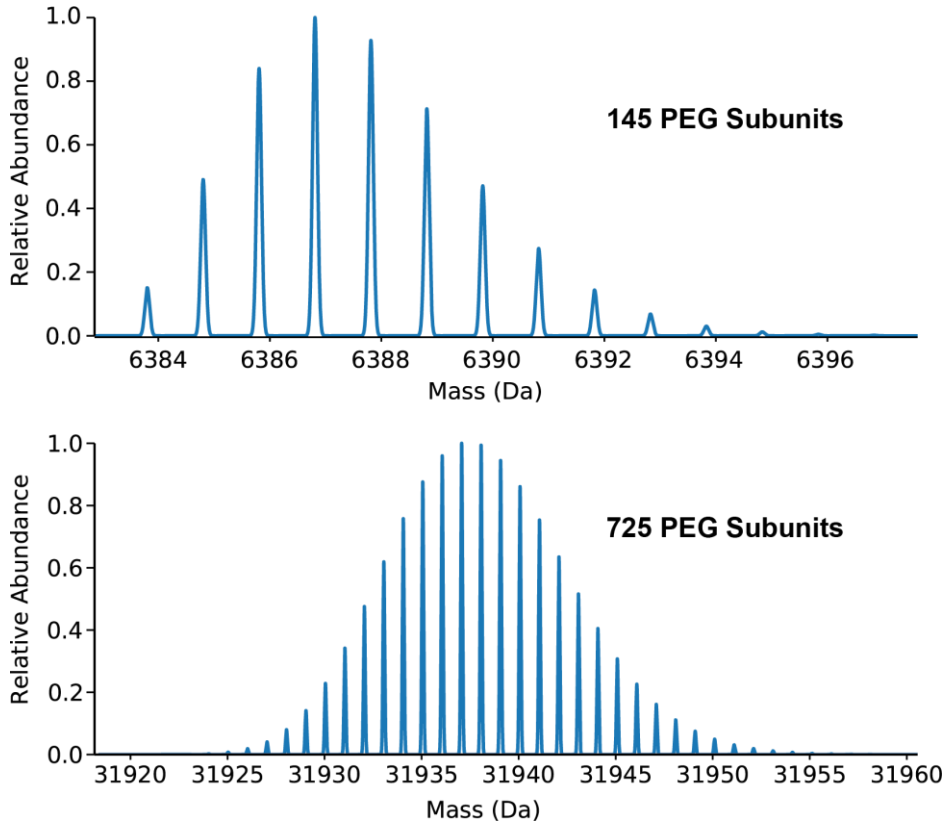


Figure S7 Calculated isotopic distributions of linear PEG with 145 ethylene glycol subunits (1X, top) and 725 subunits (5X, bottom), demonstrating the increased breadth in isoform distribution as polymer length increases.

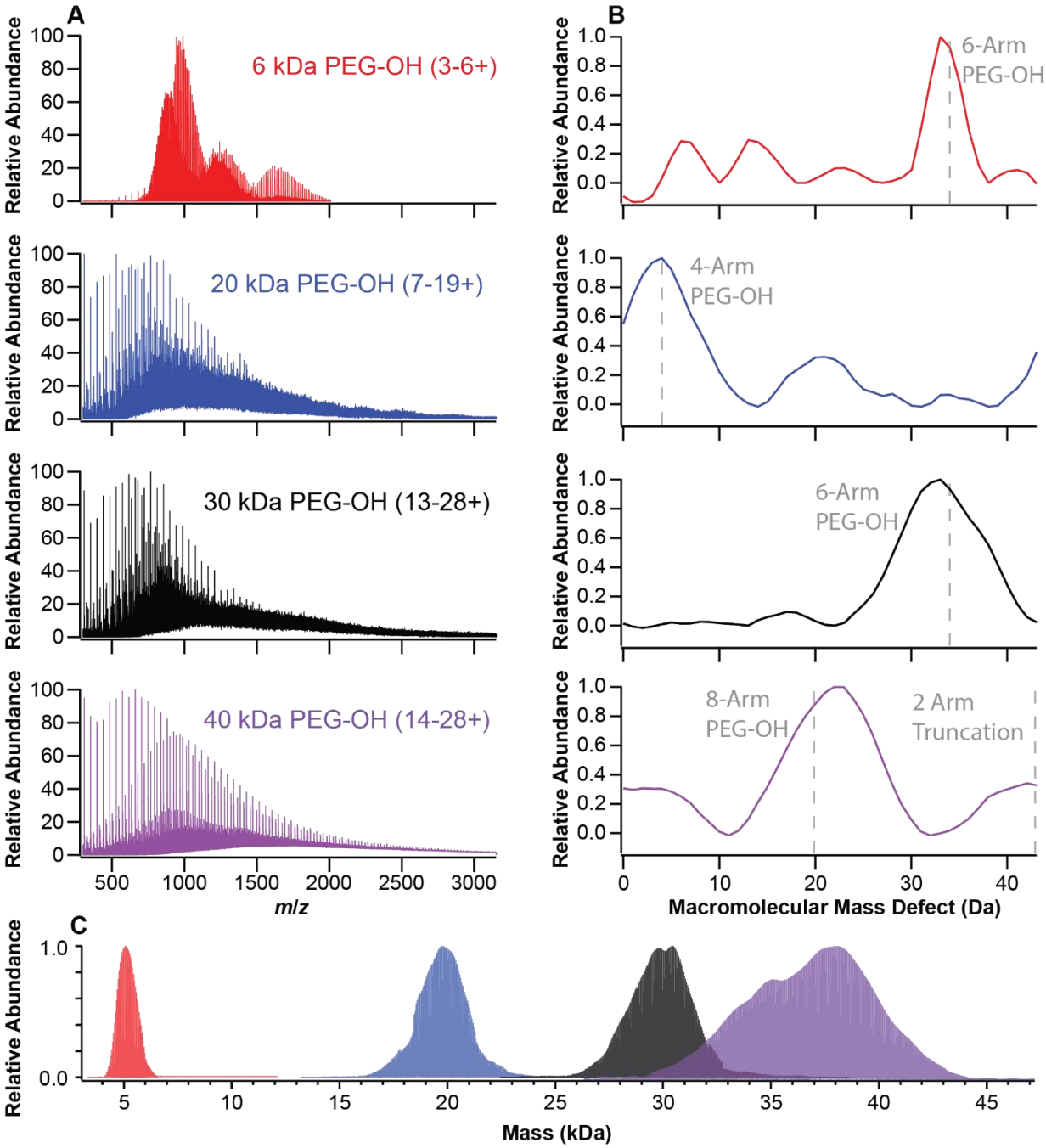


Figure S8 Deconvolution of high molecular weight PEG of various branching configurations, with purported average molecular weights of 6 kDa (red), 20 kDa (blue), 30 kDa (black), and 40 kDa (purple). Components within the PEG mass spectra (A) are differentiated by mass from their deconvolved mass reconstructions (B) and by core structures and number of end groups identified by mass defect analysis (C). The lower molecular weight shoulder in the mass reconstruction and secondary peak in the MMD profile for the 8-armed 40 kDa PEG standard polymer indicates the presence of sample impurities, likely from inhibited ethoxylation of 2 of the arms.

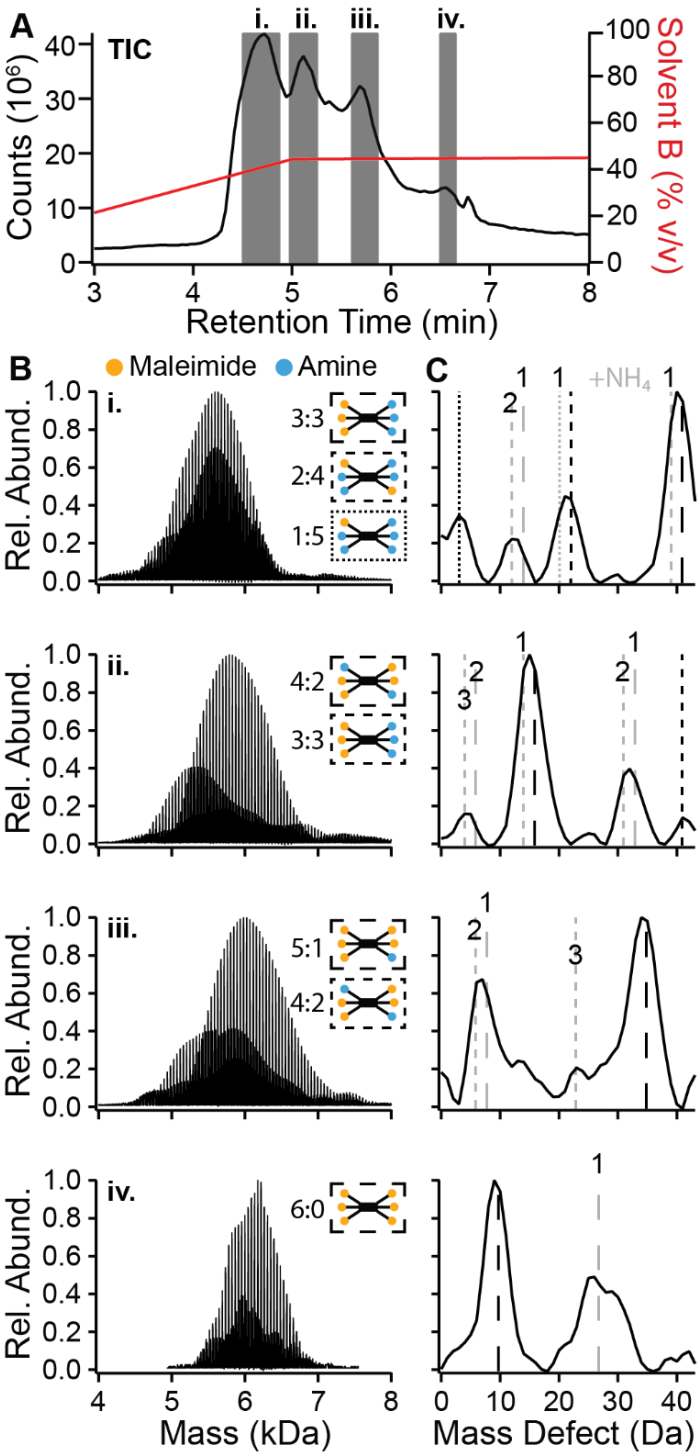


Figure S9 Ammonium adduction state assignment of asterisk labeled MMD peaks for the maleimide substitution reaction features i, ii, iii, and iv separated by LC-MS (A). The dashed line styles denote the observed end-group stoichiometries (B) for each chromatographic feature, with grey lines indicating ammonium adducted states of the respective stoichiometries (C). Grey lines are labeled with numbers to indicate the number of ammonium ions bound.

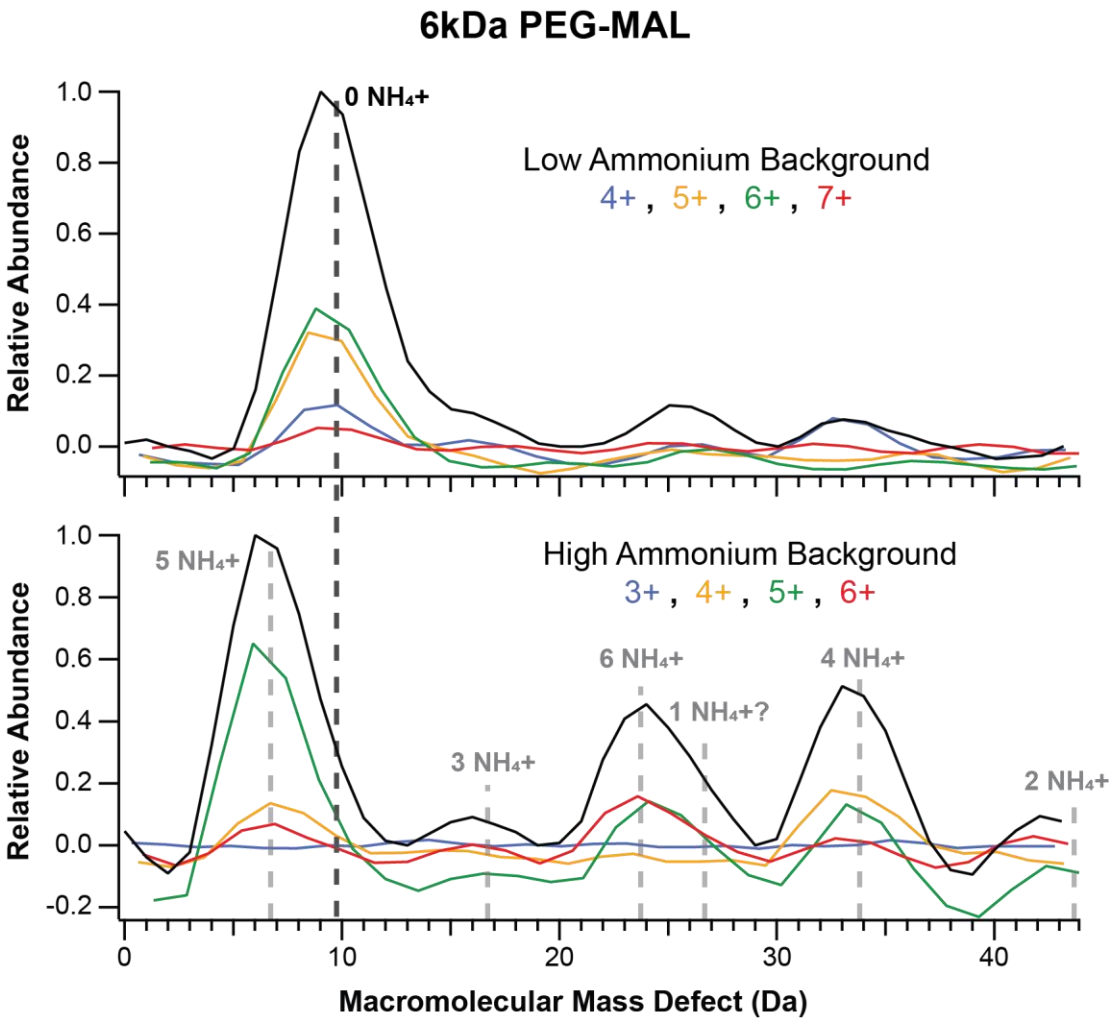


Figure S10 MMD profiles for 6-armed PEG-MAL run through two RPLC columns, one which was recently flushed with methanol (top) and one which had been run with ammonium acetate buffer (common electrospray solvent, bottom). Having a higher amount of ammonium available in solution contributes to a broader distribution of adduct states, with skewed peak shapes hinting at the possibility of overlapped features, which complicate assignment by RMSD.

Calculated Mass Defects for Ammonium Adducted Polymer Maleimidation Intermediates

Component	MW (Da)	Composition	0 NH_4^+	1 NH_4^+	2 NH_4^+	3 NH_4^+	4 NH_4^+
Core		6 MAL 0 NH3	*9.723*	*26.754*	*43.785*	*16.763*	*33.794*
Dipentaerythritol (6-arm)	248.231	5 MAL 1 NH3	*34.814*	*7.792*	*24.823*	*41.854*	*14.832*
End-groups		4 MAL 2 NH3	*15.852*	*32.883*	*5.861*	*22.892*	*39.923*
Maleimide (MAL)	195.198	3 MAL 3 NH3	*40.943*	*13.921*	*30.952*	*3.930*	*20.961*
Tertiary Amine (NH3)	44.077	2 MAL 4 NH3	*21.981*	*39.012*	*11.990*	*29.021*	*1.999*
Repeat Unit		1 MAL 5 NH3	*3.019*	*20.050*	*37.081*	*10.059*	*27.090*
Ethylene Glycol	44.053	0 MAL 6 NH3	*28.110*	*1.088*	*18.119*	*35.150*	*8.128*

Table S5 Predicted mass defect table for the maleimide substitution of 6-arm derivative PEG. The mass defects (italicized, Da) *modulo* the repeat subunit mass are organized by end-group composition (rows) and number of ammonium adducts (columns), accounting for all non-repeat components. End-group

Branched Polymer Mass Defect Peak Statistics

	PEG-OH				PEG-Ms	PEG-Amine	PEG-MAL
Average MW (kDa)	5.5	20	30	38	6	6	6
Measured MMD (Da)	33.47	3.94	33.11	21.69	17.77	28.07	10.08
Peak Width (4σ, Da)	6.97	13.73	15.94	15.97	16.16	8.24	9.86
Predicted MMD (Da)	34.014	3.988	34.014	19.987	17.941	28.110	9.723
RMSD (Da)	**0.55**	**0.050**	**0.91**	**1.70**	**0.18**	**0.042**	**0.36**

LC Retention Time (min)		MMD Peak 1	MMD Peak 2	MMD Peak 3	MMD Peak 4	**RMSD**
4.75-4.93	**Measured** MMDs (Da)	3.25	12.26	21.46	40.27	**1.01**
	Predicted MMDs (Da)	3.019	11.990	21.981	40.943	
			13.921	20.050	39.012	
4.99-5.12	**Measured** MMDs (Da)	3.87	15.357	32.32	41.30	**1.11**
	Predicted MMDs (Da)	5.861	15.852	32.883	40.943	
		3.930	13.921	30.952		
5.30-5.43	**Measured** MMDs (Da)	7.36	14.48	22.28	33.59	**1.28**
	Predicted MMDs (Da)	7.792	14.832	24.823	34.814	
		5.861		22.892	32.883	
5.74-5.82	**Measured** MMDs (Da)	8.88	27.01			**0.81**
	Predicted MMDs (Da)	9.723	26.754			
		7.792				

Table S6 Statistical data for mass defect assignments from multi-armed PEG standards and the maleimide substitution reaction. Columns for MMD profiles with multiple contributing peaks are enumerated from lowest to highest mass defect value. RMSDs for peaks with multiple assignments were taken with respect to a singular centroid rather than centroids determined by multi-peak fitting.

CHAPTER IV

Extended Methods

Brain Tissue Section Preparations for nano-DESI and Immunofluorescence

Sprague Dawley Rat brain tissues were sectioned to 12 μm thickness at -21 °C and thaw-mounted onto glass slides. Tissue sections were immersed in ethanol solutions and chloroform before nano-DESI MSI experiments, where ACN/H_2O/CH_3COOH in 65:34:1 ratio by volume was used as a working solvent. Proteins were extracted into the liquid bridge formed between two glass capillaries with working solvent spiked with bradykinin, lysozyme, beta-lactoglobulin, and equine heart myoglobin at 0.3 μM as internal standards and ionized by ESI.

The rat brain tissue sections for immunofluorescence (IF) were fixed in 4% paraformaldehyde (PFA) for 10 mins and quenched in 100 mM glycine for 10 mins before being incubated in blocking buffer for 1 h at room temperature. Tissue sections were incubated in COX7C rabbit Polyclonal primary antibody (1:100 in blocking buffer, Proteintech, Cat# 11411-2-AP) overnight at 4 °C, followed by incubation in AF647 conjugated goat anti-rabbit IgG secondary antibody (1:1000 in PBS, MyBioSource. Cat# MBS2568134) with DAPI (1:1000 in PBS, Cat# InvitrogenTM D1306) for 2 h at room temperature. The slides were then coverslipped using fluorophore mounting media.

SLURM Script Template for iFAMS Batch Deconvolution

```
#!/bin/bash

#SBATCH --partition=preempt

#SBATCH --job-name=name_of_job

#SBATCH --output=/slurm_output_file_directory/output.out

#SBATCH --error=/slurm_error_file_directory/error.err

#SBATCH --nodes=20

#SBATCH --time=01:00:00

#SBATCH --account=computing_cluster_account_name

#SBATCH --ntasks-per-node=28

#SBATCH --ntasks-per-core=1

#SBATCH --mem-per-cpu=4G


module purge

module load slurm

module load tensorflow

mpiexec -n $SLURM_NTASKS python {iFAMS_script_directory}
{batch_parameters_directory} {number of scans per raster line - 1} {pixel_spectra_directory}
{data_file_format} {number of raster lines} {output_folder_directory}
```

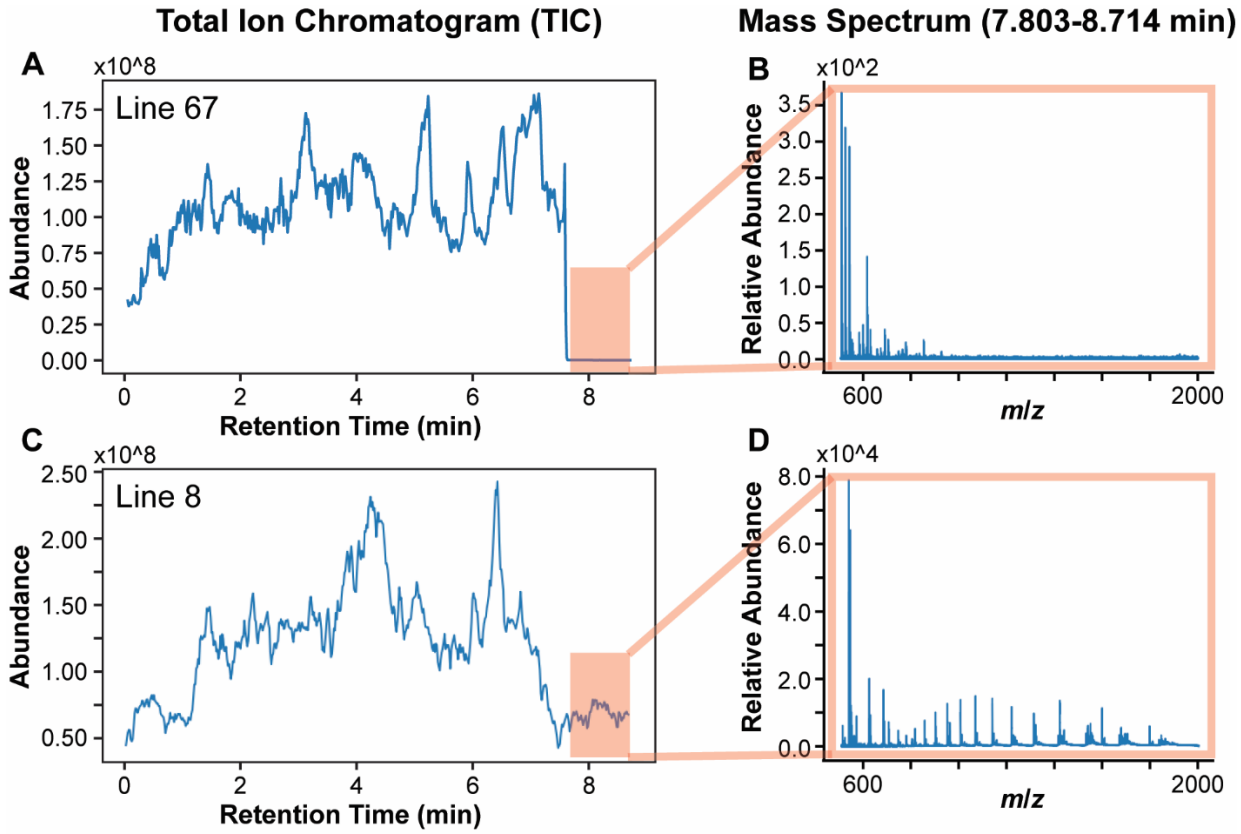


Figure S11 Comparison of total ion chromatogram (TIC) signal intensity data from two nanoDESI raster line scans (A and C). The mass spectrum collected the last few scans of line 67 (B) shows a drop of two orders of magnitude in background signal and lack of internal standard protein ions relative to the slide background collected for line 8 (D) over a similar range of scans (A and C, red boxes).

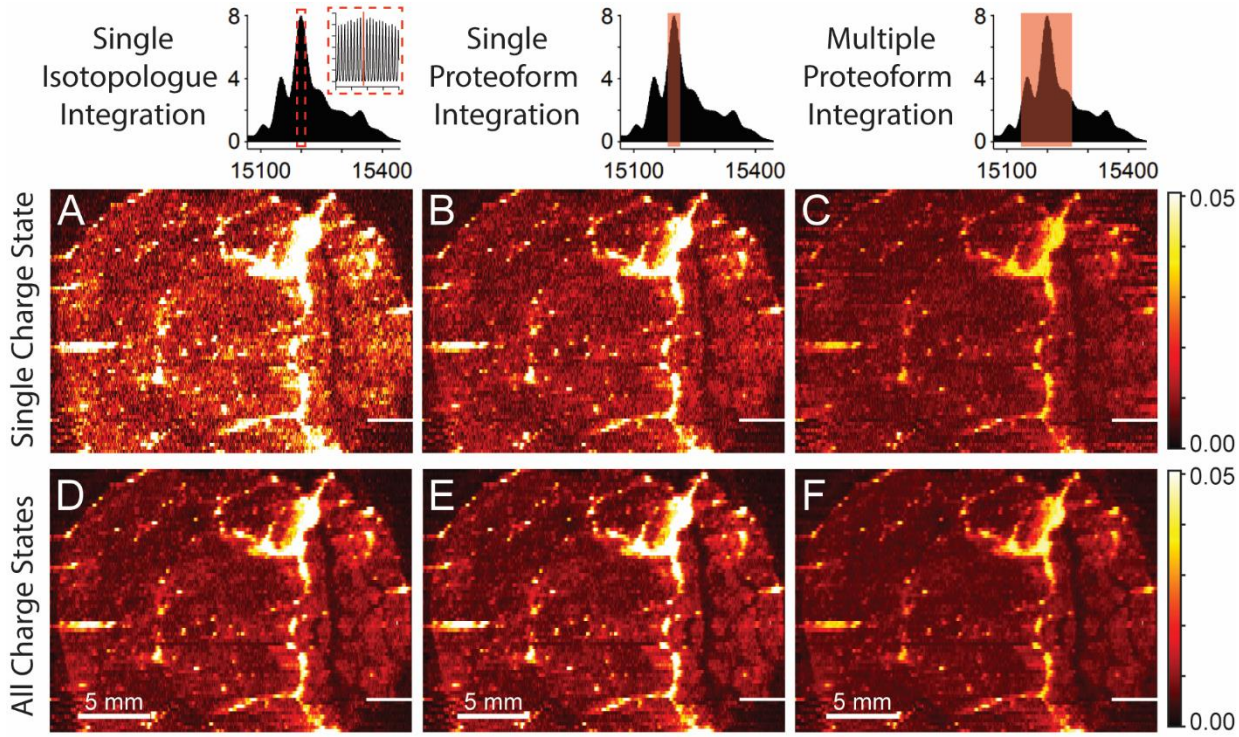


Figure S12 Protein intensity maps for hemoglobin subunit alpha (alpha-globin), normalized to the total-protein-chromatogram. Intensity maps from the single most abundant charge state (A, B, and C) exhibit higher baseline signal and noisiness (i.e., signal modulation) in regions of relatively low alpha-globin signal compared to images generated from the summed intensities across all deconvolved charge states (D, E, and F). These differences are most notable with narrower integration ranges (A, D). The bright band of rightmost pixels in line 67 are due to loss of spray (see Figure S11) Color bars on the far right and the 5 mm rulers in images D, E, and F apply to all images.

iFAMS Batch Deconvolution Parameters

128287	Batch data density
FALSE	MS domain adjustment
0	MS min
0	MS max
21.6063	noise calc frequency min
21.80671	noise calc frequency max
FALSE	MS linear baseline sub
FALSE	MS segmented baseline sub
FALSE	MS noise baseline sub
0	MS noise sub multiplyer
TRUE	smoothed deconvolved spectrum
14100	integration min x
14400	integration max x
0.1	integration min y
10	minimum distance between peaks
0.05	integration noise tolerance
FALSE	integrated segmented BC
FALSE	integrated fourier BC
FALSE	parameter match integration
3	number of peaks for bounds match
FALSE	plot real spectra
2575	frequency channels
10	oversampling

gaussian	window type
13	number of charge states
26	number of rectangles
1	number of series selected
26	series 1 final rectangle index
12	series 1 final selection index
9	selection 0 charge state
10	selection 1 charge state
11	selection 2 charge state
12	selection 3 charge state
13	selection 4 charge state
14	selection 5 charge state
15	selection 6 charge state
16	selection 7 charge state
17	selection 8 charge state
18	selection 9 charge state
19	selection 10 charge state
20	selection 11 charge state
21	selection 12 charge state
1558.762	rectangle 1 coordinate 1
1606.186	rectangle 1 coordinate 2
8.774151	rectangle 1 coordinate 3
9.217127	rectangle 1 coordinate 4
0	rectangle 1 selection number
1558.762	rectangle 2 coordinate 1

1606.186	rectangle 2 coordinate 2
17.76979	rectangle 2 coordinate 3
18.21277	rectangle 2 coordinate 4
0	rectangle 2 selection number
1402.986	rectangle 3 coordinate 1
1445.668	rectangle 3 coordinate 2
9.749057	rectangle 3 coordinate 3
10.24125	rectangle 3 coordinate 4
1	rectangle 3 selection number
1402.986	rectangle 4 coordinate 1
1445.668	rectangle 4 coordinate 2
19.74421	rectangle 4 coordinate 3
20.23641	rectangle 4 coordinate 4
1	rectangle 4 selection number
…	…
701.9928	rectangle 23 coordinate 1
723.3338	rectangle 23 coordinate 2
19.49811	rectangle 23 coordinate 3
20.48251	rectangle 23 coordinate 4
11	rectangle 23 selection number
701.9928	rectangle 24 coordinate 1
723.3338	rectangle 24 coordinate 2
39.48842	rectangle 24 coordinate 3
40.47281	rectangle 24 coordinate 4
11	rectangle 24 selection number

668.6122	rectangle 25 coordinate 1
688.937	rectangle 25 coordinate 2
20.47302	rectangle 25 coordinate 3
21.50663	rectangle 25 coordinate 4
12	rectangle 25 selection number
668.6122	rectangle 26 coordinate 1
688.937	rectangle 26 coordinate 2
41.46284	rectangle 26 coordinate 3
42.49646	rectangle 26 coordinate 4
12	rectangle 26 selection number
14102.94	low bound 1
14158.89	high bound 1
14160.93	low bound 2
14272.83	high bound 2
14280.4	low bound 3
14324.99	high bound 3
…	…

Table S7 Example batch parameter file for myelin basic protein. Ellipses show where repeated data structures were omitted for easier interpretation of the overall parameter file structure. “Batch data density” is the number of data points for data interpolation to ensure similar ranges of frequency data are analyzed for all mass spectra being batch processed. Gábor spectrogram selections are grouped by charge state and stored as ranges of cartesian coordinates in *m*/*z* and frequency. Integration parameters for the deconvolved zero-charge spectra are stored at the very end as sequential mass minima and maxima.

REFERENCES CITED

(1) 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. DOI: 10.1063/1.1670391.

(2) Fenn, J.; Mann, M.; Meng, C.; Wong, S.; Whitehouse, C. Electrospray ionization for mass spectrometry of large biomolecules. *Science* **1989**, *246* (4926), 64-71. DOI: 10.1126/science.2675315.

(3) van Berkel, W. J.; van den Heuvel, R. H.; Versluis, C.; Heck, A. J. Detection of intact megaDalton protein assemblies of vanillyl-alcohol oxidase by mass spectrometry. *Protein Sci.* **2000**, *9* (3), 435-439. DOI: 10.1110/ps.9.3.435 PubMed.

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

(5) 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? *Nature Chemical Biology* **2018**, *14* (3), 206-214. DOI: 10.1038/nchembio.2576.

(6) Siuti, N.; Kelleher, N. L. Decoding protein modifications using top-down mass spectrometry. *Nat. Methods* **2007**, *4* (10), 817-821. DOI: 10.1038/nmeth1097.

(7) Donnelly, D. P.; Rawlins, C. M.; DeHart, C. J.; Fornelli, L.; Schachner, L. F.; Lin, Z.; Lippens, J. L.; Aluri, K. C.; Sarin, R.; Chen, B.; et al. Best practices and benchmarks for intact protein analysis for top-down mass spectrometry. *Nat. Methods* **2019**, *16* (7), 587-594. DOI: 10.1038/s41592-019-0457-0.

(8) Roberts, D. S.; Loo, J. A.; Tsybin, Y. O.; Liu, X.; Wu, S.; Chamot-Rooke, J.; Agar, J. N.; Paša-Tolić, L.; Smith, L. M.; Ge, Y. Top-down proteomics. *Nature Reviews Methods Primers* **2024**, *4* (1), 38. DOI: 10.1038/s43586-024-00318-2.

(9) Tamara, S.; den Boer, M. A.; Heck, A. J. R. High-Resolution Native Mass Spectrometry. *Chem. Rev.* **2022**, *122* (8), 7269-7326. DOI: 10.1021/acs.chemrev.1c00212.

(10) Ruotolo, B. T.; Robinson, C. V. Aspects of native proteins are retained in vacuum. *Curr. Opin. Chem. Biol.* **2006**, *10* (5), 402-408. DOI: 10.1016/j.cbpa.2006.08.020.

(11) Marcoux, J.; Robinson, Carol V. Twenty Years of Gas Phase Structural Biology. *Structure* **2013**, *21* (9), 1541-1550. DOI: 10.1016/j.str.2013.08.002.

(12) Borysik, A. J.; Hewitt, D. J.; Robinson, C. V. Detergent Release Prolongs the Lifetime of Native-like Membrane Protein Conformations in the Gas-Phase. *J. Am. Chem. Soc.* **2013**, *135* (16), 6078-6083. DOI: 10.1021/ja401736v.

(13) Rolland, A. D.; Prell, J. S. Approaches to Heterogeneity in Native Mass Spectrometry. *Chem. Rev.* **2021**. DOI: 10.1021/acs.chemrev.1c00696.

(14) Wilson, J. W.; Rolland, A. D.; Klausen, G. M.; Prell, J. S. Ion Mobility-Mass Spectrometry Reveals That α-Hemolysin from *Staphylococcus aureus* Simultaneously Forms Hexameric and Heptameric Complexes in Detergent Micelle Solutions. *Anal. Chem.* **2019**, *91* (15), 10204-10211. DOI: 10.1021/acs.analchem.9b02243.

(15) van den Heuvel, R. H. H.; Heck, A. J. R. Native protein mass spectrometry: from intact oligomers to functional machineries. *Curr. Opin. Chem. Biol.* **2004**, *8* (5), 519-526. DOI: 10.1016/j.cbpa.2004.08.006.

(16) Marty, M. T.; Hoi, K. K.; Robinson, C. V. Interfacing Membrane Mimetics with Mass Spectrometry. *Acc. Chem. Res.* **2016**, *49* (11), 2459-2467. DOI: 10.1021/acs.accounts.6b00379.

(17) Marty, M. T. Nanodiscs and mass spectrometry: Making membranes fly. *Int. J. Mass Spectrom.* **2020**, *458,* 116436. DOI: 10.1016/j.ijms.2020.116436.

(18) Harvey, S. R.; Liu, Y.; Liu, W.; Wysocki, V. H.; Laganowsky, A. Surface induced dissociation as a tool to study membrane protein complexes. *Chem. Commun.* **2017**, *53* (21), 3106-3109, 10.1039/C6CC09606A. DOI: 10.1039/C6CC09606A.

(19) Han, L.; Nguyen, L.; Schmidt, E. N.; Esmaili, M.; Kitova, E. N.; Overduin, M.; Macauley, M. S.; Klassen, J. S. How Choice of Model Membrane Affects Protein–Glycosphingolipid Interactions: Insights from Native Mass Spectrometry. *Anal. Chem.* **2022**, *94* (46), 16042-16049. DOI: 10.1021/acs.analchem.2c03067.

(20) Leney, A. C.; Rezaei Darestani, R.; Li, J.; Nikjah, S.; Kitova, E. N.; Zou, C.; Cairo, C. W.; Xiong, Z. J.; Privé, G. G.; Klassen, J. S. Picodiscs for Facile Protein-Glycolipid Interaction Analysis. *Anal. Chem.* **2015**, *87* (8), 4402-4408. DOI: 10.1021/acs.analchem.5b00170.

(21) Urner, L. H.; Fiorentino, F.; Shutin, D.; Sauer, J. B.; Agasid, M. T.; El-Baba, T. J.; Bolla, J. R.; Stansfeld, P. J.; Robinson, C. V. Detergents with Scalable Properties Identify Noncanonical Lipopolysaccharide Binding to Bacterial Inner Membrane Proteins. *J. Am. Chem. Soc.* **2024**, *146* (16), 11025-11030. DOI: 10.1021/jacs.3c14358.

(22) Zhu, Y.; Odenkirk, M. T.; Qiao, P.; Zhang, T.; Schrecke, S.; Zhou, M.; Marty, M. T.; Baker, E. S.; Laganowsky, A. Combining native mass spectrometry and lipidomics to uncover specific membrane protein–lipid interactions from natural lipid sources. *Chem. Sci.* **2023**, *14* (32), 8570-8582, 10.1039/D3SC01482G. DOI: 10.1039/D3SC01482G.

(23) Valentine, S. J.; Liu, X.; Plasencia, M. D.; Hilderbrand, A. E.; Kurulugama, R. T.; Koeniger, S. L.; Clemmer, D. E. Developing liquid chromatography ion mobility mass spectometry techniques. *Expert Review of Proteomics* **2005**, *2* (4), 553-565. DOI: 10.1586/14789450.2.4.553.

(24) Mitchell Wells, J.; McLuckey, S. A. Collision-Induced Dissociation (CID) of Peptides and Proteins. In *Methods in Enzymology*, Vol. 402; Academic Press, 2005; pp 148-185.

(25) 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. DOI: 10.1016/j.cbpa.2017.11.010.

(26) Morrison, K. A.; Clowers, B. H. Fundamentals and applications of incorporating chromatographic separations with ion mobility-mass spectrometry. *Trends Anal. Chem.* **2019**, *119*, 115625. DOI: 10.1016/j.trac.2019.115625.

(27) Zheng, X.; Kurulugama, R. T.; Laganowsky, A.; Russell, D. H. Collision-Induced Unfolding Studies of Proteins and Protein Complexes using Drift Tube Ion Mobility-Mass Spectrometer. *Anal. Chem.* **2020**, *92* (10), 7218-7225. DOI: 10.1021/acs.analchem.0c00772.

(28) Phetsanthad, A.; Li, G.; Jeon, C. K.; Ruotolo, B. T.; Li, L. Comparing Selected-Ion Collision Induced Unfolding with All Ion Unfolding Methods for Comprehensive Protein Conformational Characterization. *J. Am. Soc. Mass Spectrom.* **2022**, *33* (6), 944-951. DOI: 10.1021/jasms.2c00004.

(29) Ghosh, S. K.; Walocha, J. A. Evolution of staining methods in neuroanatomy: Impetus for emanation of neuron doctrine during the turn of 20th century. *The Anatomical Record* **2024**. DOI: 10.1002/ar.25436 (acccessed 2024/07/29).

(30) Nguyen, S. N.; Sontag, R. L.; Carson, J. P.; Corley, R. A.; Ansong, C.; Laskin, J. Towards High-Resolution Tissue Imaging Using Nanospray Desorption Electrospray Ionization Mass Spectrometry Coupled to Shear Force Microscopy. *J. Am. Soc. Mass Spectrom.* **2018**, *29* (2), 316-322. DOI: 10.1007/s13361-017-1750-8.

(31) Akaji, S.; Sagawa, T.; Honda, A.; Miyasaka, N.; Sadakane, K.; Ichinose, T.; Takano, H. Post-staining Raman analysis of histological sections following decolorization. *Analyst* **2022**, *147* (20), 4473-4479, 10.1039/D2AN01138G. DOI: 10.1039/D2AN01138G.

(32) Unsihuay, D.; Hu, H.; Qiu, J.; Latorre-Palomino, A.; Yang, M.; Yue, F.; Yin, R.; Kuang, S.; Laskin, J. Multimodal high-resolution nano-DESI MSI and immunofluorescence imaging reveal molecular signatures of skeletal muscle fiber types. *Chem. Sci.* **2023**, *14* (15), 4070-4082, 10.1039/D2SC06020E. DOI: 10.1039/D2SC06020E.

(33) Suarez, S.; Ferroni, A.; Lotz, A.; Jolley, K. A.; Guérin, P.; Leto, J.; Dauphin, B.; Jamet, A.; Maiden, M. C. J.; Nassif, X.; et al. Ribosomal proteins as biomarkers for bacterial identification by mass spectrometry in the clinical microbiology laboratory. *Journal of Microbiological Methods* **2013**, *94* (3), 390-396. DOI: 10.1016/j.mimet.2013.07.021.

(34) Zhang, Y.; Jiao, J.; Yang, P.; Lu, H. Mass spectrometry-based N-glycoproteomics for cancer biomarker discovery. *Clinical Proteomics* **2014**, *11* (1), 18. DOI: 10.1186/1559-0275-11-18.

(35) Abramczyk, H.; Imiela, A.; Brożek-Płuska, B.; Kopeć, M.; Surmacki, J.; Śliwińska, A. Aberrant Protein Phosphorylation in Cancer by Using Raman Biomarkers. In *Cancers*, 2019; Vol. 11.

(36) Marty, M. T.; Baldwin, A. J.; Marklund, E. G.; Hochberg, G. K. A.; Benesch, J. L. P.; Robinson, C. V. Bayesian Deconvolution of Mass and Ion Mobility Spectra: From Binary Interactions to Polydisperse Ensembles. *Anal. Chem.* **2015**, *87* (8), 4370-4376. DOI: 10.1021/acs.analchem.5b00140.

(37) Bern, M.; Caval, T.; Kil, Y. J.; Tang, W.; Becker, C.; Carlson, E.; Kletter, D.; Sen, K. I.; Galy, N.; Hagemans, D.; et al. Parsimonious Charge Deconvolution for Native Mass Spectrometry. *J. Proteome Res.* **2018**, *17* (3), 1216-1226. DOI: 10.1021/acs.jproteome.7b00839.

(38) *MaxEnt: An Essential Maximum Entropy Based Tool for Interpreting Multiply-Charged Electrospray Data*; 1998, 1998. https://www.waters.com/webassets/cms/library/docs/an212.pdf (accessed August 19, 2021).

(39) Skilling, J. Data analysis: The maximum entropy method. *Nature* **1984**, *309* (5971), 748-749. DOI: 10.1038/309748a0.

(40) Reinhold, B. B.; Reinhold, V. N. Electrospray ionization mass spectrometry: Deconvolution by an Entropy-Based algorithm. *J. Am. Soc. Mass Spectrom.* **1992**, *3* (3), 207-215. DOI: 10.1021/jasms.8b00237.

(41) Reid, D. J.; Diesing, J. M.; Miller, M. A.; Perry, S. M.; Wales, J. A.; Montfort, W. R.; Marty, M. T. MetaUniDec: High-Throughput Deconvolution of Native Mass Spectra. *J. Am. Soc. Mass Spectrom.* **2018**, *30*, 118-127. DOI: 10.1007/s13361-018-1951-9.

(42) Marty, M. T. A Universal Score for Deconvolution of Intact Protein and Native Electrospray Mass Spectra. *Anal. Chem.* **2020**, *92* (6), 4395-4401. DOI: 10.1021/acs.analchem.9b05272.

(43) Campuzano, I. D. G.; Robinson, J. H.; Hui, J. O.; Shi, S. D. H.; Netirojjanakul, C.; Nshanian, M.; Egea, P. F.; Lippens, J. L.; Bagal, D.; Loo, J. A.; et al. Native and Denaturing MS Protein Deconvolution for Biopharma: Monoclonal Antibodies and Antibody–Drug Conjugates to Polydisperse Membrane Proteins and Beyond. *Anal. Chem.* **2019**, *91* (15), 9472-9480. DOI: 10.1021/acs.analchem.9b00062.

(44) Townsend, J. A.; Keener, J. E.; Miller, Z. M.; Prell, J. S.; Marty, M. T. Imidazole Derivatives Improve Charge Reduction and Stabilization for Native Mass Spectrometry. *Anal. Chem.* **2019**, *91* (22), 14765-14772. DOI: 10.1021/acs.analchem.9b04263.

(45) Cleary, S. P.; Li, H.; Bagal, D.; Loo, J. A.; Campuzano, I. D. G.; Prell, J. S. Extracting Charge and Mass Information from Highly Congested Mass Spectra Using Fourier-Domain Harmonics. *J. Am. Soc. Mass Spectrom.* **2018**, *29* (10), 2067-2080, journal article. DOI: 10.1007/s13361-018-2018-7.

(46) Ernst, R. R. Nuclear Magnetic Resonance Fourier Transform Spectroscopy (Nobel Lecture). *Angewandte Chemie International Edition in English* **1992**, *31* (7), 805-823. DOI: 10.1002/anie.199208053 (acccessed 2024/07/29).

(47) Coggins, B. E.; Zhou, P. Polar Fourier transforms of radially sampled NMR data. *J. Magn. Reson.* **2006**, *182* (1), 84-95. DOI: 10.1016/j.jmr.2006.06.016.

(48) Douard, M.; Bacis, R.; Rambaldi, P.; Ross, A.; Wolf, J. P.; Fabre, G.; Stringat, R. Fourier-transform lidar. *Opt. Lett.* **1995**, *20* (20), 2140-2142. DOI: 10.1364/OL.20.002140.

(49) Wu, L.; Li, Z.; Han, Y.; Mai, S.; Xing, X.; Fu, H. Y. Signal Processing Using Wavelet Transform and Short-time Fourier Transform Based on Spectral-scanning FMCW LiDAR. In *2022 Asia Communications and Photonics Conference (ACP)*, 5-8 Nov. 2022, 2022; pp 789-791. DOI: 10.1109/ACP55869.2022.10088727.

(50) Park, J.; Kim, W.; Han, D. K.; Ko, H. Voice Activity Detection in Noisy Environments Based on Double-Combined Fourier Transform and Line Fitting. *The Scientific World Journal* **2014**, *2014* (1), 146040. DOI: 10.1155/2014/146040 (acccessed 2024/07/29).

(51) Prebyl, B. S.; Cook, K. D. Use of Fourier Transform for Deconvolution of the Unresolved Envelope Observed in Electrospray Ionization Mass Spectrometry of Strongly Ionic Synthetic Polymers. *Anal. Chem.* **2004**, *76* (1), 127-136. DOI: 10.1021/ac0348266.

(52) Cleary, S. P.; Thompson, A. M.; Prell, J. S. Fourier Analysis Method for Analyzing Highly Congested Mass Spectra of Ion Populations with Repeated Subunits. *Anal. Chem.* **2016**, *88* (12), 6205-6213. DOI: 10.1021/acs.analchem.6b01088.

(53) Cleary, S. P.; Prell, J. S. Liberating Native Mass Spectrometry from Dependence on Volatile Salt Buffers by Use of Gábor Transform. *ChemPhysChem* **2019**, *20* (4), 519-523. DOI: 10.1002/cphc.201900022.

(54) Kitov, P. I.; Han, L.; Kitova, E. N.; Klassen, J. S. Sliding Window Adduct Removal Method (SWARM) for Enhanced Electrospray Ionization Mass Spectrometry Binding Data. *J. Am. Soc. Mass Spectrom.* **2019**, *30* (8), 1446-1454. DOI: 10.1007/s13361-019-02204-8.

(55) Cleary, S. P.; Prell, J. S. Distinct classes of multi-subunit heterogeneity: analysis using Fourier Transform methods and native mass spectrometry. *Analyst* **2020**, *145* (13), 4688-4697, 10.1039/D0AN00726A. DOI: 10.1039/D0AN00726A.

(56) Struwe, W. B.; Robinson, C. V. Relating glycoprotein structural heterogeneity to function – insights from native mass spectrometry. *Curr. Opin. Struct. Biol.* **2019**, *58*, 241-248. DOI: 10.1016/j.sbi.2019.05.019.

(57) 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. U. S. A.* **2018**, *115* (35), 8763-8768. DOI: 10.1073/pnas.1807439115.

(58) Wohlschlager, T.; Scheffler, K.; Forstenlehner, I. C.; Skala, W.; Senn, S.; Damoc, E.; Holzmann, J.; Huber, C. G. Native mass spectrometry combined with enzymatic dissection unravels glycoform heterogeneity of biopharmaceuticals. *Nat. Commun.* **2018**, *9* (1), 1713. DOI: 10.1038/s41467-018-04061-7.

(59) Gupta, K.; Donlan, J. A. C.; Hopper, J. T. S.; Uzdavinys, P.; Landreh, M.; Struwe, W. B.; Drew, D.; Baldwin, A. J.; Stansfeld, P. J.; Robinson, C. V. The role of interfacial lipids in stabilizing membrane protein oligomers. *Nature* **2017**, *541* (7637), 421-424. DOI: 10.1038/nature20820.

(60) Laganowsky, A.; Reading, E.; Allison, T. M.; Ulmschneider, M. B.; Degiacomi, M. T.; Baldwin, A. J.; Robinson, C. V. Membrane proteins bind lipids selectively to modulate their structure and function. *Nature* **2014**, *510* (7503), 172-175, Article. DOI: 10.1038/nature13419.

(61) van de Waterbeemd, M.; Snijder, J.; Tsvetkova, I. B.; Dragnea, B. G.; Cornelissen, J. J.; Heck, A. J. R. Examining the Heterogeneous Genome Content of Multipartite Viruses BMV and CCMV by Native Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **2016**, *27* (6), 1000-1009. DOI: 10.1021/jasms.8b05284.

(62) Agasid, M. T.; Sørensen, L.; Urner, L. H.; Yan, J.; Robinson, C. V. The Effects of Sodium Ions on Ligand Binding and Conformational States of G Protein-Coupled Receptors—Insights from Mass Spectrometry. *J. Am. Chem. Soc.* **2021**, *143* (11), 4085-4089. DOI: 10.1021/jacs.0c11837.

(63) Boeri Erba, E.; Petosa, C. The emerging role of native mass spectrometry in characterizing the structure and dynamics of macromolecular complexes. *Protein Sci.* **2015**, *24* (8), 1176-1192, https://doi.org/10.1002/pro.2661. DOI: 10.1002/pro.2661 (acccessed 2021/07/02).

(64) Ruotolo, B. T.; Marty, M. T. Native mass spectrometry for structural biology: A perspective. *Int. J. Mass Spectrom.* **2021**, *468*, 116655. DOI: 10.1016/j.ijms.2021.116655.

(65) Thompson, N. J.; Rosati, S.; Rose, R. J.; Heck, A. J. R. The impact of mass spectrometry on the study of intact antibodies: from post-translational modifications to structural analysis. *Chem. Commun.* **2013**, *49* (6), 538-548, 10.1039/C2CC36755F. DOI: 10.1039/C2CC36755F.

(66) 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. DOI: 10.1021/acs.analchem.7b03021.

(67) Marculescu, C.; Lakshminarayanan, A.; Gault, J.; Knight, J. C.; Folkes, L. K.; Spink, T.; Robinson, C. V.; Vallis, K.; Davis, B. G.; Cornelissen, B. Probing the limits of Q-tag bioconjugation of antibodies. *Chem. Commun.* **2019**, *55* (76), 11342-11345, 10.1039/C9CC02303H. DOI: 10.1039/C9CC02303H.

(68) Rolland, A. D.; Prell, J. S. Computational insights into compaction of gas-phase protein and protein complex ions in native ion mobility-mass spectrometry. *Trends Anal. Chem.* **2019**, *116*, 282-291. DOI: 10.1016/j.trac.2019.04.023.

(69) Landreh, M.; Costeira-Paulo, J.; Gault, J.; Marklund, E. G.; Robinson, C. V. Effects of Detergent Micelles on Lipid Binding to Proteins in Electrospray Ionization Mass Spectrometry. *Anal. Chem.* **2017**, *89* (14), 7425-7430. DOI: 10.1021/acs.analchem.7b00922.

(70) Landreh, M.; Marty, M. T.; Gault, J.; Robinson, C. V. A sliding selectivity scale for lipid binding to membrane proteins. *Curr. Opin. Struct. Biol.* **2016**, *39*, 54-60. DOI: 10.1016/j.sbi.2016.04.005.

(71) Gupta, K.; Li, J.; Liko, I.; Gault, J.; Bechara, C.; Wu, D.; Hopper, J. T. S.; Giles, K.; Benesch, J. L. P.; Robinson, C. V. Identifying key membrane protein lipid interactions using mass spectrometry. *Nat. Protoc.* **2018**, *13* (5), 1106-1120. DOI: 10.1038/nprot.2018.014.

(72) Patrick, J. W.; Boone, C. D.; Liu, W.; Conover, G. M.; Liu, Y.; Cong, X.; Laganowsky, A. Allostery revealed within lipid binding events to membrane proteins. *Proc. Natl. Acad. Sci. U. S. A.* **2018**, *115* (12), 2976-2981. DOI: 10.1073/pnas.1719813115.

(73) Lippens, J. L.; Nshanian, M.; Spahr, C.; Egea, P. F.; Loo, J. A.; Campuzano, I. D. G. Fourier Transform-Ion Cyclotron Resonance Mass Spectrometry as a Platform for Characterizing Multimeric Membrane Protein Complexes. *J. Am. Soc. Mass Spectrom.* **2018**, *29* (1), 183-193. DOI: 10.1007/s13361-017-1799-4.

(74) Wörner, T. P.; Snijder, J.; Bennett, A.; Agbandje-McKenna, M.; Makarov, A. A.; Heck, A. J. R. Resolving heterogeneous macromolecular assemblies by Orbitrap-based single-particle charge detection mass spectrometry. *Nat. Methods* **2020**, *17* (4), 395-398. DOI: 10.1038/s41592-020-0770-7.

(75) Olinares, P. D. B.; Kang, J. Y.; Llewellyn, E.; Chiu, C.; Chen, J.; Malone, B.; Saecker, R. M.; Campbell, E. A.; Darst, S. A.; Chait, B. T. Native Mass Spectrometry-Based Screening for Optimal Sample Preparation in Single-Particle Cryo-EM. *Structure* **2021**, *29* (2), 186-195.e186. DOI: 10.1016/j.str.2020.11.001.

(76) Marty, M. T.; Zhang, H.; Cui, W. D.; Blankenship, R. E.; Gross, M. L.; Sligar, S. G. Native Mass Spectrometry Characterization of Intact Nanodisc Lipoprotein Complexes. *Anal. Chem.* **2012**, *84* (21), 8957-8960, Article. DOI: 10.1021/ac302663f.

(77) Kostelic, M. M.; Marty, M. T. Deconvolving Native and Intact Protein Mass Spectra with UniDec. *ChemRxiv* **2020**, DOI: 10.26434/chemrxiv.13417118.v13417111. DOI: 10.26434/chemrxiv.13417118.v1.

(78) Marty, M. T.; Zhang, H.; Cui, W. D.; Gross, M. L.; Sligar, S. G. Interpretation and Deconvolution of Nanodisc Native Mass Spectra. *J. Am. Soc. Mass Spectrom.* **2014**, *25* (2), 269-277, Article. DOI: 10.1007/s13361-013-0782-y.

(79) Hoi, K. K.; Robinson, C. V.; Marty, M. T. Unraveling the Composition and Behavior of Heterogeneous Lipid Nanodiscs by Mass Spectrometry. *Anal. Chem.* **2016**, *88* (12), 6199-6204. DOI: 10.1021/acs.analchem.6b00851.

(80) Marty, M. T.; Hoi, K. K.; Gault, J.; Robinson, C. V. Probing the Lipid Annular Belt by Gas-Phase Dissociation of Membrane Proteins in Nanodiscs. *Angew. Chem. Int. Edit.* **2016**, *55* (2), 550-554, Article. DOI: 10.1002/anie.201508289.

(81) Walker, L. R.; Marzluff, E. M.; Townsend, J. A.; Resager, W. C.; Marty, M. T. Native Mass Spectrometry of Antimicrobial Peptides in Lipid Nanodiscs Elucidates Complex Assembly. *Anal. Chem.* **2019**, *91* (14), 9284-9291. DOI: 10.1021/acs.analchem.9b02261.

(82) Keener, J. E.; Zambrano, D. E.; Zhang, G.; Zak, C. K.; Reid, D. J.; Deodhar, B. S.; Pemberton, J. E.; Prell, J. S.; Marty, M. T. Chemical Additives Enable Native Mass Spectrometry Measurement of Membrane Protein Oligomeric State within Intact Nanodiscs. *J. Am. Chem. Soc.* **2019**, *141* (2), 1054-1061. DOI: 10.1021/jacs.8b11529.

(83) Kendrick, E. A Mass Scale Based on CH_2 = 14.0000 for High Resolution Mass Spectrometry of Organic Compounds. *Anal. Chem.* **1963**, *35* (13), 2146-2154. DOI: 10.1021/ac60206a048.

(84) Hughey, C. A.; Hendrickson, C. L.; Rodgers, R. P.; Marshall, A. G.; Qian, K. Kendrick Mass Defect Spectrum: A Compact Visual Analysis for Ultrahigh-Resolution Broadband Mass Spectra. *Anal. Chem.* **2001**, *73* (19), 4676-4681. DOI: 10.1021/ac010560w.

(85) Fouquet, T.; Sato, H. Extension of the Kendrick Mass Defect Analysis of Homopolymers to Low Resolution and High Mass Range Mass Spectra Using Fractional Base Units. *Anal. Chem.* **2017**, *89* (5), 2682-2686. DOI: 10.1021/acs.analchem.6b05136.

(86) Civjan, N. R.; Bayburt, T. H.; Schuler, M. A.; Sligar, S. G. Direct solubilization of heterologously expressed membrane proteins by incorporation into nanoscale lipid bilayers. *BioTechniques* **2003**, *35* (3), 556-563. DOI: 10.2144/03353rr02.

(87) Denisov, I. G.; Sligar, S. G. Nanodiscs for structural and functional studies of membrane proteins. *Nat. Struct. Mol. Biol.* **2016**, *23* (6), 481-486. DOI: 10.1038/nsmb.3195.

(88) Denisov, I. G.; Sligar, S. G. Nanodiscs in Membrane Biochemistry and Biophysics. *Chem. Rev.* **2017**, *117* (6), 4669-4713, Review. DOI: 10.1021/acs.chemrev.6b00690.

(89) Denisov, I. G.; Grinkova, Y. V.; Lazarides, A. A.; Sligar, S. G. Directed Self-Assembly of Monodisperse Phospholipid Bilayer Nanodiscs with Controlled Size. *J. Am. Chem. Soc.* **2004**, *126* (11), 3477-3487, Article. DOI: 10.1021/ja0393574.

(90) Ritchie, T. K.; Grinkova, Y. V.; Bayburt, T. H.; Denisov, I. G.; Zolnerciks, J. K.; Atkins, W. M.; Sligar, S. G. Reconstitution of Membrane Proteins in Phospholipid Bilayer Nanodiscs. *Methods Enzymol.* **2009**, *464*, 211-231. DOI: 10.1016/S0076-6879(09)64011-8 PMC.

(91) Marty, M. T. Eliminating Artifacts in Electrospray Deconvolution with a SoftMax Function. *J. Am. Soc. Mass Spectrom.* **2019**, *30* (10), 2174-2177. DOI: 10.1021/jasms.8b06232.

(92) Bleistein, N.; Handelsman, R. A. Fourier integrals and the method of stationary phase. In *Asymptotic expansions of integrals*, 1st ed.; Holt, Rinehart and Winston, 1975; pp 219-223.

(93) Harris, D. C. Quality assurance and calibration methods. In *Quantitative chemical analysis*, 8th ed.; Byrd, M. L. Ed.; Marshall, Clancy, 2010; pp 100-105.

(94) Zeng, G. L. Gibbs artifact reduction by nonnegativity constraint. *J. Nucl. Med. Technol.* **2011**, *39* (3), 213-219. DOI: 10.2967/jnmt.110.086439 PubMed.

(95) Kostelic, M. M.; Zak, C. K.; Jayasekera, H.; Marty, M. T. Assembly of Model Membrane Nanodiscs for Native Mass Spectrometry. *Anal. Chem.* **2021**, *93* (14), 5972-5979. DOI: 10.1021/acs.analchem.1c00735.

(96) Murphy, A. J.; Hoang, A.; Aprico, A.; Sviridov, D.; Chin-Dusting, J. Anti-Inflammatory Functions of Apolipoprotein A-I and High-Density Lipoprotein Are Preserved in Trimeric Apolipoprotein A-I. *J. Pharmacol. Exp. Ther.* **2013**, *344* (1), 41-49. DOI: 10.1124/jpet.112.199257.

(97) Cerofolini, L.; Giuntini, S.; Carlon, A.; Ravera, E.; Calderone, V.; Fragai, M.; Parigi, G.; Luchinat, C. Characterization of PEGylated Asparaginase: New Opportunities from NMR Analysis of Large PEGylated Therapeutics. *Chem. Eur. J.* **2019**, *25* (8), 1984-1991. DOI: 10.1002/chem.201804488.

(98) Ferrige, A. G.; Seddon, M. J.; Jarvis, S.; Skilling, J.; Aplin, R. Maximum entropy deconvolution in electrospray mass spectrometry. *Rapid Commun. Mass Spectrom.* **1991**, *5* (8), 374-377. DOI: 10.1002/rcm.1290050810.

(99) Ekladious, I.; Colson, Y. L.; Grinstaff, M. W. Polymer–drug conjugate therapeutics: advances, insights and prospects. *Nature Reviews Drug Discovery* **2019**, *18* (4), 273-294. DOI: 10.1038/s41573-018-0005-0.

(100) Knol, W. C.; Pirok, B. W. J.; Peters, R. A. H. Detection challenges in quantitative polymer analysis by liquid chromatography. *J. Sep. Sci.* **2021**, *44* (1), 63-87. DOI: 10.1002/jssc.202000768.

(101) Jarrold, M. F. Applications of Charge Detection Mass Spectrometry in Molecular Biology and Biotechnology. *Chem. Rev.* **2022**, *122* (8), 7415-7441. DOI: 10.1021/acs.chemrev.1c00377.

(102) De Bruycker, K.; Welle, A.; Hirth, S.; Blanksby, S. J.; Barner-Kowollik, C. Mass spectrometry as a tool to advance polymer science. *Nature Reviews Chemistry* **2020**, *4* (5), 257-268. DOI: 10.1038/s41570-020-0168-1.

(103) Scalf, M.; Westphall, M. S.; Smith, L. M. Charge Reduction Electrospray Mass Spectrometry. *Anal. Chem.* **2000**, *72* (1), 52-60. DOI: 10.1021/ac990878c.

(104) Bagal, D.; Zhang, H.; Schnier, P. D. Gas-Phase Proton-Transfer Chemistry Coupled with TOF Mass Spectrometry and Ion Mobility-MS for the Facile Analysis of Poly(ethylene glycols) and PEGylated Polypeptide Conjugates. *Anal. Chem.* **2008**, *80* (7), 2408-2418. DOI: 10.1021/ac7020163.

(105) Huang, L.; Gough, P. C.; DeFelippis, M. R. Characterization of Poly(ethylene glycol) and PEGylated Products by LC/MS with Postcolumn Addition of Amines. *Anal. Chem.* **2009**, *81* (2), 567-577. DOI: 10.1021/ac801711u.

(106) Fouquet, T. N. J. The Kendrick analysis for polymer mass spectrometry. *J. Mass Spectrom.* **2019**, *54* (12), 933-947. DOI: 10.1002/jms.4480.

(107) Swansiger, A. K.; Marty, M. T.; Prell, J. S. Fourier-Transform Approach for Reconstructing Macromolecular Mass Defect Profiles. *J. Am. Soc. Mass Spectrom.* **2021**, *33* (1), 172-180. DOI: 10.1021/jasms.1c00317.

(108) Yang, S. H.; Chen, B.; Wang, J.; Zhang, K. Characterization of High Molecular Weight Multi-Arm Functionalized PEG–Maleimide for Protein Conjugation by Charge-Reduction Mass Spectrometry Coupled to Two-Dimensional Liquid Chromatography. *Anal. Chem.* **2020**, *92* (12), 8584-8590. DOI: 10.1021/acs.analchem.0c01567.

(109) Nasioudis, A.; Heeren, R. M. A.; van Doormalen, I.; de Wijs-Rot, N.; van den Brink, O. F. Electrospray Ionization Tandem Mass Spectrometry of Ammonium Cationized Polyethers. *J. Am. Soc. Mass Spectrom.* **2011**, *22* (5), 837-844. DOI: 10.1007/s13361-011-0101-4.

(110) Townsend, J. A.; Marty, M. T. What's the defect? Using mass defects to study oligomerization of membrane proteins and peptides in nanodiscs with native mass spectrometry. *Methods* **2023**, *218*, 1-13. DOI: 10.1016/j.ymeth.2023.07.004.

(111) Unsihuay, D.; Mesa Sanchez, D.; Laskin, J. Quantitative Mass Spectrometry Imaging of Biological Systems. *Annu. Rev. Phys. Chem.* **2021**, *72* (Volume 72, 2021), 307-329. DOI: 10.1146/annurev-physchem-061020-053416.

(112) Swales, J. G.; Hamm, G.; Clench, M. R.; Goodwin, R. J. A. Mass spectrometry imaging and its application in pharmaceutical research and development: A concise review. *Int. J. Mass Spectrom.* **2019**, *437*, 99-112. DOI: 10.1016/j.ijms.2018.02.007.

(113) Yin, R.; Kyle, J.; Burnum-Johnson, K.; Bloodsworth, K. J.; Sussel, L.; Ansong, C.; Laskin, J. High Spatial Resolution Imaging of Mouse Pancreatic Islets Using Nanospray Desorption Electrospray Ionization Mass Spectrometry. *Anal. Chem.* **2018**, *90* (11), 6548-6555. DOI: 10.1021/acs.analchem.8b00161.

(114) Dunham, S. J. B.; Ellis, J. F.; Li, B.; Sweedler, J. V. Mass Spectrometry Imaging of Complex Microbial Communities. *Acc. Chem. Res.* **2017**, *50* (1), 96-104. DOI: 10.1021/acs.accounts.6b00503.

(115) Spengler, B. Mass Spectrometry Imaging of Biomolecular Information. *Anal. Chem.* **2015**, *87* (1), 64-82. DOI: 10.1021/ac504543v.

(116) Norris, J. L.; Caprioli, R. M. Analysis of Tissue Specimens by Matrix-Assisted Laser Desorption/Ionization Imaging Mass Spectrometry in Biological and Clinical Research. *Chem. Rev.* **2013**, *113* (4), 2309-2342. DOI: 10.1021/cr3004295.

(117) Caprioli, R. M.; Farmer, T. B.; Gile, J. Molecular Imaging of Biological Samples: Localization of Peptides and Proteins Using MALDI-TOF MS. *Anal. Chem.* **1997**, *69* (23), 4751-4760. DOI: 10.1021/ac970888i.

(118) Zavalin, A.; Yang, J.; Hayden, K.; Vestal, M.; Caprioli, R. M. Tissue protein imaging at 1 μm laser spot diameter for high spatial resolution and high imaging speed using transmission geometry MALDI TOF MS. *Anal. Bioanal. Chem.* **2015**, *407* (8), 2337-2342. DOI: 10.1007/s00216-015-8532-6.

(119) Laskin, J.; Lanekoff, I. Ambient Mass Spectrometry Imaging Using Direct Liquid Extraction Techniques. *Anal. Chem.* **2016**, *88* (1), 52-73. DOI: 10.1021/acs.analchem.5b04188.

(120) Roach, P. J.; Laskin, J.; Laskin, A. Nanospray desorption electrospray ionization: an ambient method for liquid-extraction surface sampling in mass spectrometry. *Analyst* **2010**, *135* (9), 2233-2236. DOI: 10.1039/C0AN00312C.

(121) Venter, A.; Nefliu, M.; Graham Cooks, R. Ambient desorption ionization mass spectrometry. *Trends Anal. Chem.* **2008**, *27* (4), 284-290. DOI: 10.1016/j.trac.2008.01.010.

(122) Otsuka, Y.; Kamihoriuchi, B.; Takeuchi, A.; Iwata, F.; Tortorella, S.; Matsumoto, T. High-Spatial-Resolution Multimodal Imaging by Tapping-Mode Scanning Probe Electrospray Ionization with Feedback Control. *Anal. Chem.* **2021**, *93* (4), 2263-2272. DOI: 10.1021/acs.analchem.0c04144.

(123) Ma, S.; Leng, Y.; Li, X.; Meng, Y.; Yin, Z.; Hang, W. High spatial resolution mass spectrometry imaging for spatial metabolomics: Advances, challenges, and future perspectives. *Trends Anal. Chem.* **2023**, *159*, 116902. DOI: 10.1016/j.trac.2022.116902.

(124) Hale, O. J.; Cooper, H. J.; Marty, M. T. High-Throughput Deconvolution of Native Protein Mass Spectrometry Imaging Data Sets for Mass Domain Analysis. *Anal. Chem.* **2023**, *95* (37), 14009-14015. DOI: 10.1021/acs.analchem.3c02616.

(125) Meldrum, K. L.; Swansiger, A. K.; Daniels, M. M.; Hale, W. A.; Kirmiz Cody, C.; Qiu, X.; Knierman, M.; Sausen, J.; Prell, J. S. Gábor Transform-Based Signal Isolation, Rapid Deconvolution, and Quantitation of Intact Protein Ions with Mass Spectrometry. *Anal. Chem.* **2024**, *96* (23), 9512-9523. DOI: 10.1021/acs.analchem.4c00978.

(126) Källback, P.; Nilsson, A.; Shariatgorji, M.; Andrén, P. E. msIQuant – Quantitation Software for Mass Spectrometry Imaging Enabling Fast Access, Visualization, and

Analysis of Large Data Sets. *Anal. Chem.* **2016**, *88* (8), 4346-4353. DOI: 10.1021/acs.analchem.5b04603.

(127) Hu, H.; Laskin, J. Emerging Computational Methods in Mass Spectrometry Imaging. *Advanced Science* **2022**, *9* (34), 2203339. DOI: 10.1002/advs.202203339 (acccessed 2024/07/25).

(128) Hu, H.; Helminiak, D.; Yang, M.; Unsihuay, D.; Hilger, R. T.; Ye, D. H.; Laskin, J. High-Throughput Mass Spectrometry Imaging with Dynamic Sparse Sampling. *ACS Measurement Science Au* **2022**, *2* (5), 466-474. DOI: 10.1021/acsmeasuresciau.2c00031.