FOURIER TRANSFORM ANALYSIS FOR THE CHARACTERIZATION OF MASS HETEROGENEITY OF INTACT PROTEIN COMPLEXES USING NATIVE MASS SPECTROMETRY.

by

SEAN PATRICK CLEARY

A DISSERTATION

Presented to the Department of Chemistry and Biochemistry
and the Graduate School of the University of Oregon
in partial fulfillment of the requirements
for the degree of
Doctor of Chemistry

March, 2020

DISSERTATION APPROVAL PAGE

Student: Sean Patrick Cleary

Title: Fourier Transform Analysis for the Characterization of Mass Heterogeneity of Intact Protein Complexes using Native Mass Spectrometry.

This dissertation has been accepted and approved in partial fulfillment of the requirements for the Doctor of Chemistry degree in the Department of Chemistry and Biochemistry by:

Marina Guenza	Chairperson
James Prell	Advisor
Mike Harms	Core Member
Tristan Ursell	Institutional Representative

and

Kate Mondloch	Interim Vice Provost and Dean of the Graduate School

Original approval signatures are on file with the University of Oregon Graduate School.

Degree awarded March 2020

DISSERTATION ABSTRACT

Sean Patrick Cleary

Doctor of Chemistry

Department of Chemistry and Biochemistry

March 2020

Title: Fourier Transform Analysis for the Characterization of Mass Heterogeneity of Intact Protein Complexes using Native Mass Spectrometry.

Heterogeneous mass populations, or molecules that have molecular weight distributions for their mass measurement, appear in many contexts throughout chemistry, including multi-subunit protein complexes, lipid-bound membrane proteins, and polymers. The average mass and dispersity of these molecules are useful and pertinent parameters to measure and investigate, as these properties can have dramatic effects on the physical properties of the overall system. Mass spectrometry has emerged as a powerful tool to probe these parameters, but conventional mass spectrometry can be problematic for heterogeneous biological samples, as the experiment requires transfer to the gas phase and can sometimes require harsh ionization conditions. Native mass spectrometry can overcome these limitations, in that it can maintain the native stoichiometry and structure of biomolecular complexes into the gas phase, but as the dispersity and size of these molecules increases, it can become increasingly difficult to measure the average mass and dispersity due to mass spectral congestion. This congestion of peaks can often obfuscate determination of charge state, total mass, or subunit mass using conventional mass spectrometry analysis methods.

Here, research is presented dedicated to the development of a Fourier transform-based method that can be used to deconvolve highly congested mass spectra for a variety of different heterogeneous mass populations. The method is parameter-free and requires no initial guesses of charge states, total mass, or subunit mass, thus giving it a unique advantage over other established techniques. First, a 1-dimensional Fourier analysis is introduced that can probe the subunit mass, charge states, and subunit mass dispersity for a variety of different molecules. The method is further advanced by discussing the advantages of using higher harmonic frequencies in the Fourier spectrum, particularly for mass spectra with low signal-to-noise and poor resolution. A short-time Fourier transform-based method is then introduced, and is demonstrated to be useful for extracting signal from native-like protein ions even in the presence of a large salt-cluster background. Finally, the theoretical and practical implications for investigating mass populations with two or more different subunits is explored. This dissertation includes previously published co-authored material.

CURRICULUM VITAE

NAME OF AUTHOR: Sean Patrick Cleary

GRADUATE AND UNDERGRADUATE SCHOOLS ATTENDED:

University of Oregon, Eugene
San Francisco State University

DEGREES AWARDED:

Doctor of Chemistry, 2020, University of Oregon
Bachelor of Science, Chemistry, 2014, San Francisco State University

AREAS OF SPECIAL INTEREST:

Native Mass Spectrometry
Signal Processing
Membrane Lipid-Protein Chemistry
Python Data Analysis

PROFESSIONAL EXPERIENCE:

Graduate Teaching Fellow, University of Oregon, 2014-present

GRANTS, AWARDS, AND HONORS:

Graduate Teaching Fellowship, Chemistry, 2014 to present

PUBLICATIONS:

Cleary, S. P. and Prell, J. S. "Liberating Native Mass Spectrometry from Dependence on Volatile Salt Buffers by Use of Gábor Transform" *ChemPhysChem* **2019,** *20*, 519-523.

Cleary, S. P., et al. "Extracting Charge and Mass Information from Highly Congested Mass Spectra Using Fourier-Domain Harmonics" *J. Am. Soc. Mass Spectrom.* **2018,** *29*, 2067-2080.

Cleary, S. P., Thompson, A. M. and Prell, J. S. "Fourier Analysis Method for Analyzing Highly Congested Mass Spectra of Ion Populations with Repeated Subunits" *Anal. Chem.* **2016,** *88,* 6205-6213.

ACKNOWLEDGMENTS

I would like to express my upmost gratitude to Professor James S. Prell, my advisor, whose patience and countless white board discussions undoubtedly made this research possible, and for his assistance in the preparation of this manuscript. In addition, I would like to thank my committee members, Professors Marina Guenza, Mike Harms, and Tristan Ursell, for constantly challenging me though insightful questions and input at committee meetings. I would like to thank all the members of the Prell lab, past and current, for their useful input on my research, as well as the Wong lab for polite conversations in the office. This research was supported in part by the National Institutes of Health under Award Number R21AI125804, and by the National Science Foundation (award number 1752994). The content is solely the responsibility of the author and does not necessarily represent the official views of the National Science Foundation or the National Institutes of Health.

I dedicate this to my grandparents.

To Shelia Perry, who was my biggest supporter, and the first person I called the day I got into grad school.

To Fredrick Karl (with a K) Perry, who is undoubtedly the source for my fascination with science and engineering.

To Christopher Paul Cleary, the 1st, whose bravery in moving his family to different country for better opportunities lead to me being where I am today.

And to Evelyn Cleary, for being the role model of kindness, loyalty, and dedication that I constantly strive to be.

P.S. Don’t worry Grandpa Perry, my truck is fine and it’s still raining here in Oregon.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Table Page

LIST OF SCHEMES

Scheme Page

CHAPTER I

INTRODUCTION

Heterogeneous mass populations, or molecules whose mass measurement is a molecular weight distribution, appear in many contexts throughout chemistry and biochemisty. The variety of different molecules that fit this definition are numerous and ubiquitous, and can include lipid-protein complexes,[1, 2] polymers,[3-6] biotherapeutics,[7, 8] and many more. Properties such as the average mass or the dispersity of the masses are important properties to investigate for these molecules, as they can have dramatic effects on the physical properties of the overall system. For example, the dispersity of masses for a polymer population can dramatically affect its glass transition temperature, viscosity and strength, and resistance and wear.[9] Furthermore, these characteristics are not restricted to molecules made up of only covalent interactions. For example, it is theorized that the lipid environment surrounding transmembrane proteins can play an important role in the overall function of the protein or protein complex.[10-12] Here, the complex of the protein and the lipids immediately surrounding the protein can be treated as a heterogeneous mass population. Because these interactions are difficult to study and are poorly understood, the development of novel methods is necessary to begin to fully comprehend the properties of cellular membranes and other likewise systems.[13, 14] The main focus of this dissertation will be dedicated to the development of one such method, but before discussing this method, it is important to first discuss and introduce the

advantages and disadvantageous of current methods for studying heterogeneous mass populations.

Methods to Study the Average Mass and Dispersity of a Heterogeneous Mass Population.

While currently there are many established methods for studying the properties of heterogeneous mass populations, the three most popular techniques include size exclusion chromatography,[15] nuclear magnetic resonance,[16] and mass spectrometry.[17] While a short list will be discussed here, it should be noted that this is not a complete list, but rather a short list of the more popular techniques in academia and industry.

Size Exclusion Chromatography. Size exclusion chromatography (SEC) is a chromatographic method that separates molecules based on their size and shape, and is currently perhaps the most popular industrial method for measuring the average size and dispersity of polymers. A SEC experiment is a liquid chromatographic method, meaning the analyte of interest is first solubilized in a liquid mobile phase and is then passed by a stationary phase in a column, which is typically nonreactive porous beads for SEC. Smaller molecules get trapped in the pores of the beads while larger molecules simply pass by, resulting in larger molecules being eluted first. A calibration curve of analytes of known size can then be made and used to measure the size of an unknown analyte. The molecular weight and dispersity of molecule are typically estimated based on this size measurement. As an example of data obtained by an SEC experiment, shown in Figure 1 is a calibration curve of known analytes (black trace), as well as an analyte of unknown size measured with an internal calibration curve (red trace). As analytes are eluted from the column, they are detected by some physical property, which is UV absorbance for

Figure 1. Each peak (labeled with letters A through E) in Figure 1 represents a different analyte which decreases in size going from left to right.

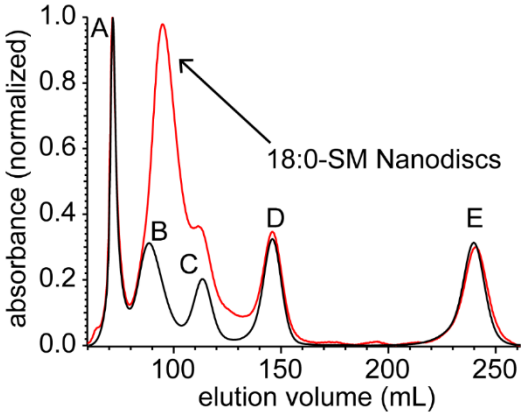


Figure 1. Size exclusion chromatography of gel filtration standards (A = bovine thyroglobulin, B = bovine γ-globulin, C = chicken ovalbumin, D = equine myoglobin, E = vitamin B_{12}) without (black) and with (red) addition of 18:0-SM Nanodiscs.

The advantages of using SEC when compared with other techniques is that the instrumentation required to perform an SEC experiment is often simpler and less expensive than other established methods. Furthermore, there can be short and well-defined separation times, which can lead to good sensitivity, and the experiment can be set up such that no sample is loss, as the analytes can be collected and separated at the end of the experiment. SEC is also particularly popular for biological molecules, as the liquid mobile phase can often be a biologically relevant buffer. For studying heterogeneous mass populations, both the average mass and the dispersity of the mass can be estimated by using the top and the width of the peak respectively. However, it is important to note that one of the major disadvantages in using SEC to measure the average mass and dispersity is that the mass is not actually being measured. Instead, SEC measures the size of the molecule, and uses this measurement to estimate the mass. Furthermore, theoretically, SEC measures an estimated hydrodynamic radius, meaning one of the fundamental assumptions of SEC is that all of molecules being measured are spherical, which can be a problematic assumption for many different systems. Finally, in comparison with the other methods discussed here, SEC is typically limited in resolution

to separate finer details of heterogeneous mass populations, as the distribution often appears as one solid peak rather than a distribution of peaks for all of the masses present in the spectrum. Thus, while SEC can be useful for estimating the average mass and dispersity of heterogeneous mass populations, other popular techniques can often produce more accurate and insightful results.

Nuclear Magnetic Resonance. Nuclear Magnetic Resonance (NMR) is another popular technique for measuring the average mass of a heterogeneous mass population. NMR is a spectroscopic technique that works by aligning the nuclear magnetic spins of molecules with a strong magnetic field, and measuring the energy needed to flip the alignment of spin using radio waves. This can give descriptive details about the chemical composition of molecules, such as the location of different functional groups within the molecule, as the atoms of different functional groups require a differing amount of energy to flip the nuclear magnetic spin. NMR is another popular technique for polymer analysis, as a technique known as head group analysis can be used to measure the average mass of the molecule. As an example how this is typically done, shown in Figure 2 is a ^{13}C NMR spectrum of a polyethylene glycol (PEG) 10k molecule in DMSO. ^{13}C NMR indicates how many different carbons are contained within the molecule, and in the case of the PEG polymer, there are three different types: Two head group carbons, shown as "a" and "b" in Figure 2, and the chain carbons, which composes all of carbons contained within the chain of the polymer and is shown as "c" in Figure 2,. These appear as three distinct peaks in the NMR spectrum, along with a fourth peak for the solvent (see Figure 2). Integration of the peaks is related to the average amount of each carbon present in the molecule. Therefore, the average number of subunits contained within the chain be

measured using the integration of the three peaks in the NMR spectrum, which can then be used to measure the average mass.

One major advantage of using NMR is that finer detail about a heterogeneous mass population can be learned from analyzing the spectrum. For example, molecules that have dispersity from two differing subunits, such as copolymers, can also be analyzed by NMR using the head group analysis described above. Using this analysis, not only can the average mass be measured, but the relative amount of each subunit can also be determine. This type of analysis is not feasible with a method like SEC. However, one major disadvantage of using NMR is that the sensitivity of the instrument is often lower than other methods, and therefore, a higher concentration of analyte is often needed to obtain reasonable data. This can become problematic for biological samples, as the required concentration can sometimes approach the millimolar range, a range that is often infeasible for many protein samples to remain solubilized in solution. Furthermore, NMR measures the bulk average of that analyte contained within the sample. This can also be problematic if the sample is not highly purified, as is the case for the polymer in Figure 2. As will be discussed below, the average molecular weight of the fully formed polymer was found to be closer to 13 kDa rather than the 10 kDa measurement found by NMR. This discrepancy is attributed to smaller molecular

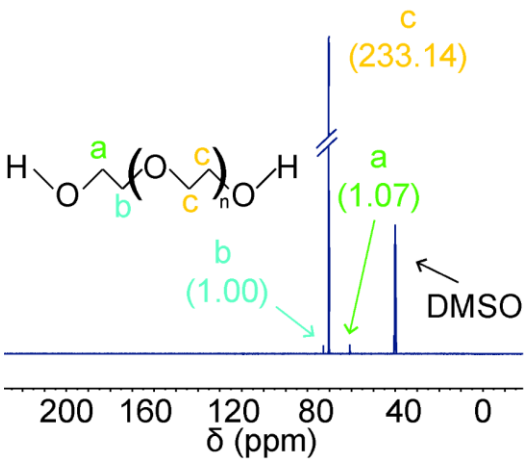


Figure 2. ^{13}C-NMR spectrum of the PEG 10k polymer sample in dimethyl sulfoxide (DMSO). Colored text indicates chemical structure assignment of observed peaks. Numbers in parentheses indicate the integrated area for each peak found, relative to peak b.

weight contaminates found in the sample, which are taken into account when average mass measurement is made by NMR. Thus, if the sample is not highly purified, the average mass can be skewed toward an inaccurate measurement.

Mass Spectrometry Mass spectrometry is another popular technique for measuring the average mass and dispersity of heterogeneous mass populations. Mass spectrometry works by first charging the molecule through some form of ionization, and measuring the ions mass to charge ratio using a number of different mass analyzers. For example, a popular mass analyzer is the Time-of-Flight mass analyzer, which measures how long it takes a charged particle to transverse a drift cell under vacuum. This time measurement is associated with two physical properties: The charge of the ion, where the higher the charge, the faster it will make it through the cell. And the weight of the ion, where the higher the molecular weight, the slower it will make it through the drift cell. Other versions of mass analyzers exist, but the common result is a spectrum the reports an ion's mass/charge ratio. This means that unless the amount charge on the ion (or, as it will be referred to here, the charge state) is 1, then knowledge of the amount of charge is needed prior to performing a mass measurement.

Mass spectrometry has some distinct advantages for measuring the average mass and dispersity of heterogeneous mass populations, the first of which is, it is the only technique listed here that actually measures the mass. Furthermore, the resolution of a mass spectrum is typically much higher than SEC, in that each mass associated with a disperse complex (i.e. PEG with 100 subunits, PEG with 101 subunits, ect…) appear as individual peaks in the mass spectrum rather than the one large peak often seen with SEC. This is shown in Figure 3, which is the mass spectrum of the PEG 10 kDa polymer

in Figure 2. Furthermore, because the ions are plotted as function of the mass to charge ratio, mass spectrometry can overcome the bulk average limitation of NMR, as lower molecular weight contaminates often have different *m/z* values than the fully formed polymer. Finally, mass spectrometry is much more sensitive than either SEC or NMR, where it is often feasible to obtain quality spectra using nanomolar concentrations of analytes. One disadvantage however, particularly from the perspective of studying biological samples such as the lipid-protein complexes, is that mass spectrometry requires transfer to the gas phase to perform the analysis. For a lipid-protein complex, this can certainly seem problematic, as the aqueous environment is often needed for self-assembly of the lipid bilayer.

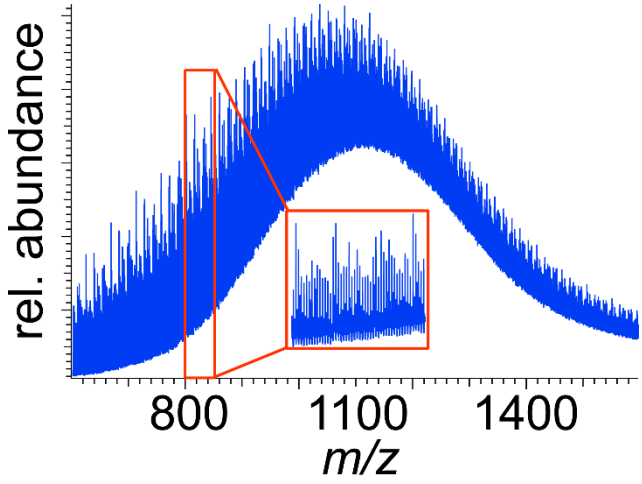


Figure 3. Mass spectrum of a polyethylene glycol Polymer. Red inset shows *m/z* ranges from 800 to 850 *m/z*. The manufacture reported average mass for this polymer is 10 kDa.

Fortunately, a recent development in mass spectrometry has led to the ability to maintain non-covalently bound complexes into the gas phase, including lipid-protein complexes. This development is a technique known as native mass spectrometry. As will be discussed below, because of the ability to measure the mass and maintain non-covalent interactions, native mass spectrometry will be the main analytical method used to investigate the average mass and dispersity of heterogeneous mass populations throughout this dissertation.

Native Mass Spectrometry

Native mass spectrometry is an analytical technique that attempts to maintain and preserve the native structure and stoichiometry of biomolecules into the gas phase. The advantages of using this technique are numerous, including the advantages already discussed for mass spectrometry, but with the added benefit of being able to quickly determine the native stoichiometry of biomolecules composed of primarily non-covalent interactions. As an example, anthrax lethal toxin is a multi-subunit pore forming toxin that was recently analyzed by native mass spectrometry. The protein complex is a decamer composed of two different protein subunits; lethal factor (LF), which composes three of the subunits, and protective antigen (PA) which composes seven. This structure was determined in 2015 by cryo-EM,[18] but it should be noted that this is a membrane protein, and it is rare to achieve such a high resolution cryo-EM for a lipid-protein complex. As is shown in Figure 4, the same stoichiometry can be determined using native mass spectrometry. The protocol for determining this is a typical workflow for native mass spectrometry, in that the mass of the intact protein complex is first measured, followed by a mass measurement of the individual components (i.e. the mass of PA and LF of anthrax). Using the mass of the entire complex and the mass of the

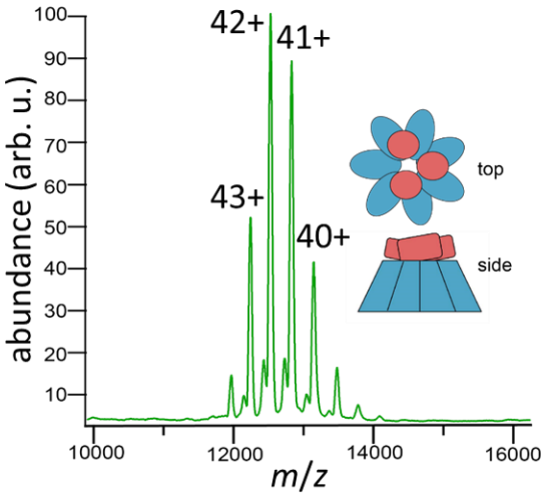


Figure 4. Mass spectrum of fully formed anthrax lethal toxin. Anthrax lethal toxin is made of 10 subunits, 7 of which are Protective Antigen (blue subunit) and 3 of which are Lethal Factor (red subunit). The mass spectrum above can be used to determine this native stoichiometry by using the mass of the fully formed intact complex.

individual components, the stoichiometry of the individual components can then be determined, which when performed for the spectrum in Figure 4, results in same ratio found by cryo EM, 7 PA's and 3 LF's. While this workflow is typical, it is important to note that there are some key experimental parameters that need to be carefully controlled to run this experiment, which will be discussed below.

Electrospray Ionization Electrospray ionization (ESI) is the ionization source typically required for native mass spectrometry. This is because ESI is known as a "soft" ionization source, in that it does not require fragmenting the molecule during the ionization process, which is preferable for maintaining non-covalent interactions of biological molecules. In order to understand how ESI is able to maintain non-covalent interactions into the gas phase, it is important to discuss some of the mechanistic details of the ionization process. The analyte is first solubilized (in the case of a biological molecule, the solution is typically a native-like buffer) and placed into a glass capillary. The solution is then charged, typically using a platinum wire, and a potential difference is applied between the capillary and the front of the mass spectrometer. This causes the charged solution to be pulled out from the capillary in the form of charged droplets, and if the concentration is held sufficiently low, a small number of these droplets will contain the analyte of interest.

As the droplet makes its way to the front of the instrument, desolvation of the droplet occurs, and charges on the outside of the droplet move closer. This creates a columbic potential between like charges that will eventually overcome the surface tension of the droplet. When this occurs, the droplet breaks apart into smaller droplets through columbic fission. After this occurs, the analyte of interest is now contained in

one of the smaller droplets, and the process is repeated until the droplet is fully desolvated. At this point, the charges are placed on the analyte of interest, and if the experiment is set up correctly, the fully intact ionized analyte will make its way into the instrument. Since mass spectrometry requires the instrument to be held under vacuum, this can aid in maintaining the native structure and stoichiometry of analytes, as the vacuum conditions keep the analyte kinetic trapped in a native like state (i.e. the interactions holding the complex together are much stronger under vacuum than in solution). It is also important to note that the amount of charge deposited onto an analyte during the ionization process can vary. Therefore, an ESI spectrum will typically have multiple peaks for the same analyte, where each peak in the spectrum is the mass of the analyte divided by a different amount of charge. Thus, in order to measure the mass of an analyte from an ESI spectrum, knowledge of the charge state for a peak is needed.

Ion Mobility Separation. While from a theoretical standpoint, the idea of trapping protein complexes in a native-like structure through ESI makes sense, it is important to mention that this has also been shown experimentally to be the case. Ion mobility separation (IMS) is additional experiment that can be performed in the mass spectrometer prior to the transfer of the ion to the detector. Importantly, IMS can measure the shape of the ion in the gas phase. The technique works by first placing the ions into a drift cell while a neutral buffer gas is being flowed the opposite direction. The time it takes to transverse the drift cell is related to two different physical properties of the ions: The charge of the molecule, where the more charge an ion has, the faster it will move through the cell. And importantly, the size of the molecule, where the larger a molecule is, the slower it will make it through the drift cell, resulting from more

collisions with the neutral buffer gas. When measured against a calibration curve of analytes of known size, IMS can measure an ion's rotationally averaged collisional cross section (CCS), a parameter that is closely related to its overall size. This CCS measurement can also be measured in the solution or estimated through computation, and when compared with measurements taken by IMS, it has been found the collisional cross sectional measurements typically agree within 3% of their solution phase values.[19]

This sets up native mass spectrometry as technique with a large amount of potential to analyze highly heterogeneous mass populations, all while overcoming the limitations associated with SEC, NMR and conventional mass spectrometry. Unfortunately, while the technique can indeed perform the analysis, native mass spectrometry can often suffer from a data analysis flaw, which will be described below.

Native ESI Mass Spectrometry Produces Complex Data Sets for Heterogeneous Mass Populations

While native mass spectrometry has been shown to be a powerful analytical method for a variety of different protein complexes,[20-23] it can be difficult to use for highly heterogeneous mass populations such as polymers and protein-lipid complexes. The origin of this difficulty comes from the interpretation of the data. As with other mass spectrometry methods, a native mass spectrum can report every mass associated with a heterogeneous mass population, but determining these masses can be time consuming process, as a heterogeneous mass population can often produce 10's to even hundreds of different mass peaks in the spectrum. Combining this observation with the necessity of using ESI means there will also be multiple charge states present in the mass spectrum.

This results in a spectrum where each charge state can contain its own distribution of masses. To put it another way, the ESI mass spectrum of a highly heterogeneous mass population is often found to be a distribution of distributions. To further this difficulty, assigning masses for all of these peaks can often be infeasible if the average mass changes between different charge state distributions. If the average mass changes between different charge states, then overlap can occur for adjacent charge state distributions. This can result in a superposition of many peaks, an observation that has been seen with protein-lipid complexes[1, 24, 25] and is seen with the polymer mass spectrum in Figure 3. This makes it difficult to uniquely identify a peak with a specific mass.

With this in mind, it should not be understated that the information desired of heterogeneous mass populations is still present in the mass spectrum. Therefore, native mass spectrometry still has the potential to identify the different properties of heterogeneous mass populations if methods can be developed that could simplify the analysis of the spectrum. This idea has led to a surge in research dedicated to this very concept, including a method that will be the main focus of this dissertation. Before discussing this however, it is important to first describe the current state methods available to deconvolve native mass spectra.

Methods for Reducing the Complexity of Native Mass Spectrometry Data

Both experimental and data analysis based methods have been developed to simplify and deconvolve the mass spectra of heterogeneous mass populations. While not a complete list, described below is list of some of the more popular methods.

Ion Mobility Separation. As was described above as a means to measure the collisional cross section of ions in the gas phase, IMS can also separate out and simplify overlapping charge state distributions in the mass spectrum. One of the main challenges in analyzing the mass spectrum of heterogeneous mass populations is the potential of multiple charge state distributions overlapping each other and forming a superposition of peaks, making it difficult to uniquely identify each peak with a particular mass. IMS can overcome this limitation, as the process of measuring the drift time (i.e. the time it takes for an ion to transverse the IMS cell, see above section) can be plotted against the mass spectrum to produce a 2-dimensional spectrum which can separate two signals overlapping in the mass spectrum. This has been demonstrated with reasonable success for polymers, as was demonstrated by Trimpin and coworkers,[5, 6] and an example this type of spectrum is shown in Figure 5 for the PEG 10 kDa polymer in Figures 2 and 3.

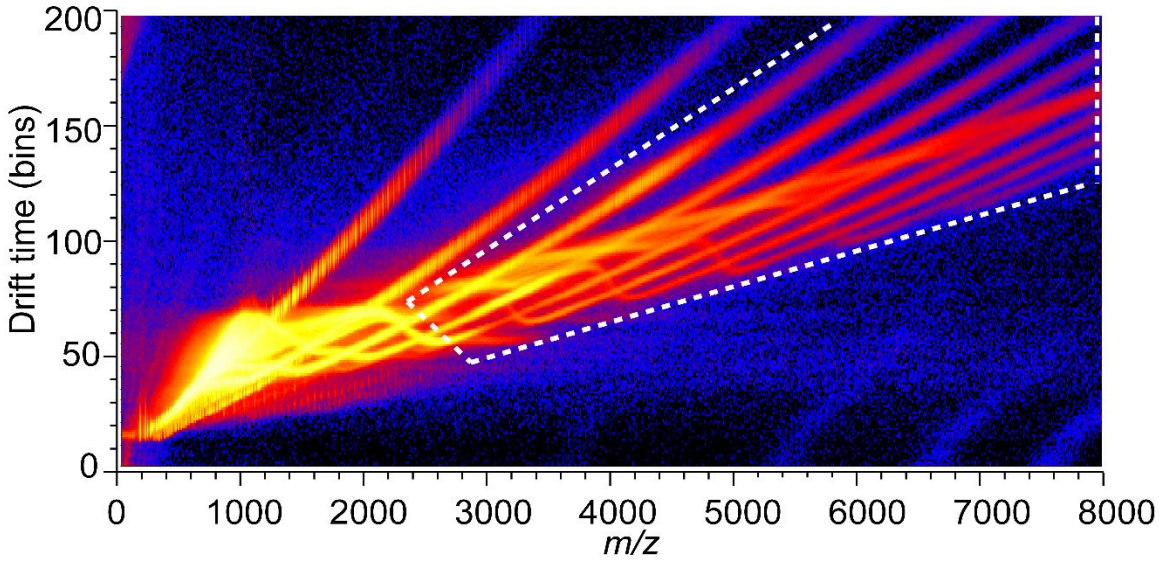


Figure 5. Ion mobility spectrum of a polyethylene glycol polymer. The mass spectrum (x-axis) is plotted against the drift time of (y-axis). The white dotted line demonstrates segregated signals that have been separated by ion mobility when previously overlapped in the mass spectrum. The average molecular weight of the polymer is reported as 10 kDa

Signals captured in this spectrum can also be isolated and plotted back into the mass spectrum, thus allowing the user to separate out charge state distributions into their own mass spectrum. However, while IMS can help separate out overlapping distributions, the technique can still leave the time consuming process of manually identifying 10's to 100's of unidentified mass peaks. Because of this, the other three methods discussed here focus more on the development of different algorithms to automatically assign peaks, overcoming this limitation associated with IMS and traditional mass spectrometry techniques.

Maximum Entropy. Maximum entropy (MaxEnt) is an algorithm that was introduced in 1989,[26] and has since gone through many iterations and created additional programs that are based on the same underlying method. MaxEnt has been used by many laboratories in both academia and industry, is often considered the gold-standard for mass spectral data deconvolution. While many different versions of MaxEnt exist, most of the versions work on the same basic premise, in that the algorithm iteratively searches for a charge state and mass assignment that maximizes an entropy parameter based on deviations from predicted mass spectral peak positions. Importantly, one of the fundamental assumptions of MaxEnt is that each charge state has identical mass or mass dispersity. For heterogeneous mass populations, this assumption can often be problematic, as the average mass can be different for the charge state distributions present in a spectrum. In cases such as these, this can lead to a failed convergence or incorrect charge state and mass assignments. While the method can be avoided if this information is known prior to analysis, one of the main difficulties with MaxEnt and other fitting algorithms is being able to discern this knowledge if the spectrum is highly congested.

Thusly, users of these algorithms are advised caution prior to performing the analysis to not over interpret results.

Bayesian Deconvolution. Bayesian deconvolution is a newer data deconvolution technique that was introduced as "UniDec" by Marty and coworkers in 2015.[14] UniDec is another fitting algorithm, but differs from MaxEnt by using a Bayesian statistical approach rather than maximum entropy. The underlying algorithm works by fitting delta functions to the mass spectrum based on Bayesian priors, such as ranges for mass and charge states, and finding the most likely fit by iteratively fitting the modeled data sets until a convergence is found. Importantly, because a range of masses can be used as a prior, UniDec does not suffer from the assumption that every charge state has to contain the same mass or mass distribution. This advantage has allowed UniDec to be successful in analyzing mass spectra with charge state distributions that change in the average mass, such as lipid-protein complexes. However, because UniDec is based off of Bayesian statistics, in order for the algorithm to be successful, knowledge of the sample and spectrum is required prior to performing the analysis. As is the case with MaxEnt, one of the major drawbacks of using fitting algorithms to analyze heterogeneous mass populations is that this information can sometimes be difficult to know prior to using the algorithm, in particular for spectra that are highly congested. This was demonstrated in a publication in 2018 using model data,[25] where lacking this information produced a result that do not resemble the known data. Indeed, as is the case with other fitting algorithms, the user is advised to not over interpret results found with the program, as inputting the incorrect Bayesian priors has the potential to produce incorrect results that can seem reasonable to a naïve user.

Fourier Transformation. Fourier transformation (FT) is another deconvolution method that has been shown to be effective at analyzing the mass spectra of heterogeneous mass populations. Rather than fitting data based on a statistical parameter, the FT method instead treats the mass shift associated with the addition of subunit (i.e. PEG with 100 subunits versus 101 subunits) as a frequency in the mass spectrum. This frequency is related to both the subunit mass and the charge state associated with the spacing, and can be probed by Fourier transforming the mass spectrum. The result of this in another spectrum with an equally spaced series of peaks, where each peak corresponds to a charge state present in the mass spectrum divided by the shift in mass (or subunit mass). This method was introduced in 2004 by Cook and coworkers,[4] where it was shown to have success at analyzing polymer mass spectra.

When compared with other deconvolution methods, the main advantage of the FT method is that no assumptions or parameters are needed to perform the analysis. Because of the linearity of FT, the data that is being analyzed in the FT spectrum is simply the original data in the mass spectrum plotted in a different space. This allows FT to avoid the issue of over interpreting fitted data sets associated with other fitting algorithms. As for analyzing heterogeneous mass populations, a benefit of using the FT method is that it can often simplify the highly complex data seen with the mass spectra of these systems. As was mentioned above, one of the main difficulties in interpreting the mass spectra of a heterogeneous mass population is the time consuming process of uniquely identifying every peak associated with different mass. FT takes all of peaks, and sorts each peak according to its respective charge state, often resulting in turning 100's of peaks into only

a select few. This is even the case when there is a high amount of overlap between adjacent charge state distributions.

Until recently, the main limitation with FT was the potential of higher harmonic overlap. As an example to describe this limitation, unique identification of the charge states present in a polymer mass spectra from the original publication in 2004 was difficult due to higher harmonic overlap in the frequency spectrum. This is because peaks that appear in the Fourier spectrum have the potential of producing higher harmonic frequency peaks, particularly if the mass spectrum is well resolved. This means that if a peak appears at the 3+ charge state in the FT spectrum, another peak may appear at that 6+, 9+ and any other multiple of the 3+ frequency. This concept is discussed in more detail in Chapter 3, but the result of this is it becomes difficult to uniquely say if the mass spectrum has a 3+ charge state, a 6+ charge state, or both. Thus, unique identification of charge states presents a challenge to using the FT method for highly heterogeneous mass populations.

That being said, if the harmonic overlap issue could be resolved, the FT method appears to have a lot of potential for analyzing heterogeneous mass populations. As it will be shown in this dissertation, it can indeed be overcome, and over the next few chapters, I will demonstrate research that I have conducted to expand and understand the theoretical implications of the Fourier transform method. In Chapter 2, I will set the theoretical groundwork for a typical 1d Fourier transformation, and show that the method can measure the subunit mass, charge states, and mass dispersity for multiple highly heterogeneous mass populations. This chapter contains co-authored material from James S. Prell and Avery M. Thompson. I further the theoretical groundwork in Chapter 3 by

demonstrating all of the additional information that can be learned from a Fourier analysis, including how this new information can useful for other fitting algorithms, such as the Bayesian analysis described above. This chapter includes co-authored material from James S. Prell, Huilin Li, Dhanashri Bagal, Joseph A. Loo, and Iain D. G. Campuzano. In Chapter 4, I will introduce a novel 2 dimensional Fourier analysis approach, which can overcome the limitation of higher harmonic overlap, and allow researchers to study biological molecules from more atypical biochemical buffers. This chapter includes co-authored material from James S. Prell. Finally, in Chapter 5, I will discuss the potential of what can be learned from performing the analysis heterogeneous mass complexes with two different subunits, and how mechanistic detail can be learned about the complex from performing the analysis. This will include future co-authored material from James S. Prell.

CHAPTER II

1-DIMENSIONAL FOURIER ANALYSIS FOR DECONVOLVING MASS SPECTRA OF HETEROGENEOUS MASS POPULATIONS

Includes co-authored material from:

> Cleary, S. P., Thompson, A. M. and Prell, J. S. "Fourier Analysis Method for Analyzing Highly Congested Mass Spectra of Ion Populations with Repeated Subunits" *Anal. Chem.* **2016,** *88,* 6205-6213.

Introduction

Electrospray ionization mass spectrometry (ESI-MS) can be a powerful tool for studying large ions, including native-like protein complexes and other large assemblies.[27-30] ESI-MS is a soft ionization technique capable of maintaining native-like structures of protein complexes with many subunits, and the topology, stoichiometry, shape, and other properties of the resulting ions can in many cases be characterized using MS and/or ion mobility spectrometry (IMS).[31-34] For very large or heterogeneous native-like biomolecular complexes, accurate mass and charge determination can be very challenging owing to low practical resolution caused by solvent and/or salt adduction and a high density of mass spectral peaks.[35, 36]

As was mentioned in Chapter 1, complexes with differing numbers of repeated subunits appear in many contexts, including native membrane protein-lipid,[12, 37] chaperone-target,[38, 39] and hetero-oligomeric protein assemblies,[40, 41] as well as synthetic polymers.[42] However, ESI mass spectra of these complexes can be exceptionally challenging to analyze when the stoichiometry of the repeated subunit varies over a wide range, due to strongly overlapped distributions of peaks that represent different charge states and subunit stoichiometries, especially when individual peaks exhibit broadening

from solvent or salt adductions. These problems are exacerbated for ions with high charge states, where peak spacing may approach the practical resolution of the mass spectrum. Adducts can often be removed to some extent through collisional activation. This "collisional clean-up" method is commonly used to improve the resolution, homogeneity, and transmission efficiency of the ions.[43, 44] However, analysis of fragile native assemblies can be challenging due to unwanted dissociation of labile ligands of interest during collisional "clean-up" and transfer to the mass analyzer.[24, 44, 45]

As was mentioned in Chapter 1, deconvolution algorithms have been developed to assist in assigning complex mass spectra of these types. Maximum entropy-based algorithms iteratively search for a charge state and mass assignment that maximizes an entropy parameter based on deviations from predicted mass spectral peak positions, assuming identical mass distributions for each charge state.[26, 46] However, for many heterogeneous ion populations, mass distributions may be different for each charge state, which can lead to failed convergence or incorrect charge state and mass assignments using this method. In contrast, Bayesian statistical analysis is another deconvolution approach that can be useful with heterogeneous ion populations because different mass distributions can be modeled for different charge states, and a fit to the mass spectrum can be generated from a maximum likelihood estimation.[14, 47] Typically, a range of charge states, total mass, and/or subunit mass are input as initial parameters, potentially making this method challenging to apply if this information is not known. Furthermore, if the likelihood function has more than one local maximum, the result of the algorithm may not be unique and may in principle depend on the initial guesses of parameters. Therefore, in the case of very heterogeneous ion populations, it is desirable to develop

deconvolution algorithms that do not require initial guesses of charge state and mass distributions and allow for different mass distributions for each charge states.

Here I present a deconvolution method that is parameter-free and can be used to extract mass spectral information with little need for collisional clean-up of ions prior to mass analysis. This method is based on Fourier transform, and I describe its use in measuring the charge state distribution, accurate subunit mass, total mass, and subunit stoichiometry distribution for electrosprayed analytes containing varying numbers of a repeated subunit. The method does not require mass spectral smoothing or initial guesses for charge states, total mass, or subunit mass. As proof of principle and to demonstrate its versatility, the method is applied to ubiquitin with multiple adductions of sodium and potassium, large single and mixed polymers, and native self-assembled lipid-protein bilayers (Nanodiscs). The method is widely applicable to many different types of ions and ion assemblies containing repeated subunits.

Methods

Sample preparation. All samples were prepared using ultrapure (18.0 MΩ) water. Bovine ubiquitin (Ubq) was purchased from Sigma-Aldrich and used without further purification. Aqueous Ubq solutions were prepared containing 25 μM Ubq and 1 mM sodium chloride or potassium chloride to promote adduction of sodium and potassium to Ubq ions. Polyethylene glycol (PEG) and polyacrylic acid (PAA) were purchased from Sigma-Aldrich and used without further purification. PEG and PAA were prepared as 1 mg/mL aqueous solutions.

Nanodiscs containing palmitoyloleoylphosphatidylcholine (POPC), dioleoylphosphatidylcholine (DOPC), dimyristoylphosphatidylcholine (DMPC),

dipalmitoylphosphatidylcholine (DPPC), or 18:0-sphingomyelin (18:0-SM) were prepared according to a method adapted from that of Sligar and co-workers.[48, 49] Briefly, all lipids were purchased from Avanti Polar Lipids as 5 mg/mL solutions in chloroform, dried until opaque with dry nitrogen gas, and re-suspended to a final concentration of 50 mM in a pH 7.4 aqueous buffer containing 100 mM sodium cholate (Sigma-Aldrich), 20 mM Tris (Bio-Rad), 100 mM sodium chloride, and 0.5 mM ethylenediaminetetraacetic acid (EDTA). Membrane scaffold protein MSP1D1 (Sigma-Aldrich) was reconstituted in pH 7.4 aqueous buffer (20 mM Tris, 100 mM sodium chloride, 0.5 mM EDTA, 0.01% sodium azide) to a concentration of ~200 μM. Lipid suspensions were mixed with MSP1D1 solutions and additional buffer to a final concentration 100 μM in MSP1D1 and 6.5 mM POPC, 6.0 mM DOPC, 8.0 mM DMPC, 9.0 mM DPPC, or 7.5 mM 18:0-SM, and incubated for 1 hr at 4 °C (POPC and DOPC), room temperature (DMPC), or 37 °C (DPPC and 18:0-SM). Nanodisc self-assembly was initiated by addition of 50% volume equivalent of BioBeads SM-2 (Bio-Rad) that had been previously rinsed and sonicated three times in methanol followed by three times in the reconstitution buffer. After 4-12 hr, the Nanodisc-containing supernatants were removed from the BioBeads and buffer-exchanged into 200 mM ammonium acetate (Sigma-Aldrich) using Micro Bio-Spin 6 columns (Bio-Rad) immediately before IM-MS analysis.

Mass spectrometry. All mass spectrometry analysis was performed with a Synapt G2-S*i* ion mobility mass spectrometer (Waters Corp.) using a static nanoelectrospray ionization (nanoESI) source. NanoESI emitters were prepared by pulling borosilicate capillaries (ID 0.78 mm, Sutter Instruments) to a tip ID of ~1 μm using a Flaming-Brown P-97 micropipette puller (Sutter Instruments). For each sample,

~3-5 μL of solution was loaded into an emitter, which was placed approximately 3-5 mm from the entrance of the mass spectrometer. A platinum wire inserted into the solution was used to apply an electrical potential of 1.0-1.2 kV relative to instrumental ground to initiate electrospray. The ion source was held at a temperature of 60 °C for all experiments except those with Nanodiscs, for which it was equilibrated to ambient temperature. Mass spectra from sodiated and potassiated Ubq were collected in Resolution mode, and mass spectra for all other analytes were collected in Sensitivity mode for maximal signal-to-noise. Mass calibration was performed using the distribution of singly-charged $Cs^+(CsI)_n$ peaks in mass spectra obtained for 100 mM aqueous CsI solutions. Argon Trap gas was introduced at a flow rate of 5 mL/min in all experiments. Trap/Transfer Collision Energy settings of 20/2 V, respectively, were used for all analytes other than Nanodiscs, for which Trap/Transfer Collision Energies of 40/10, 75/50, or 175/75 were used to create "minimal", "mild", or "strong" collisional activation conditions, respectively. For all analytes other than Nanodiscs, mass spectra collected continuously for ~3 min were summed, whereas for Nanodiscs, ~10 min of continuously collected data were summed. All mass spectral data were processed using MassLynx v. 4.1 (Waters Corp.) before further analysis using a Fourier transform-based algorithm (see below).

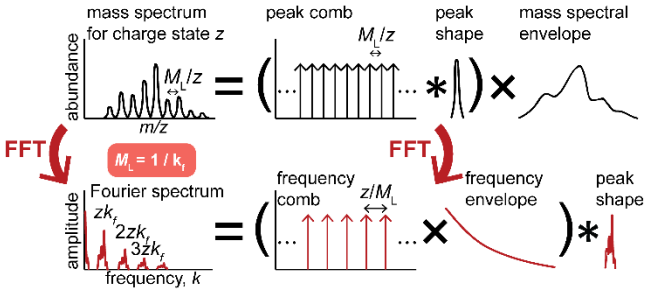
Analysis methods. A Fast Fourier Transform (FFT)-based algorithm, illustrated in Scheme 1, was used to deconvolve and analyze mass spectra. The key concept to the FFT algorithm is that the distribution of mass spectral peaks for an individual charge state z of an analyte containing various numbers n of a repeated subunit with mass M_L can be approximated as a comb of equally-spaced peaks multiplied by a mass spectral envelope

function. This envelope function describes the relative abundances of peaks with charge state z representing ions with different n. When M_L is much smaller than the total mass of the ion, the shape of each mass spectral peak with the same z but different n in the comb varies only slightly. The mass spectrum for a given z can thus be represented mathematically as:

$$s(m/z) = [c(m/z)*p(m/z)]\times f(m/z)$$

where $s(m/z)$ is the abundance of the ion with m/z, $c(m/z)$ is a comb function with comb spacing M_L/z, $p(m/z)$ is a function that approximately describes the shape of each individual peak in the comb, $f(m/z)$ is the mass spectral envelope function, * indicates convolution, and × indicates multiplication (see Scheme 1, top).

Scheme 1. Key concepts behind the Fourier transform algorithm



The FFT of $c(m/z)$ is another comb $C(k)$ of evenly spaced peaks centered about zero in the Fourier "frequency" domain, or k-domain, where k represents the periodicity of peaks in the mass spectrum and is equal to z/M_L for a given charge state z. The FFT of $s(m/z)$, which I call $S(k)$, can be written as:

$$S(k) = [C(k)\times P(k)]*F(k)$$

where P(k) is the FFT of p(m/z) and is a slowly decaying function of k, and F(k) is the FFT of f(m/z) and has a width inversely proportional to the width of f(m/z) (Scheme 1, bottom). For the purposes of the mass spectral analysis method outlined here, the important consequences of the FFT operation are: 1) the comb of mass spectral peaks associated with charge state z transforms to another comb of peaks in the k-domain, 2) the mass spectral envelope for ions with a *broad distribution* of n transforms to a *narrow peak shape* in the k-domain, 3) the shape of each k-domain peak is the IFFT of the individual m/z envelope function corresponding to its charge state, and 4) because FFT is linear, the sum of comb-shaped mass spectra for different values of z transforms to a sum of k-domain comb spectra. In particular, for ions with sequential charge states z_{min}, (z_{min}+1), (z_{min}+2), …, z_{max}, as is common for many ions produced by ESI, the FFT of the mass spectrum contains comb peaks spaced by $\Delta k = 1/M_L$. I term this value of Δk the "fundamental frequency," k_f, of the mass spectrum and note that its inverse is the average mass of the repeated subunit, M_L. Thus, the mass of the repeated subunit and the charge states represented in the mass spectrum have a straightforward mathematical relationship to the spacing and distribution of peaks in the k-domain spectrum, which is the central idea used in the detailed analysis described below.

For mass spectra of the type under discussion, basic questions that one may want to answer include: What is the mass of the repeated subunit (M_L)? What charge state distribution is present in the mass spectrum? For a given charge state, z, what is its associated mass spectral distribution, f(m/z)? Once the answers to these questions are known, the total mass distribution for charge state z can be obtained by re-scaling the m/z-axis by z, and if the "base mass" of the ion (the total mass of all substituents other

than the repeated subunit) is known, the distribution of the number of repeated subunits for charge state z can be determined from this total mass distribution. A method for determining M_L, the charge state distribution, and f(m/z) for each charge state is outlined below, including a 2-dimensional algorithm for challenging cases in which k-domain spectra are highly congested. Many computational implementations of the algorithm are possible and our implementation is described in greater detail in the Appendix A.

Determination of M_L (= 1/k_f). For a given charge state, z, k-domain peaks appear at integer multiples of zk_f, and the spacing between the first harmonic peaks for adjacent charge states is k_f. Thus, k_f can be estimated as the average spacing between the centroids of adjacent peaks in the k-domain comb. M_L is then estimated as the reciprocal of k_f. Once k_f has been determined, the charge state associated with a peak that is not an overtone peak is its frequency divided by k_f.

Determination of charge-state distribution. The k-domain comb for a given charge state includes overtone peaks at integer multiples of zk_f, thus a 1+ ion will contribute to every peak of the k-domain comb. In general, for two charge states z_1 and z_2, if z_2 is a multiple of z_1, the first comb peak in the k-domain associated with z_2 will overlap with an overtone peak associated with z_1. Conversely, all peaks with k-values that are not integer multiples of the k-value of any other peak must correspond to a charge state present in the mass spectrum. For example, if a k-domain peak at zk_f is present, and this peak does not coincide with overtone of any other peak, then charge state z must contribute to the mass spectrum. Thus, the distribution of charge states contributing to the mass spectrum can be straightforwardly obtained under these conditions, which are relatively common for electrospray mass spectra of large ions. If the k-domain spectrum

does not decay too quickly at high *k*, the assignment of charge states can be validated, for example, by verifying the presence of a similarly-shaped distribution of corresponding overtones in the *k*-domain, for example, at the second harmonic peak of each of the putative charge states (see below).

Determination of the mass spectral envelope, f(*m/z*), for a given charge state, *z*. If the *k*-domain peak at zk_f has been determined to be associated exclusively with charge state *z*, i.e., it does not coincide with an overtone peak from another charge state and does not significantly overlap with adjacent peaks, the mass spectral envelope for charge state *z* can be estimated as the inverse Fast Fourier Transform (IFFT) of the *k*-domain peak at zk_f. Here, I compute the IFFT of this peak by windowing the *k*-domain spectrum with a symmetric rectangular window from $\left(z-\frac{1}{2}\right)k_f$ to $\left(z+\frac{1}{2}\right)k_f$. Care must be used in choosing the window width, because "ringing" artifacts (due to the IFFT of the window function itself) may appear in the reconstructed f(*m/z*) if the window is too narrow, and contributions from adjacent peaks will be present if the window is too wide.

2-dimensional algorithm for challenging cases. The above procedure can be challenging to implement when *k*-domain peaks overlap significantly, which can occur when the *m/z* envelope for a particular charge state is narrower than $M_L/1$. However, if the mass spectral distributions for adjacent charge states do not completely overlap, a helpful strategy for analyzing the mass spectrum using FFT can be to multiply the mass spectrum by a window function (for example, a Gaussian) before performing the FFT. A 2-dimensional plot representing the windowed FFT of the mass spectrum as a function of the center of the window can then be used to separate out the contributions in the *k*-domain from each charge state and facilitate determination of k_f (see Appendix A).

Caution must be exercised in reconstructing f(*m/z*) for different *z* by IFFT from these 2-dimensional spectra, because the *k*-domain peaks are convolved with the FFT of the Gaussian window. If the ion population is suspected to contain oligomeric states with overlapping *k*-domain spectra, caution should be exercised in assigning charge states, and additional separation by chromatography or ion mobility may be useful.

Results and Discussion

Choice of Samples. In order to assess the strengths and limitations of the Fourier algorithm, three different types of sample were used. To demonstrate the consistency of results from the Fourier algorithm with those easily obtained from conventional mass spectral analysis, the masses of sodium and potassium adductions to Ubq (~8.5 kDa) were measured using the Fourier spectrum alone or the mass spectrum alone. Aqueous long-chained polymer ions for which the charge state envelope and subunit masses could not be straightforwardly determined from the mass spectrum were also analyzed with the Fourier algorithm, and charge state assignments were confirmed with the aid of ion mobility separation. Finally, mass spectra for ~100-200 kDa lipid-protein Nanodisc samples incorporating several different types of lipids were acquired. Nanodisc ions represent an exceptionally challenging case of heterogeneity not easily resolved by chromatography or ion mobility, and the Fourier spectra for these ions were used to assess the reproducibility of Nanodisc samples in triplicate preparations as well as the dependence of Nanodisc lipid stoichiometry on lipid identity and collisional activation conditions. A Nanodisc sample prepared using 18:0-sphingomyelin, a lipid-raft

associated lipid, was also analyzed with the FFT method to determine for the first time the native lipid stoichiometry and effective bilayer surface area of this lipid in Nanodiscs.

Sodiated and Potassiated Ubq. Sodium and potassium cations often displace protons as charge-bearing adducts and represent a simple example of a repeating subunit for aqueous biomolecular ions produced with ESI. These adductions can lead to a broad mass distribution at a given charge state for many biomolecular ions, which can be problematic in measuring accurate masses. To illustrate the effectiveness of the FFT algorithm for determining accurate masses of repeated subunits in a case where they can be directly verified from the mass spectrum, the FFT algorithm was applied to the mass spectra of extensively sodiated or potassiated Ubq with native charge states.

Mass spectra and the corresponding Fourier spectra for these Ubq samples are shown in Figure 6. A small population of ions with a single sodium adduction and varying numbers of potassium adductions was observed in the mass spectrum of potassiated Ubq. While the mass spectra are isotope-resolved and can be assigned without the aid of the FFT algorithm, mass and charge state analysis was straight forward from the Fourier spectra as well. The mass of the subunit is greater than twice the width (i.e., twice the standard deviation) of the *m/z* envelope for each charge state present in the mass spectrum. Due to the properties of Fourier transform, the *k*-domain peaks for individual charge states thus overlap and are poorly resolved. Resolution of the charge states is achieved by use of the 2-dimensional analysis described in the Methods section (and in general detail in the Appendix A). In order to determine accurate adduct masses and charge states present in the Ubq mass spectrum, the analysis used a total of 31 evenly spaced Gaussian windows in the mass spectrum with a width (2σ) of 50 *m*/*z*. Identical

parameters (windows and mass spectral range) were used for both the sodiated and potassiated Ubq spectra. The resulting 2-dimensional spectra are shown in Figure 7. Charge states for both the sodiated and potassiated Ubq spectra were found to be 4+ and 5+ for the monomer and 6+ and 7+ for the non-specific dimer of Ubq. The mass difference due to displacement of a proton by a sodium cation was found to be 22.05(2) u compared to a calculated (^{23}Na − ^{1}H) exact mass of 21.982 u. The potassium adduct was found to have a mass of 38.11(2) u, compared to a calculated exact (^{39}K − ^{1}H) mass of 38.090 u. These results correspond to relative errors of 3000 and 500 ppm for the adduct masses, respectively, or an error in the relative mass difference between adjacent sodiation or potassiation states of the Ubq ions of 8 or 1 ppm, respectively. By comparison, corresponding accurate mass differences of 21.973(3) and 37.940(3) u were measured directly from the mass spectra.

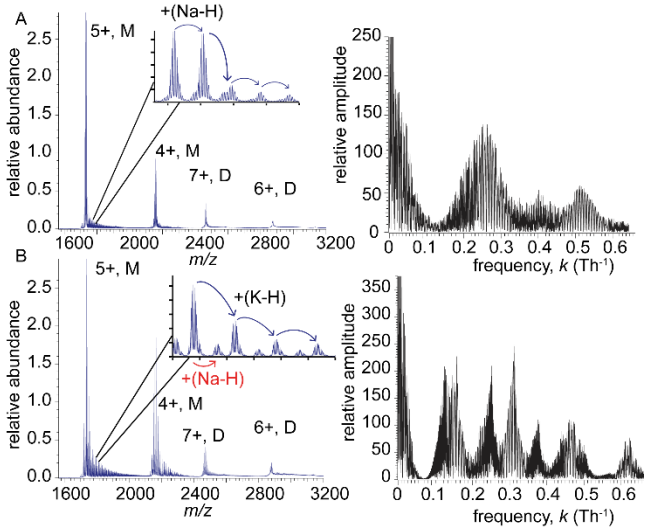


Figure 6. Native mass spectra (left) and FFT spectra (right) of sodiated (A) and potassiated (B) Ubq, including monomers ("M") and non-specific dimers ("D"). Insets in mass spectra illustrate sequential adductions of sodium or potassium ions. Arrows show sequential repeated adductions of sodium or potassium to Ubq. The high-intensity peak near zero frequency in the FFT spectra has been truncated for clarity.

PEG and PAA. Long-chain polymers are prototypical molecules with repeated covalently bound subunits that can be difficult to separate chromatographically and to analyze by electrospray ionization mass spectrometry, because polymers with many different chain lengths may exist in the same sample, and many charge states may be produced upon ionization. The resulting mass spectra can be highly congested, comprising hundreds of densely spaced peaks with no obvious abundance pattern from which charge states or masses can be straightforwardly determined without the aid of methods such as charge stripping.[42, 50] To assess the utility of the FFT algorithm in analyzing the charge states and subunit masses for ions with this type of heterogeneity, nanoESI mass spectra were acquired for aqueous PEG, PAA, and a sample containing both polymers (Figure 8). Above *m*/*z* ~500, a highly complex distribution of peaks spanning more than 1000 *m*/*z* is present for both analytes. These high-*m*/*z* distributions comprise a sequence of peaks of

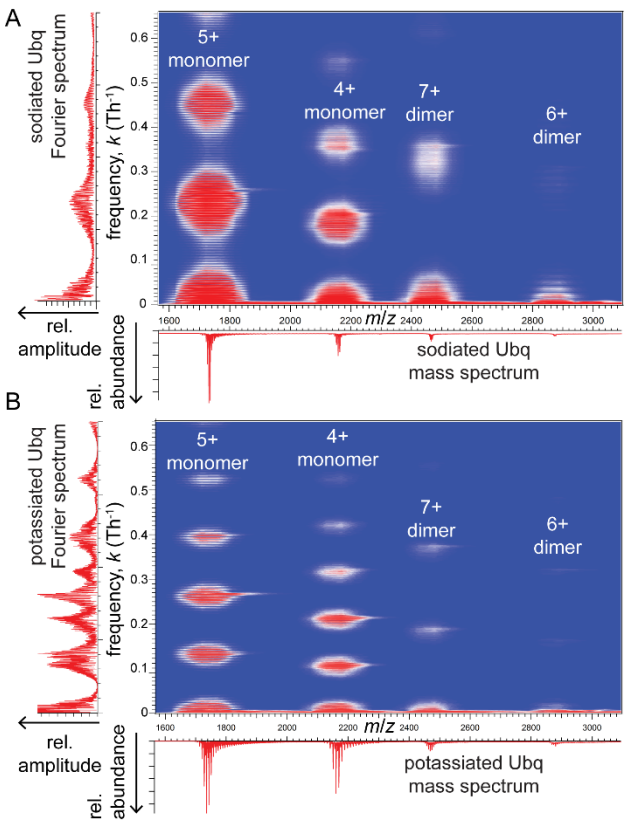


Figure 7. 2-dimensional Fourier analysis for sodiated (A) and potassiated (B) Ubq mass spectra. Bottom traces are nanoESI mass spectra, and the left traces are 1-dimensional FFT spectra. 2-dimensional plots axes correspond to the center of the mass spectrum Gaussian window (horizontal axis) and Fourier frequency (vertical axis). High-intensity peaks near zero frequency in the 1- and 2-dimensional FFT spectra have been truncated for clarity.

relatively high abundance spaced by approximately 1 *m*/*z* in addition to a plethora of more closely-spaced peaks of lower abundance. These mass spectra are similar to previously reported ESI mass spectra of aqueous long-chain PEG, for which highly heterogeneous, overlapped distributions of ions were also observed over similar ranges of *m*/*z*.[17, 50]

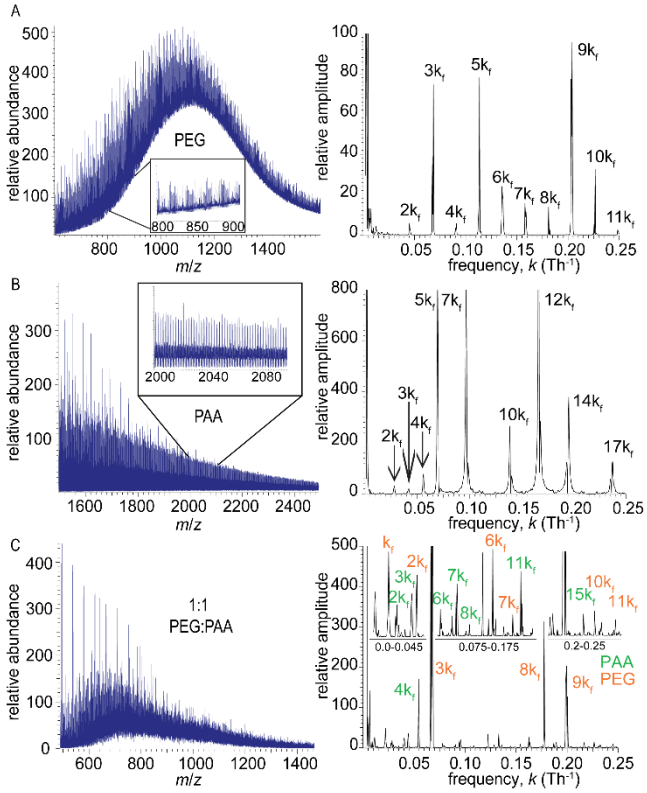


Figure 8. Mass spectra (left) and FFT spectra (right) of PEG (A), PAA (B), and a 1:1 wt:wt mixture of PEG and PAA (C) generated from aqueous solution using nanoESI. The high-intensity peak near zero frequency in the FFT spectra has been truncated for clarity.

The FFT spectra (Figure 8A and 8B) for each single-polymer analyte consist of a single comb of many sharp, evenly-spaced peaks. Fundamental frequencies, k_f, for the PEG and PAA samples were computed from the 2nd through 8th *k*-domain harmonic peaks to be $2.267(5)\times10^{-2}$ *z*/*m* and $1.387(4)\times10^{-2}$ *z*/*m*, respectively. From these values, accurate subunit masses for the analytes were readily determined to be 44.04(1) and 72.07(5) u for PEG and PAA, corresponding to errors of 450 and 650 ppm, respectively, as compared to the calculated formula weight of C_2H_4O (44.02 u) and $C_3O_2H_4$ (77.02 u). While metal cation adducts such as sodium and potassium may certainly be present in the mass spectra of these polymer samples, signal due to such adductions is not clearly

visible in the FFT spectra and is likely overwhelmed by signal from the repeated polymer subunits.

Because many sharp peaks are present in the mass spectra for both analytes, their *k*-domain spectra exhibit many harmonics with a complex structure containing information about the ions' charge states. For PEG, discernable peaks are present at every harmonic of k_f between the 0^{th} and 22^{nd}. Because the fundamental peak is very low in absolute intensity, it can be concluded that 1+ ions constitute only a very small fraction of the ion population. Conversely, ions with charge states between 2+ and 22+ that are prime numbers must be present for PEG. For PAA, discernable peaks are present at all prime-number harmonic of k_f between the 0^{th} and 22^{nd} other than the 11^{th} and 13^{th}. For both polymers, it is difficult to assess from the Fourier spectra whether ions with composite-valued charge states are present due to potential signal from overtones of prime-valued multiples of k_f. Prior chromatographic separation or ion mobility separation may facilitate assignment for composite number charge states for some polymer samples, and this harmonic overlap can be overcome through the use of a short-time Fourier Transform (see chapter 4).

Finally, a mass spectrum was acquired for a mixture of the two polymers to illustrate that the algorithm can be used to determine the subunit masses of two different polymers present in the same sample (Figure 3C). A nanoESI mass spectrum for a 1:1 wt:wt aqueous mixture of the two polymers was acquired under the same conditions as for the single-polymer experiments. There is a strong 3+ peak for PEG with high amplitude in the *k*-domain spectrum that dominates the rest of the peaks. This corresponds to the highly abundant comb of peaks visible in the mass spectrum that is

spaced by 14.67 *m*/*z*. The *k*-domain contains many peaks that are not evenly spaced, indicating the presence of at least two different repeated subunits in the ion population. Two distinct combs are present with very different relative intensities at each multiple of k_f. In particular, high-intensity peaks are observed at $3k_f$, $8k_f$, and $9k_f$ for PEG, and, by contrast, only at $4k_f$ for PAA. Interestingly, the relative *k*-domain intensities observed for the mixed-polymer sample differ substantially from those observed from either single-polymer sample, an effect likely attributed to competing ionization during the electrospray process for the mixed-polymer sample. This conclusion is borne out by the observation that the mass spectrum of the mixed-polymer sample is not proportional to the sum of the mass spectra of the single-polymer samples under the same experimental conditions.

Nanodiscs. Protein-lipid Nanodiscs are an example of large biomolecular assemblies with repeating subunits that can be difficult to analyze by ESI mass spectrometry. Briefly, Nanodiscs are discoidal non-covalent biomolecular assemblies a few nanometers across, consisting of a phospholipid bilayer surrounded by two amphipathic helical membrane scaffold proteins (MSPs). For the MSPs used here (MSP1D1), native-like Nanodisc ions typically have masses near 150 kDa, as found using other deconvolution algorithms.[14, 24, 51] Their native mass spectra typically exhibit a high degree of overlap between the heterogeneous mass distributions for each charge state distributions contained in the ion population. For a given native-like charge state, the distribution in the number of lipids has been reported to be roughly ±5 for DMPC, POPC, DOPC, and DPPC.[24, 51] The Fourier algorithm was applied to mass spectra of Nanodisc ions containing POPC, DOPC, DMPC, DPPC, or 18:0-SM, and the mass of

each lipid, the charge state distribution for the ions, and the mass distribution and lipid stoichiometry distribution for each charge state were determined. Three separate preparations of DMPC Nanodics were analyzed to test the reproducibility of the results. Much of the large "baseline" in these spectra is attributed to the resolving power of the mass spectrometer i.e., it arises in large part from the summed "tails" of the mass spectral peaks of ions at many different closely spaced *m*/*z* values. This "baseline" contains signal from the analytes, and the FFT algorithm was therefore applied to the raw spectrum of the Nanodiscs without "baseline" subtraction or smoothing procedures.

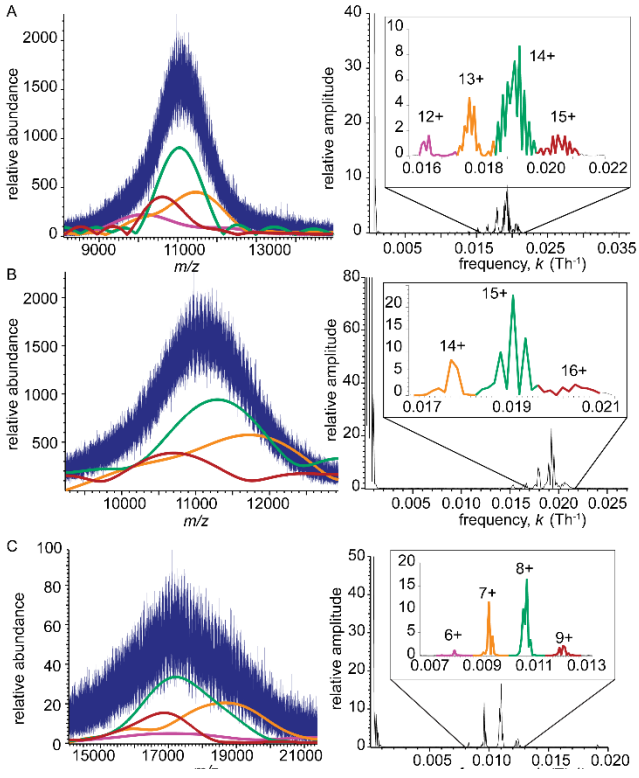


Figure 9. Mass spectra (left) and corresponding FFT spectra (right) for Nanodiscs containing 18:0-SM (A), DOPC (B), and DPPC (C). IFFT of the charge-state specific peaks in the Fourier domain (insets) are shown as envelope functions of the same color as overlays in mass spectra. The high-intensity peak near zero frequency in the FFT spectra has been truncated for clarity.

Example mass spectra and the corresponding Fourier spectra for these Nanodiscs are shown in Figure 9 (additional Nanodisc mass spectra and Fourier spectra can be found in Figures A1 through A4 in Appendix A). While they are very challenging to identify by conventional means directly from the mass spectrum, charge states are easily identified in the *k*-domain. *k*-domain peaks were found to be well separated with little to no overlap, a result attributed to the large width of the mass spectral envelope function compared to the

mass of the repeated subunit (the lipid, typically 650-800 u; much smaller repeated "subunits," such as water, ammonium, or sodium adductions, were not resolved). Furthermore, while peak shapes in the k-domain were found to be different, different charge states can have different peak shapes due to differences in the distribution of subunits for each charge state, i.e. the charge state specific *m/z* envelope. Charge states for Nanodisc samples containing different lipids were found to be similar when measured under identical instrumental conditions, although they vary somewhat depending on the type of lipid used for Nanodisc assembly. Lastly, very low-frequency peaks, including the fundamental, were not observed in the Fourier spectra, an observation that allows for an unambiguous assignment of charge states. The mass of each lipid was determined from the Fourier spectra to be within two standard deviations of their theoretical values for all 13 Nanodisc samples analyzed (and within one standard deviation for 11 of these). The results of these experiments (not including DMPC Nanodisc triplicate experiments, discussed below) under three different collisional activation conditions are summarized in Table 1. These results demonstrate that native Nanodisc ions containing different types of lipids can be readily distinguished using the FFT algorithm with no prior knowledge of the lipid mass.

Individual *k*-domain peaks were subjected to inverse Fast Fourier Transform (IFFT) to determine the envelope functions (i.e. mass spectra) for each individual charge state (Figure 9). Assuming a base mass of 49450 u (see Appendix A), the distribution of lipids in the Nanodiscs was determined for each sample (Tables 1 and 2) and found to be in good agreement with previously reported values.[51, 52] "Reconstruction" of the mass spectrum from the IFFT data is achieved by summing the IFFT of the *k*-domain data with

Table 1. Charge States, Lipid Masses, and Lipid Stoichiometries Determined for Nanodisc Ions Using Fourier Transformed Mass Spectra

Lipid	Collisional Activation Condition	Theoretical Average Mass (u)	Measured Mass (u) from IFFT Data	*z*	Lipid Stoichiometry
18:0-SM	Minimal	731.1	730.9 ± 0.2	19+	153 ± 8
				20+	173 ± 10
				21+	165 ± 7
18:0-SM	Mild	731.1	730.9 ± 0.5	13+	140 ± 15
				14+	152 ± 8
				15+	168 ± 10
DMPC	Mild	677.9	678.8 ± 0.8	13+	142 ± 12
				14+	151 ± 10
				15+	157 ± 13
DOPC	Mild	786.1	789.0 ± 3.0	14+	140 ± 10
				15+	153 ± 9
				16+	160 ± 10
DPPC	Mild	734.0	734.1 ± 0.2	13+	142 ± 12
				14+	147 ± 13
POPC	Mild	760.1	760.1 ± 0.4	12+	111 ± 10
				13+	117 ± 11
18:0-SM	Strong	731.1	731.0 ± 1.6	8+	134 ± 9
				9+	143 ± 8
DMPC	Strong	677.9	677.0 ± 0.6	5+	66 ± 10
				6+	75 ± 10
DPPC	Strong	734.0	733.0 ± 1.0	7+	107 ± 14
				8+	124 ± 12
				9+	141 ± 10
POPC	Strong	760.1	760.4 ± 0.4	6+	61 ± 6
				7+	82 ± 16
				8+	103 ± 8

a value of k within (k_f/2) of 0 and those from the identified charge-state peaks, as shown in Figure 10 for DPPC Nanodiscs. Inclusion of higher-order harmonic peaks in the reconstruction changes the reconstructed mass spectrum only slightly, indicating that remaining deviations from the raw mass spectrum are almost entirely due to random noise. This is explained by the nearly flat (i.e., "white") noise profile in the Fourier spectrum in the region of interest (Figure 10) and represents an instance of the "Fellgett

advantage" offered by Fourier transform that is exploited in many spectroscopic techniques.[53]

To observe the effects of collisional activation for Nanodiscs, a step often used to "clean up" ion populations with poor mass spectral resolution by removing small adducts, samples were collected in different collisional activation conditions which I term "minimal", "mild", and "strong" activation (see Methods). After "minimal" activation, Nanodisc ions were found to have *m*/*z* values between 8000 and 9000, with charge states between 17+ and 21+. For "mild" (resp., "strong") activation conditions, Nanodisc ion *m*/*z* values ranged from 10,000 to 12,000 (resp., 19,000 to 23,000), with charge states from 12+ to 16+ (resp., 5+ to 9+). Furthermore, stronger activation conditions generally correlated with lower charge states and a lower average number of lipids and wider distribution in the number of lipids contained within the Nanodiscs (Table 1). For example, 18:0-SM Nanodiscs had charge states from 19-21+, 13-15+, and 8-9+ after minimal, mild, and strong collisional activation, with an average number of lipids of 153-173, 140-168, and 134-143, respectively. These results are

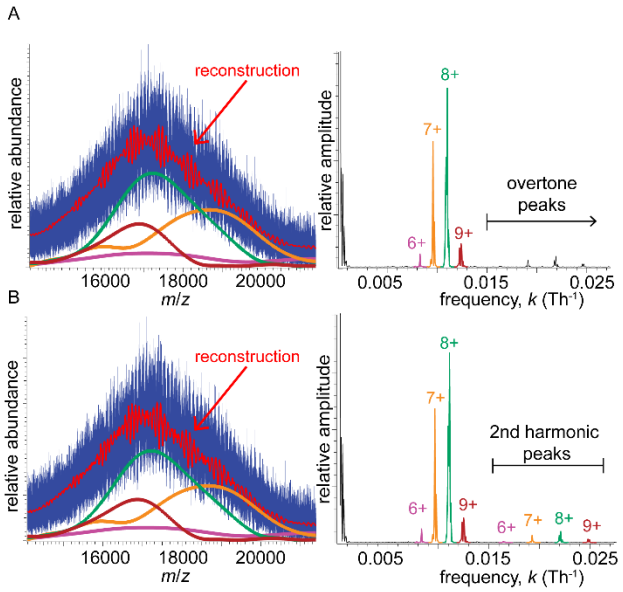


Figure 10. Reconstructions of mass spectrum (left, blue trace) for DPPC Nanodisc ions acquired under "strong" collisional activation conditions. Colors associated with charge states in the FFT spectra (right) correspond to like-colored charge-state-specific mass spectral envelopes in the mass spectra (left). The bright red traces overlaid over raw mass spectra are reconstructions using the FFT data for first harmonic peaks only (A), or both first and second harmonic peaks (B).

consistent with previously reported trends for ion mobility-separated Nanodisc mass spectra analyzed with Bayesian deconvolution. [14]

To assess the reproducibility of these Nanodisc experiments, a triplicate experiment was performed for Nanodiscs containing DMPC under "minimal" collisional activation conditions. The results of this experiment are summarized in Figure 11 and Table 2. Charge states are similar, but not identical, for each sample, and *k*-domain peak shapes differ between the three samples. However, upon IFFT of the individual *k*-domain peaks back to mass spectrum, the charge-state specific mass spectral envelope functions appear to be remarkably similar between the three samples, with two of the replicates having nearly identical mass spectra for the most-abundant common charge states (19+ and 20+; Figure 6, top and middle). For these same charge states, the third sample has slightly different reconstructed charge-state specific mass spectral envelopes (Figure 11, bottom), but the average and standard deviation in the number of lipids are nearly the same among all three samples. The average number and

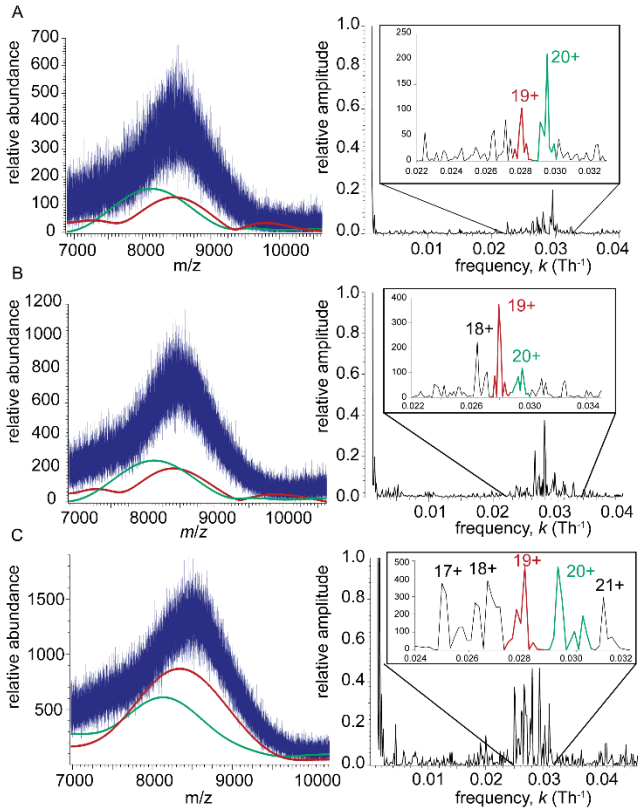


Figure 11. Mass spectra (left) and FFT spectra (right) for triplicate experiments with DMPC Nanodiscs acquired under "minimal" collisional activation conditions (see text). The high-intensity peak near zero frequency in the FFT spectra has been truncated for clarity.

Table 2. Lipid Masses and Stoichiometries for Common Observed Charge States in DMPC Nanodisc Triplicate Experiment.

Triplicate Sample	Collisional Activation Conditions	Theoretical Average Mass (u)	Measured Mass (u) from IFFT Data	*z*	Lipid Stoichiometry
DMPC 1	Minimal	677.9	678.4 ± 1.0	19+	166 ± 9
				20+	168 ± 15
DMPC 2	Minimal	677.9	678.2 ± 0.8	19+	164 ± 11
				20+	163 ± 18
DMPC 3	Minimal	677.9	678.1 ± 0.4	19+	163 ± 11
				20+	169 ± 13

standard deviation in the number of lipids for these samples are similar to those previously reported for native-like DMPC Nanodiscs analyzed with an FT-ICR,[51] Orbitrap,[24] or Synapt[14] mass spectrometers.

To demonstrate further the utility of the FFT algorithm, it was also used here to obtain the first reported native stoichiometry of Nanodiscs assembled with sphingomyelin, a major raft-associated lipid, as well as the effective bilayer surface area of this lipid. Under "minimal" activation conditions, the total number of lipids contained within the 18:0-SM Nanodiscs was found to be 163 ± 15. This value corresponds to a mean lipid bilayer surface area of 54 ± 5 $Å^2$ per 18:0-SM lipid, assuming a total bilayer surface area of 4400 $Å^2$ for MSP1D1 Nanodiscs.[52] The result is in good agreement with simulated aqueous 18:0-SM bilayers above their melting transition temperature, where the effective lipid surface area was found to be ~53 $Å^2$.[54] The effective Stokes radius of the 18:0-SM Nanodiscs was determined by size exclusion chromatography to be 4.1 ± 1.0 nm, within error of previously reported values for Nanodiscs containing other phospholipids (Figure 1).[55]

Conclusions

The one- and two-dimensional forms of the Fourier transform algorithm described here can be very useful to extract mass spectral information for ions with broad distributions of mass and charge arising from differing stoichiometries of a repeated subunit. Using this algorithm, effects of competitive ionization for electrospray of mixed-polymer samples were observed, and the native stoichiometry of 18:0- sphingomyelin-MSP1D1 Nanodiscs and the effective bilayer surface area of this lipid in Nanodiscs were measured for the first time. The algorithm can be applied to many different types of analyte, from polymers to non-covalent biomolecular assemblies, and mass and charge analysis using the algorithm appears to especially facile with broad mass spectra that represent a wide distribution of subunit stoichiometries. Major advantages of the algorithm are that it does not require guesses of subunit mass, ion charge states, or total mass, and peak-tail baseline subtraction and smoothing are not necessary when using this algorithm. Additionally, the algorithm can work well without the need for extensive collisional clean-up, potentially resulting in more accurate native stoichiometry determination for weakly bound assemblies such as biomolecular assemblies. Disadvantages to the Fourier method include complications in interpretation arising from overlap of Fourier domain peaks with adjacent peaks or overtones of other peaks, which are both less likely to occur for very large ions with broad mass spectra. These disadvantages could in principle be mitigated somewhat by using ion mobility separation prior to mass analysis or other separation methods.

Future application of this method could aid in characterizing biomolecular ions with many repeated post-translational modifications or ligands as well as lipid-protein

assemblies. Furthermore, while the algorithm has been demonstrated for complexes that contain one type of repeated subunit, Chapter 5 explores its use in identifying multiple different types of repeated subunit contained within the same ions. Techniques, such as Cation-to-Anion Proton Transfer Reactions (CAPTR),[56] charge stripping,[42] reactions with gas phase[57] or solution phase additives,[58] or electron transfer dissociation[59] that lead to improved resolution in the mass spectrum without leading to subunit dissociation could potentially enable more accurate mass determination using the Fourier algorithm, because the resulting broader mass spectra can yield better-resolved peaks in the Fourier domain.

While the Fourier spectrum is effective at determining the subunit mass, charge states, and charge-state-specific stoichiometry for the mass spectra of heterogeneous mass populations, there is still other valuable information that can be learned that is not discussed in this chapter. Higher harmonic frequencies are a common occurrence in the Fourier transformed spectra, and in the next chapter, a detailed description will be given as to what can be learned from these additional signal components. This includes how to determine the shape of a peak from a highly congested mass spectrum, how the higher harmonics can be used as a smoothing filter for the mass spectrum, and how all this information combined can be useful for other fitting algorithms such as UniDec (see Chapter 1).

CHAPTER III

HIGHER HARMONIC ANALYSIS FOR EXTREME CASES OF MASS SPECTRAL CONGESTION

Includes co-authored material from:

> Cleary, S. P., et al. "Extracting Charge and Mass Information from Highly Congested Mass Spectra Using Fourier-Domain Harmonics" *J. Am. Soc. Mass Spectrom.* **2018,** *29*, 2067-2080.

Introduction

Native electrospray ionization mass spectrometry (ESI-MS) can be a powerful tool for investigating the stoichiometry of large, multi-subunit biomolecular assembly ions.[2, 19, 21, 29, 51, 60-75] However, as polydispersity, size, and complexity of these biomolecular ions increase, accurate mass and charge determination can become very challenging, because salt or other co-solute adduction[76, 77] and a high density of mass spectral peaks[5, 6, 13, 14, 17, 51, 78-80] reduce practical resolution. Although separation methods such as chromatography and ion mobility separation can improve resolution of ion subpopulations differing sufficiently in chemical properties or shape,[6, 34, 42, 62, 81-86] many types of analytes can still be difficult to separate with these methods, necessitating development of new methods for determining charge states and assembly stoichiometries. In addition to charge-stripping or -reducing approaches, which can increase peak spacing,[42, 87] fitting-based mass deconvolution algorithms have been shown to greatly facilitate charge state determination and mass analysis.[14, 47, 88-90] However, to obtain accurate results from these fitting algorithms, the user is typically required to input estimates for some initial parameters, such as mass and charge state ranges and widths of the mass spectral peaks, that are close to the true values. It can thus be highly

advantageous to have accurate estimates of these parameters before implementing the algorithms in order to obtain reliable results. Alternatively, Fourier Transform-based deconvolution approaches require minimal data processing and parameter guessing.[4, 13, 91] For example, Fourier transform requires only linear data interpolation and specification of the minimum allowed number of data points separating frequency-domain peaks.[13]

Here, I extend the previously-introduced Fourier-Transform-based method (chapter 2) for characterizing disperse assembly ion mass spectra,[13] focusing on the use of information contained in higher harmonic peaks in the Fourier domain, including mass spectral peak shape information. I illustrate the use of higher harmonics in studying lipid-protein Nanodiscs, which are self-assembled discoidal non-covalent assemblies consisting of a phospholipid bilayer surrounded by two amphipathic helical membrane scaffold proteins (MSPs).[92] Nanodiscs have been used in numerous biochemical applications to study isolated, embedded membrane proteins and protein complexes[92, 93] as well as in native IM-MS to study lipid binding to peripheral and transmembrane protein complexes.[1, 79, 94]

A common occurrence in native ESI-MS analysis of Nanodiscs and other large, disperse ions, is a relatively large baseline in the raw mass spectrum.[10, 13, 60, 78, 85, 95] This results from the overlap from the superposition of many closely-spaced peaks and can occur even for mass spectra acquired on high-resolution quadrupole–time-of-flight (QTOF) instruments.[13, 78, 95] In such cases, it can be difficult to intuit what part of the mass spectral signal constitutes the baseline and what part contains information that can be used to determine ion masses and charge states. Data pre-processing software for many commercial mass spectrometers, including some Orbitrap and Fourier Transform-

Ion Cyclotron Resonance (FT-ICR) mass spectrometers, often performs baseline subtraction, apodization, or other signal processing that can alter peak shapes before data is displayed to the user.[96] After first discussing how Fourier Transform (FT) can be used to extract the information-rich portion of mass spectra for assembly ions, I show how analysis of harmonic peaks in the Fourier spectrum can be used to mitigate artifacts resulting from overlap or low signal-to-noise of peaks in the Fourier spectrum. Previously analyzed native Nanodisc mass spectra acquired on Orbitrap and FT-ICR mass spectrometers[60] are used as benchmarks. Advantages of this FT method are then illustrated for a mass spectrum of native-like "large" Nanodiscs containing over 300 lipids acquired on a QTOF mass spectrometer, which represents an extreme example of mass spectral congestion and is also the first reported mass spectral analysis of this type of "empty" Nanodisc. Finally, I show how information learned from the higher harmonic Fourier analysis can be used to characterize mass spectral peak shapes as well as to determine values for input parameters to improve the quality of results obtained from a Bayesian deconvolution method, UniDec.[14]

Methods

Nanodisc preparation. Nanodiscs containing dimyristoylphosphatidylcholine (DMPC) or dipalmitoylphosphatidylcholine (DPPC) were prepared according to a method adapted from that of Sligar and coworkers.[55] Briefly, all lipids (Avanti Polar Lipids Inc., Alabaster, AL, USA) were prepared as 5 mg/mL solutions in chloroform, dried until opaque with dry nitrogen gas, and resuspended to a final concentration of 50 mM in a pH 7.4 aqueous buffer containing 100 mM sodium cholate (Sigma-Aldrich, St.

Louis, MO, USA), 20 mM Tris (Bio-Rad, Hercules, CA, USA), 100 mM sodium chloride, and 0.5 mM ethylenediaminetetraacetic acid (EDTA). Membrane scaffold protein MSP1D1 or MSP1E3D1 (Sigma-Aldrich, St. Louis, MO, USA), for "small" or "large" Nanodiscs, respectively, was reconstituted in pH 7.4 aqueous buffer (20 mM Tris, 100 mM sodium chloride, 0.5 mM EDTA, 0.01% sodium azide) to a concentration of ~200 μM. Lipid suspensions were mixed with MSP1D1 solutions and additional buffer to a final concentration 50 μM in MSP1D1, 4.0 mM in DMPC, and 25 mM in cholate. MSP1E3D1 Nanodiscs were prepared in a similar fashion, differing only in the final lipid concentration, which was 9.0 mM for DPPC. The samples were incubated for 1 h at room temperature for DMPC (37 °C for DPPC). Nanodisc self-assembly was initiated by 1000:1 (vol:vol) dialysis into the Tris buffer, and BioBeads SM-2 (Bio-Rad, Hercules, CA, USA) were added to the dialysis buffer after being previously rinsed and sonicated three times in methanol followed by three times in the reconstitution buffer. Nanodisc samples were removed from dialysis after 24 hours and were purified by size exclusion chromatography. Fractions containing Nanodiscs were pooled together and concentrated to a final concentration ~10 μM in Nanodiscs. Concentration was determined by UV absorbance of the MSP1D1 or MSP1E3D1 protein, and divided by two (due to the presence of the two scaffold proteins) to obtain Nanodisc concentration.

Mass spectrometry. Orbitrap-EMR (Amgen, Thousand Oaks, CA, USA). Experiments were performed on a modified Exactive Plus instrument (ThermoFisher Scientific, Bremen, Germany) equipped with a nanoelectrospray ionization (nESI) source. All critical instrument voltages and pressures were as follows: The capillary voltage 0.8-1.5 kV. Ions formed by nESI were transmitted through a heated stainless steel

capillary (4.25 cm long ion transfer tube) maintained at 200 °C into an S-Lens stacked ring ion guide with an applied RF-amplitude (peak-to-peak) of 200 V. Ions then traveled through a transport multipole and entered the HCD cell where they were stored at high pressure before they were returned to the C-trap. This feature allows efficient trapping and desolvation of large protein ions and dramatically improves sensitivity. Nitrogen gas was used in the C-trap as well as the HCD cell. Using a trapping gas pressure setting of 7.0 (software-determined) the C-trap pressure was approximately 2×10^{-4} mbar and the UHV pressure (Orbitrap analyzer) was 7.5×10^{-10} mbar. The voltage offsets on the transport multipoles were manually tuned to increase the transmission of large complexes (C-trap entrance lens; 0 V, bent flatapole DC, 4 V; inter-flatapole DC 4 V; injection flatapole DC, 4 V). An in-source CID voltage of 50 V to 120 V and an HCD voltage of 20 V to 100 V were required to achieve efficient sample desolvation. Raw transients were processed using enhanced Fourier Transform [96] for converting the time-domain data into frequency before *m*/*z* conversion; this data pre-processing is standard for this generation of Orbitrap instruments. The instrument was set at a nominal resolving power of 70,000 at *m*/*z* 200, and mass spectra were acquired for 2 minutes by averaging 10 microscans per analytical scan. Data were analyzed using XcaliburTM2.2. No additional data processing was performed before Fourier-Transform analysis.

Quadrupole–Time-of-Flight mass analyzer (University of Oregon, Eugene, OR, USA). All QTOF mass spectrometry experiments were performed with a Synapt G2-S*i* ion mobility mass spectrometer (Waters Corp. Milford, MA, USA) using a static nESI source. nESI emitters were prepared by pulling borosilicate capillaries (i.d. 0.78 mm, Sutter Instruments) to a tip i.d. of ~1 μm using a Flaming-Brown P-97 micropipet puller

(Sutter Instruments, Novato, CA, USA). For each sample, ~3−5 μL of solution was loaded into an emitter, which was placed approximately 1-2 mm from the entrance of the mass spectrometer. A platinum wire inserted into the solution was used to apply an electrical potential of 0.6-1.0 V relative to instrumental ground to initiate electrospray. The ion source was equilibrated to ambient temperature. Mass spectra were collected in Sensitivity mode to maximize signal-to-noise. Argon gas was introduced into the Trap at a flow rate of 5 mL/min in all experiments. Trap/Transfer collision energy settings of 100/5 V, respectively, were used for all Nanodiscs. Nanodisc mass spectra were collected for 20 min and the continuously collected data were summed. All mass spectral data were processed without smoothing using MassLynx v. 4.1 (Waters Corp., Milford, MA, USA).

Fourier Transform-Ion Cyclotron Resonance (FT-ICR) (UCLA, Los Angeles, CA, USA). All FT-ICR native-MS experiments were performed using SolariX 15 Tesla instrument (Bruker Daltonics, Billerica, MA, USA). The nESI capillary voltage was set to 0.6 to 0.8 kV in positive ionization mode. The temperature of drying gas was 100 °C and the flow rate was 2.5 L/min. The RF amplitude of the ion-funnels was 300 V peak-to-peak, and the applied voltages were 210 V and 6 V for funnels 1 and 2, respectively. The voltage of skimmer 1 was 50 and the skimmer 2 voltage was 20 V. The lowest values of RF frequencies were used in all ion-transmission regions: multipole 1 (2 MHz), quadrupole (1.4 MHz), and transfer hexapole (1MHz). Ions were accumulated for 500 ms in the hexapole collision cell before being transmitted to the infinity ICR cell. A time-of-flight of 2.5 ms was used. Vacuum pressures for different regions were: source 2 mbar; quadrupole 2×10^{-6} mbar; ultra-high vacuum chamber 2.6×10^{-9} mbar. To assist in ion desolvation, mild collision-induced dissociation (CID) was performed in the hexapole

collision cell by collision with argon at a voltage of 4 V. 1000 scans were averaged for each spectrum using 256,000 data points (transient length 0.085 s) per scan in magnitude mode. The MS Control software was Compass solariXcontrol, version 1.5.0, build 103, and the data were apodized using a full-sine-bell function. The mass spectrometer was externally calibrated with a 50 μg/mL solution cesium iodide in 1:1 (vol:vol) acetonitrile:water over the *m*/*z* range 100 to 20,000.

Computational work. All FT-based analysis was performed using the Prell group's home-built program, iFAMS (interactive Fourier Analysis for Mass Spectra) v. 4.2. All mass spectra are symmetrized before FT is performed to yield real-valued Fourier spectra for ease of graphical presentation. Signal-to-noise is calculated as the maximum amplitude of a peak divided by the root-mean-square white noise at baseline in a neighborhood of the peak. Unless otherwise specified, all other data analysis was performed using Igor Pro v. 6.37 (WaveMetrics, Inc., Lake Oswego, OR, USA).

Theory

The principle of the Fourier Transform (FT)-based mass spectrum analysis method used here for deconvolving heterogeneous mass populations is described in detail in chapter 2, as well as the following references[4, 13, 91] and therefore will only be briefly explained here. The key concept to the FT algorithm is that a distribution of mass spectral peaks for an individual charge state *z* of an analyte containing various numbers of a repeated subunit with mass m_s can be described as a comb of equally-spaced peaks multiplied by a mass spectral envelope function. Mathematically the mass spectrum (s(*m*/*z*)) can be decomposed into three separate functions (Figure 12A): a comb function

(c(*m/z*)) with spacing m_s/z between adjacent delta functions, a peak shape function (p(*m/z*)) that is convolved with the comb function and describes the typical shape of the peaks in the comb, and an envelope function (e(*m/z*)) that describes the relative abundances of the peaks in the comb, i.e., the stoichiometry distribution. This relationship can be described symbolically in the following way:

1. $s(m/z) = [c(m/z)*p(m/z)]\times e(m/z)$

where the symbols * and × represent convolution and multiplication, respectively.

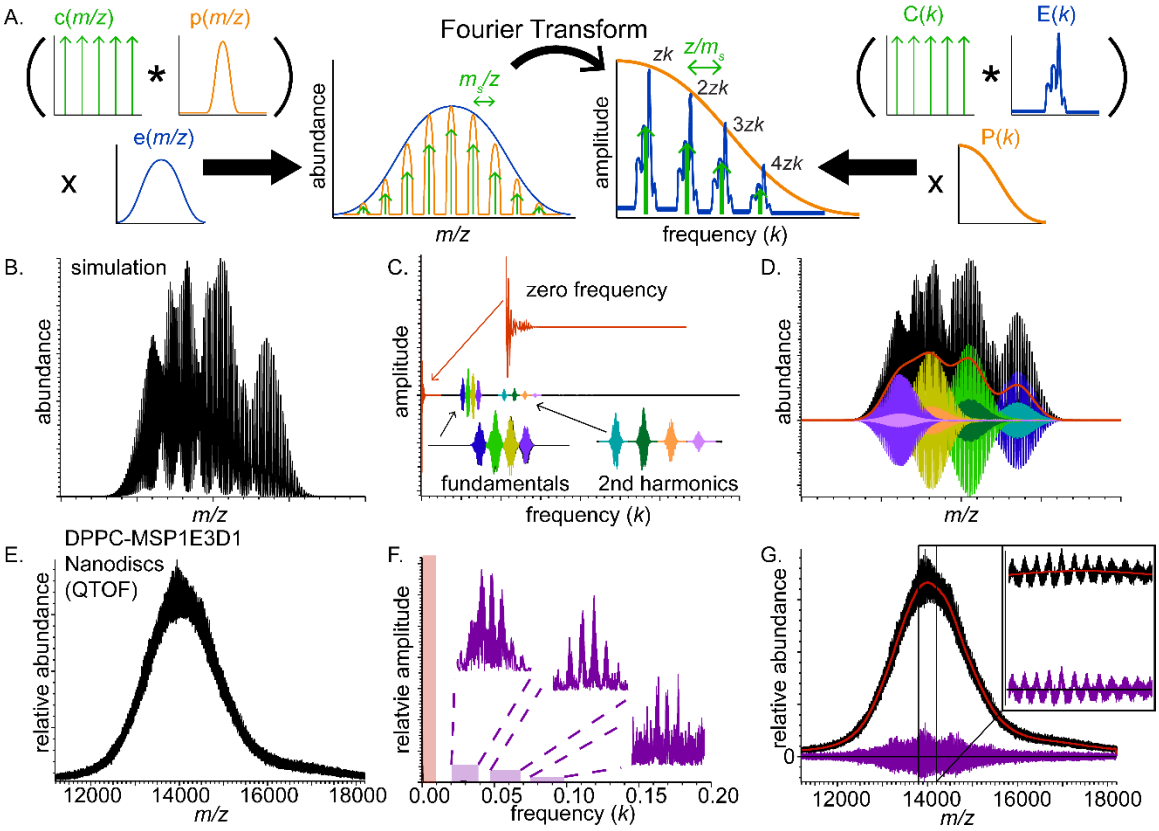


Figure 12. (A) Graphical depiction of mathematical decomposition of a nESI mass spectrum (left) and its Fourier Transform (right) for a polydisperse ion population with a repeated subunit. (B) Simulated mass spectrum, (C) corresponding Fourier spectrum, and (D) reconstructed FT-baseline (red) and charge-state-specific mass spectra with colors identical to their corresponding Fourier-domain peaks from C. (E) DPPC-MSP1E3D1 Nanodisc mass spectrum acquired on an QTOF mass spectrometer, (F) corresponding Fourier spectrum with highlighted zero-frequency band (red) and harmonic peak series (violet), and (G) reconstructed low-information signal (red) and FT-baseline-subtracted mass spectral signal (violet)

The FT of Eq. 1 is (hereafter referred to as the "Fourier spectrum"):

$$2. \quad S(k) = [C(k) \times P(k)] * E(k)$$

where $k = z/\Delta m$ is the frequency corresponding to peak spacing $\Delta m/z$ the mass spectrum. For a given charge state, the peaks in S(k) each have the shape E(k), which is the FT of e(m/z), and the comb of peaks in S(k) decays as P(k), which is the FT of p(m/z). The FT of the entire mass spectrum, which may include several charge states, is the sum of the Fourier spectra corresponding to each charge state present. The fundamental peaks in the Fourier spectrum for each charge state are spaced by $1/m_s$, thus m_s is found by computing the reciprocal of the fundamental peak spacing. The charge states present in the ion population are determined by multiplying the frequencies of the fundamental peaks in the Fourier spectrum by m_s, and the stoichiometry distribution e(m/z) can be found for a particular charge state by inverse Fourier Transforming its corresponding fundamental peak in the Fourier spectrum. The relationship between these characteristics of the mass spectrum and its Fourier Transform is illustrated for one charge state series in Figure 1A.

P(k), which describes the decay of the harmonic frequency peaks for a given charge state, is the FT of p(m/z), the average peak shape for the mass spectral peaks for that charge state. The inverse Fourier Transform of P(k) is therefore p(m/z), the width of which reflects the width of each peak in the comb associated with the chosen charge state.

Results and Discussion

In extremely congested mass spectra with poor resolution, a significant baseline is often observed that may have a complex shape. Beyond a constant or linear baseline subtraction, other forms of baseline fitting and subtraction are often used in mass spectral

analysis software that can model curved baseline shapes. These algorithms include fitting a baseline to a polynomial[60] or smoothing a spliced sequence of step functions that measure local minima in the raw mass spectrum,[47] and data points with negative abundance values after baseline fitting and subtraction are often removed or set to zero abundance. These methods can be highly effective when adjacent peaks of interest are well-separated and the baseline originates primarily from chemical interferents, such as salt or small molecule clusters.[26, 88] In cases where many peaks of interest overlap strongly, a significant baseline, perhaps even larger in magnitude than the modulation depth of the spectrum, can arise due to the superposition of the tails of these peaks.[24, 78, 97] In such instances, it is not always easy to assess to what extent the above-described baseline subtraction methods distort the true peak shapes, centroids, or relative abundances of ions in the mass spectrum.

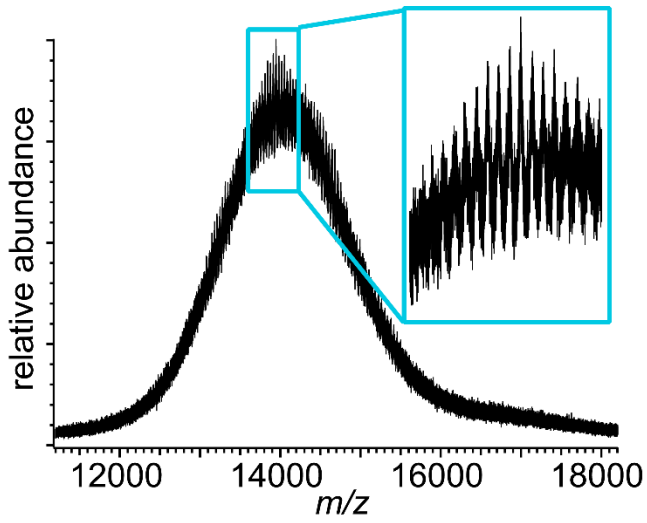


Figure 13. Mass spectrum of lipoprotein Nanodisc ions assembled with DPPC lipid and membrane scaffold protein MSP1E3D1. Inset shows signal modulation in *m/z* range 13,600-14,200

For example, Figure 13 shows a QTOF native mass spectrum of "large"-diameter (12.9 nm wide[52]) Nanodiscs assembled using membrane scaffold protein MSP1E3D1 and DPPC lipids. This and other sizes of Nanodiscs are widely used in biochemical studies of isolated membrane protein complexes.[1, 92, 98, 99] A very large, curved baseline is observed that results from the overlap of many poorly-resolved peaks

attributed to individual charge states and lipid stoichiometries. Poorly-resolved mass spectra of disperse native-like ion populations,[5, 17] such as this, motivate a careful analysis of the baseline and signal modulation to determine what portion of the mass spectral signal can be used to reconstruct ion mass, charge, and stoichiometry information, and how this can be done reliably. After establishing the self-consistency of an FT-based method for analyzing similar Nanodisc mass spectra acquired with significantly higher resolution and other instruments, results from this method for this lower-resolution mass spectrum in Figure 13 are presented below.

FT-based approaches for baseline characterization and noise filtering in mass spectra. Fourier filtering is a well-known procedure in signal processing and many spectroscopic techniques that can facilitate detection and characterization of periodic signals even in a noisy environment as well as removal of low-frequency baselines.[100-102] A raw signal is first Fourier transformed, signals at frequencies of interest are identified, and these signals are extracted using frequency windows before inverse Fourier transforming to yield the filtered signal in the original domain. Figure 12 illustrates a simulated mass spectrum for a population of assembly ions containing a range of subunit stoichiometries and charge states as well as the corresponding FT spectrum. Signal in the Fourier spectrum is observed near zero frequency and at several series of equally-spaced peaks corresponding to the fundamental and higher harmonics belonging to each charge state (Figure 12C). Because FT is a linear operation, the Fourier spectrum is a sum of each charge-state-specific signal in the Fourier domain. White noise and other non-periodic signals in the mass spectrum are spread out across all frequencies in the Fourier spectrum. While the spacing and shapes of the fundamental and harmonic peaks contain

information about charge state, subunit mass, and stoichiometry distribution,[13] it is extremely challenging to extract this information from the frequency band around 0 in the Fourier spectrum because this band contains a superposition of signals from all of the charge states. This band is thus much less information-rich than the other peaks in the Fourier spectrum, and it is reasonable to define its inverse Fourier Transform (IFT) as the "low-information signal" of the mass spectrum (see Figure 12D, red trace).

The remaining signal after subtracting the low-information signal in the mass spectrum thus contains the most useful information, and it typically takes on both positive and negative values that oscillate about zero (see Figure 12D). It should be noted that this signal is generally not the same as the signal obtained by subtracting a low-order polynomial baseline or by using other common mass spectral domain baseline fitting approaches.[14, 47] Its FT corresponds to the entire Fourier spectrum outside the zero-frequency band. As shown for the simulated, noise-free mass spectrum in Figure 13, for each charge state, the envelope of the IFT of each harmonic has identical shape but different scaling, and the low-information signal resembles the sum of these envelopes (Figure 12D). A similar decomposition of the experimental mass spectrum from Figure 13, which contains white noise, is shown in Figure 12E-G.

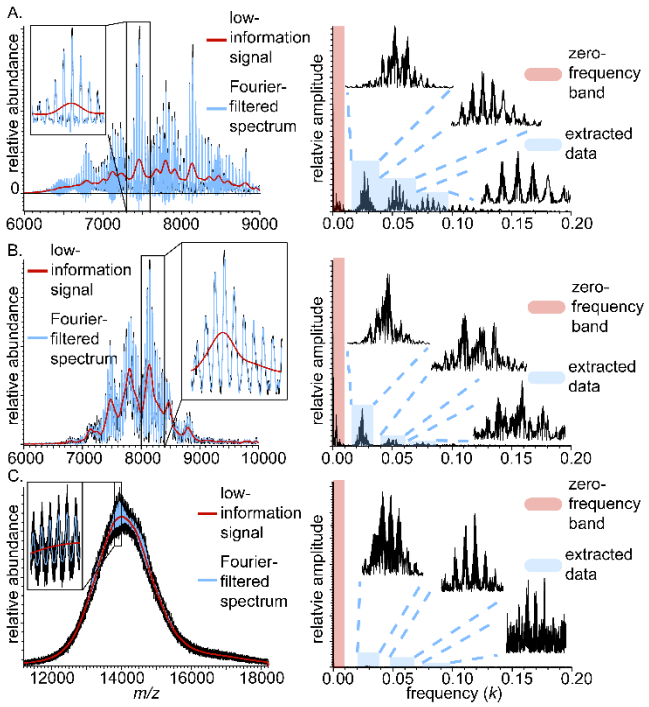


Figure 14. Mass spectra (left) and corresponding Fourier spectra (right) for native-like DMPC-MSP1D1 Nanodiscs measured on (A) Orbitrap and (B) FT-ICR mass spectrometers, and (C) a DPPC-MSP1E3D1 Nanodisc mass spectrum measured on a QTOF mass spectrometer. Fourier-filtered mass spectra (blue overlays) are calculated using the first three harmonics, as well as the Fourier baseline (red) the corresponding to the zero-frequency band. Detailed peak structure illustrated in insets

Fourier filtering has the potential advantage over simpler low- or high-pass filtering methods in that noise outside and between frequency bands of interest can be eliminated, and signal at several frequencies can be characterized simultaneously. Fourier filtering is equivalent to removal of all signal and noise outside the zero-frequency band and Fourier domain peaks. Figure 14 shows the low-information signal and Fourier-filtered mass spectrum reconstructions for previously reported mass spectra of DMPC-MSP1D1 NDs ionized using nESI on an Orbitrap–EMR and 15 Tesla FT-ICR,[60] and for DPPC-MSP1E3D1 NDs acquired using a QTOF (same data as in Figure 13). The FT-ICR and Orbitrap mass spectra have nearly completely resolved peaks and very small low-information signals, whereas the QTOF mass spectrum exhibits a very large low-information signal, in agreement with visual intuition. The root-mean-square (RMS) random noise in the Fourier spectrum for the QTOF data is essentially constant as a function of frequency, whereas it decreases with increasing frequency (up to the

Nyquist frequency) for the FT-ICR and Orbitrap data. These differences in mass spectral resolution and frequency-dependence of the RMS random noise can be explained in part by noting that Orbitrap pre-processing software combines magnitude- and absorption-mode data to improve apparent mass spectral peak resolution,[96] and both the Orbitrap and FT-ICR data (which is analyzed here in magnitude mode) are apodized during pre-processing to remove ringing artifacts caused by a finite sampling period. The small depth of modulation for the Fourier-filtered QTOF mass spectrum (Figure 14C, blue trace) as compared to the FT-baseline subtracted mass spectrum in Figure 12G (violet trace) is consistent with the relatively large random white noise and illustrates the utility of Fourier-filtering in removing this type of noise. A comparison of Fourier filtering with other common noise filtering techniques (Savitzky-Golay, moving-average, and median filters) is shown in Appendix B (Figure B1) and illustrates that Fourier filtering typically removes much more low- and high-frequency noise from the mass spectrum than do these other filters and has the additional advantage of leaving Fourier-domain peak amplitudes unaltered.

When multiple harmonics are resolved in the Fourier spectrum, there are in principle multiple pathways by which to determine the subunit mass, charge state distribution, and charge-state-specific mass spectra for this type of ion population, though different pathways may have unique advantages for real data exhibiting Fourier peak overlap or low signal-to-noise. In addition to the spacing and shape of individual peaks in the Fourier spectrum, the relative scaling of the harmonics contains information about the shape and width of peaks in the mass spectrum. Extraction of information from harmonic peaks is described below for Nanodisc mass spectra.

Consistency of subunit mass and charge state determinations using different Fourier-domain harmonic peak series. In chapter 2, FT-based analysis of mass spectra with well-resolved fundamental peaks in the Fourier domain was illustrated.[13] In all three Fourier spectra shown in Figure 14, the fundamental peaks overlap significantly, thus stoichiometry information obtained by directly inverse Fourier transforming the fundamental peaks may be unreliable, because overlap of signal from adjacent Fourier-domain peaks and use of frequency windows that are too narrow are potential sources of error in Fourier filtering. Figure B2 illustrates the relationship between Fourier-domain peak separations, IFT window width, and reconstructed mass spectral ringing artifacts (which are less than 5% of the maximum signal when Fourier-domain peak separation is at least ~1.5 times the sum of adjacent peak widths). A potential solution to these problems is to use higher harmonic peaks rather than fundamental peaks to determine subunit stoichiometry distributions, because they are more widely spaced and less to prone to overlap, though it is important to note that higher harmonic peaks can have lower signal-to-noise than their corresponding fundamentals. Thus, a trade-off between overlap-induced artifacts and artifacts introduced by noise must be considered when performing this analysis.

Figure 15 illustrates this approach, where charge-state-specific mass spectrum reconstructions for the fundamental, second, and third harmonics are shown for the baseline-resolved Orbitrap mass data from Figure 14A. The fundamental peaks have extensive overlap, but charge states 15-23+ are still found using the FT method, in agreement with previously published results using UniDec.[60] A subunit mass of 678.5 ± 3.6 Da is calculated using the FT method, which is close to the known mass of DMPC

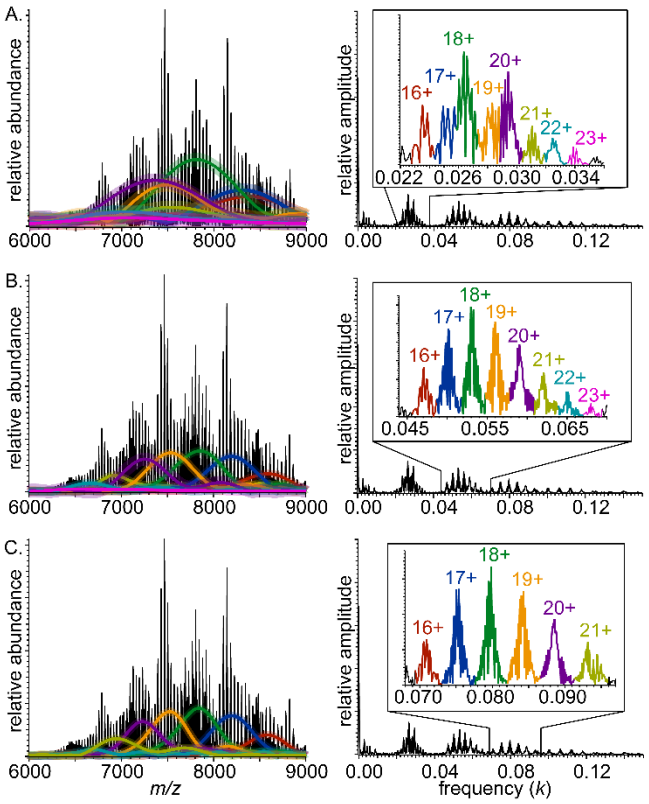


Figure 15. Mass spectrum of DMPC-MSP1D1 Nanodiscs acquired on an Orbitrap mass spectrometer (left) and corresponding Fourier spectrum (right) for (A) fundamentals, (B) second harmonics, and (C) third harmonics. IFT of the charge-state specific peaks in Fourier spectra (insets) are shown as overlaid envelope functions of the same color in mass spectra. Faint color surrounding envelope functions represents uncertainty of envelope functions calculated from average noise amplitude in the Fourier domain

(677.993 Da). These results are also consistent with the second harmonic series, which also indicate charge states 15-23+ and a more accurate and precise subunit mass of 678.2 ± 1.0 Da. Using the third harmonics, the FT method calculates similar subunit mass (677.3 ± 0.8 Da) but only identifies charge states 16-21+, due to overlap of the higher charge states with the 4th harmonics. Thus, overlap of two different charge states from two different harmonic series can in certain cases represents a limitation in the higher harmonic analysis. These results are summarized in Table B1.

While the charge states and subunit mass measurements are similar for the fundamental, second, and third harmonics, a notable difference can be seen in their reconstructed mass spectral envelope functions, particularly when using the fundamentals. The reconstructed charge-state-specific envelope functions using the fundamental peaks are visibly wider than those for the 2nd and 3rd harmonics, a result attributed to strong overlap of the fundamental peaks in the

Fourier spectrum (Figure 15). The higher harmonic peaks in the Fourier spectrum are more widely spaced and are better resolved, leading to narrower corresponding charge-state-specific envelope reconstructions (Figure 15). The corresponding standard deviations in the number of lipids present (Table B1) are consistently lower than those found using the fundamental peaks, and lipid statistics are nearly identical when using the 2^{nd} and 3^{rd} harmonics for charge states 16-21+. Notably, all of these Fourier-domain peaks have signal-to-noise greater than 10:1 and inter-peak separation exceeding 1.5 times the sum of adjacent peak widths. Subunit statistics for Fourier-domain peaks that meet these two criteria are therefore likely to be accurate in general even for mass spectra for which only one set of harmonic peaks meets them, as is often the case in the other Nanodisc spectra presented here. Similar results for mass spectra exhibiting even greateroverlap of the fundamental peaks are illustrated in Figure B3 and Table B1 in Appendix Figure 16 shows results for the much more congested QTOF mass spectrum from Figure 13, in which the mass spectral peaks are far from baseline-resolved. FT analysis of the fundamental frequencies indicates charge states 18-24+ and average sub-unit mass of 732. ± 2. Da, which is within one standard deviation of the average mass of DPPC (733.562 Da). However, the 19 and 20+ fundamental peaks overlap strongly, so their corresponding charge-state-specific mass spectra cannot be reliably reconstructed from these peaks alone. By contrast, the second harmonic peak sequence is baseline-separated. The same range of charge states, 18-24+, is identified, and a more accurate and precise average sub-unit mass of 733.0 ± 0.8 Da is determined. The average number and standard deviation in the number of lipid subunits for these charge states determined using the fundamental and second harmonic peaks are very similar. Together with the

results for DMPC-MSP1D1 Nanodisc ions described above, these results show that higher harmonic frequencies can often beneficial in determining charge states, subunit mass, and charge-state-specific mass spectra for disperse assembly ion populations. Similar results were obtained for the spectrum in Figure 14B and are shown in Figure B4 in Appendix B.

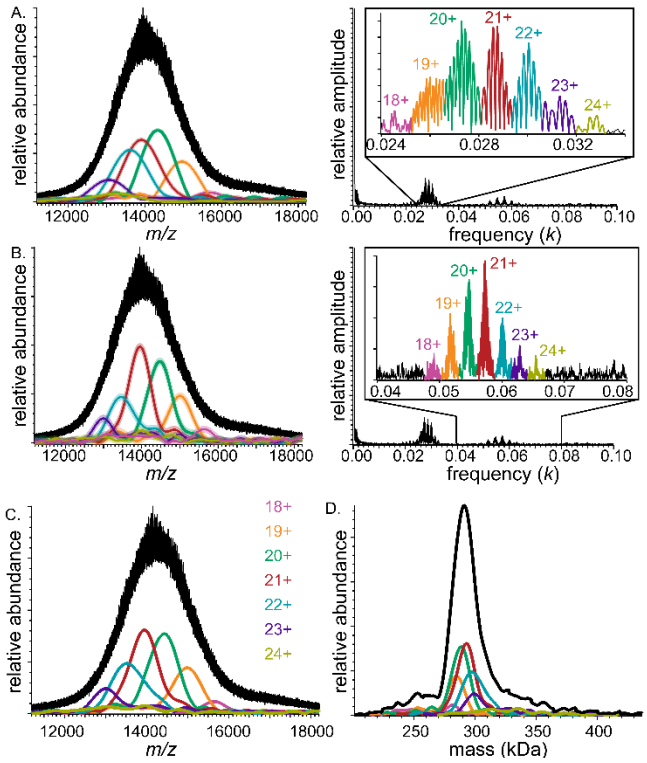


Figure 16. Mass spectrum of DPPC-MSP1E3D1 Nanodiscs acquired on a QTOF mass spectrometer (left) and corresponding Fourier spectrum (right) for (A) fundamentals and (B) second harmonics. IFT of the charge-state specific peaks in Fourier spectra (insets) are shown as overlaid envelope functions of the same color in mass spectra. Faint color surrounding envelope functions represents uncertainty of envelope functions calculated from average noise amplitude in the Fourier domain. (C) Harmonic-averaged reconstruction of envelope functions. (D) Zero-charge spectrum (black), calculated from harmonic-averaged spectra for all charge states

In addition to signal overlap and artifacts introduced by using overly narrow Fourier frequency windows (see Figure B2), artifacts in reconstructions of charge-state-specific envelope functions can occur if the white noise present in the Fourier spectrum obscures the true shape of higher harmonic peaks that have relatively low signal-to-noise. Averaging the shape of two or more reconstructed spectra from different harmonics belonging to the same charge state is a potential strategy for reducing artifacts of this type. “Harmonic-average” charge-state-specific mass spectra for the data from Figure 13 were reconstructed by

directly averaging the IFT of the fundamental and higher harmonic peaks for each charge state (Figure 16C). The harmonic-averaged spectral envelopes widths are slightly wider than those for the second harmonics, but contain fewer periodic "ripples" in the tails of the spectra due to averaging with the data from the fundamentals. A "zero-charge" mass spectrum for the entire ion population was also calculated from these harmonic-averaged mass spectra (Figure 16D) to illustrate the mass distribution for the entire ion population. (Zero-charge spectra for the Orbitrap and FT-ICR spectrum in Figures 14A and 14B can be found in Figures B5 and B6.) The average ion mass found by this method is ~290 kDa, which, assuming two scaffold proteins are present per ion, means that the DPPC-MSP1E3D1 Nanodiscs in the ion population represented in Figure 13 have an average of ~306 lipids. Using this number, an average condensed-phased area per lipid head group in the Nanodisc of ~60 A^2 can also be estimated, assuming that each leaflet has a diameter of 2 nm smaller than the diameter of the assembly.[52] This result agrees well with previous computational simulations of model DPPC bilayers.[103]

Characterizing peak width and unresolved adductions in the mass spectrum and Fourier domain. When baseline-resolved mass spectral peaks can be attributed to only one charge state, peak broadening caused by unresolved adductions (such as solvent molecules and small cosolute ions) for each peak can in principle be determined from the width of the peaks in the mass spectrum and (separately-measured) instrumental resolving power. If peaks from more than one charge state overlap or there is a large baseline, peak width determination directly from the mass spectrum can be more challenging. Statistics for the average peak shape for each charge state instead can be determined from analysis of the higher harmonic peaks in the Fourier spectrum, which

have amplitudes that decay as the FT of the average mass spectral peak shape for the corresponding charge state (P(*k*); see Figure 12A). Then, p(*m*/*z*) for each charge state is simply the IFT of P(*k*). However, for a typical ESI mass spectrum with multiple overlap charge states, it is typically infeasible to determine this zero-frequency contribution for individual charge states, but it may be reasonable to assume that the zero-frequency amplitude has a fixed relationship to the amplitudes of all the other harmonic peaks. For example, if p(*m*/*z*) is Gaussian in shape, as in the simulated spectrum in Figures B7A and B7B, its Fourier Transform, P(*k*), is also a Gaussian, and the FWHM of P(*k*) is inversely proportional to that of p(*m*/*z*). The FWHM of p(*m*/*z*) can then be determined from P(*k*) without knowing its amplitude at zero frequency, provided enough harmonics have sufficient signal-to-noise to confidently fit P(*k*) to a Gaussian.

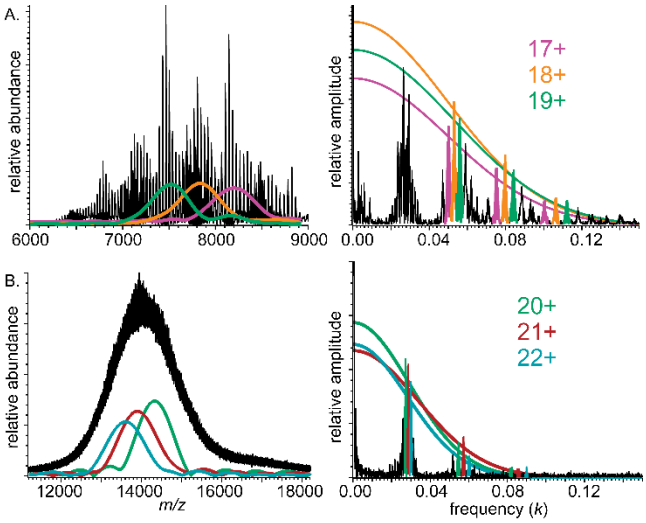


Figure 17. Mass spectra (left) and corresponding Fourier spectra (right) of (A) DMPC-MSP1D1 Nanodiscs acquired using an Orbitrap mass spectrometer and (B) DPPC-MSP1E3D1 Nanodiscs acquired using a QTOF mass spectrometer. Charge-state-specific mass spectral envelope functions (left), and Gaussian frequency decay functions (P(*k*), right) are shown with same color

The baseline-resolved Orbitrap mass spectrum in Figure 15A has roughly Gaussian peaks and is in this respect similar to the modeled spectrum in Figure B7B. Average peak-width statistics for each charge state can thus be readily estimated from reconstruction of P(*k*). This is shown in Figure 17A for charge states 17, 18, and 19+

using the second, third, and fourth harmonics (the fundamentals, as well as fifth and higher order harmonics, could not be used due to overlap). Using these parameters, fitting a Gaussian to the higher harmonics results in a reconstructed average FWHM, in *m/z*, of 8.1 ± 0.1, 7.9 ± 0.1, and 7.5 ± 0.1 for the 17, 18, and 19+ charge states respectively. Note that these and all peak width uncertainties reported below reflect fitting to a forced Gaussian shape and do not include uncertainty reflecting exact, i.e., non-Gaussian, peak shapes. These peak width values are close to forced-Gaussian FWHM measurements determined directly from the mass spectrum (7.8 ± 0.6, 6.4 ± 0.6, and 6.3 ± 0.6 for charge states 17, 18, and 19+), with the small discrepancies likely arising from the slightly asymmetric shape of the mass spectral peaks. This result is consistent with a small degree of salt adduction and the slight deviation of the data from the Gaussian fits shown in Figure 17A.

A similar analysis was performed for the poorly resolved QTOF spectrum from Figure 13 assuming Gaussian shape peaks (Figure 17B). The FWHM, in *m*/*z*, for charge states 20, 21, and 22+ are found to be 13.7 ± 0.2, 12.2 ± 0.1, and 13.6 ± 1.0, respectively. This is within error of values found by measuring the most abundant peak in each charge state after an initial smoothing and baseline subtraction (13. ± 1., 12. ± 1., and 13. ±1., respectively). Notably, both the directly measured and FT-reconstructed peak FWHM are less than half the inter-peak spacing for each charge state (~17-18, in *m*/*z*, for these ions). These results are summarized in Table 3.

Table 3. Mass Spectral Peak Widths for Native-Like Nanodisc Ions Determined Using P(*k*) and Directly from Mass Spectrum

Analyte (Instrument)	***z***	**FWHM found using P(*k*)**	**FWHM from Mass Spectrum**
DPPC-MSP1E3D1	20+	13.7 ± 0.2	13. ± 1.
(QTOF)	21+	12.2 ± 0.1	12. ± 1.
	22+	13.6 ± 1.0	13. ± 1.
DMPC-MSP1D1	17+	8.1 ± 0.1	7.8 ± 0.6
(Orbitrap)	18+	7.9 ± 0.1	6.4 ± 0.6
	19+	7.5 ± 0.1	6.3 ± 0.6

Potential errors in peak-width determination can occur when using this method if the mass spectral peak shapes are very poorly resolved or highly non-Gaussian, especially if a large baseline is subtracted before peak shape analysis. The extent of these potential errors was investigated using model spectra with peak FWHM ranging from 115% to 180% of the inter-peak spacing for each charge state. For example, Figure B7A shows a mass spectrum in which the peak FWHM is 115% of the inter-peak spacing. The FT approach described above yields a FWHM, in *m*/*z*, of 6.5, whereas direct measurement of the FWHM in the mass spectrum after baseline subtraction is ~5.1, slightly smaller than the true value (5.9). When the true FWHM are increased, the resulting mass spectra have even larger baselines. If these curved (i.e., polynomial-fitted, not FT) baselines are subtracted, the resulting spectrum is ostensibly baseline-resolved, and directly measuring the FWHM of the peaks leads to nearly identical measurements for the FWHM (5.3, 5.4, and 5.6, respectively), increasingly far from the correct values (7.1, 8.2, and 9.4), but all very similar to half the inter-peak spacing (5.1). This error trend arises from the fact that the signal modulation with respect to the low-information signal in the mass spectrum asymptotically approaches a simple sinusoid at the

fundamental frequency as the width of the individual peaks increases. In sharp contrast, Fourier transforming the model mass spectra, with or without subtracting the curved baseline, results in much more accurate FWHM measurements because the shape of P(*k*) is easily measured in the Fourier spectra. These results illustrate the robustness of this method and indicate one should exercise caution in interpreting mass spectra after curved baseline subtraction, especially as relates to peak width and resolution, if peak widths are similar to or larger than inter-peak spacing. Peak width analysis can be even more challenging in the case that the mass spectral peak shape is far from Gaussian (see Appendix B, especially Figure B7C and B7D).

Use of parameters determined from FT algorithm to improve results of Bayesian mass spectral fitting. Disperse mass populations with repeated subunits from native mass spectra, such as the Nanodisc mass spectra investigated here, can in many cases be deconvolved using other strategies besides Fourier Transform, such as Bayesian fitting, the approach implemented in UniDec/MetaUniDec,[14, 60, 104] or MaxEnt.[26] These fitting algorithms typically require the user to input some initial parameters, such as charge state, peak width, and mass estimates, and a modeled data set based on those parameters is iteratively fit to the experimental data set until convergence is achieved. (The MaxEnt algorithm further assumes that the stoichiometry distribution for each charge state is identical, which may not be the case for some disperse ion populations, such as the Nanodiscs investigated here; see Table B1.) When using these fitting algorithms, choosing the right initial parameters can therefore be highly advantageous for extracting accurate information, especially when the mass spectra have low signal-to-noise or the goodness-of-fit function used in the algorithm contains multiple local

extrema. These effects are illustrated for simulated DPPC-MSP1E3D1 Nanodisc mass spectra in Figure B8 with signal-to-white-noise ratio 20:1 and a realistic increase in average lipid content of 10 lipids per charge state. Despite the low signal-to-noise ratio, nearly perfect reconstruction of the exact zero-charge mass spectrum is achieved using the FT approach alone, and dramatic improvement of reconstructions using UniDec is observed when the charge state range is limited using output from the FT method and a Fourier-filtered mass spectrum is used as input.

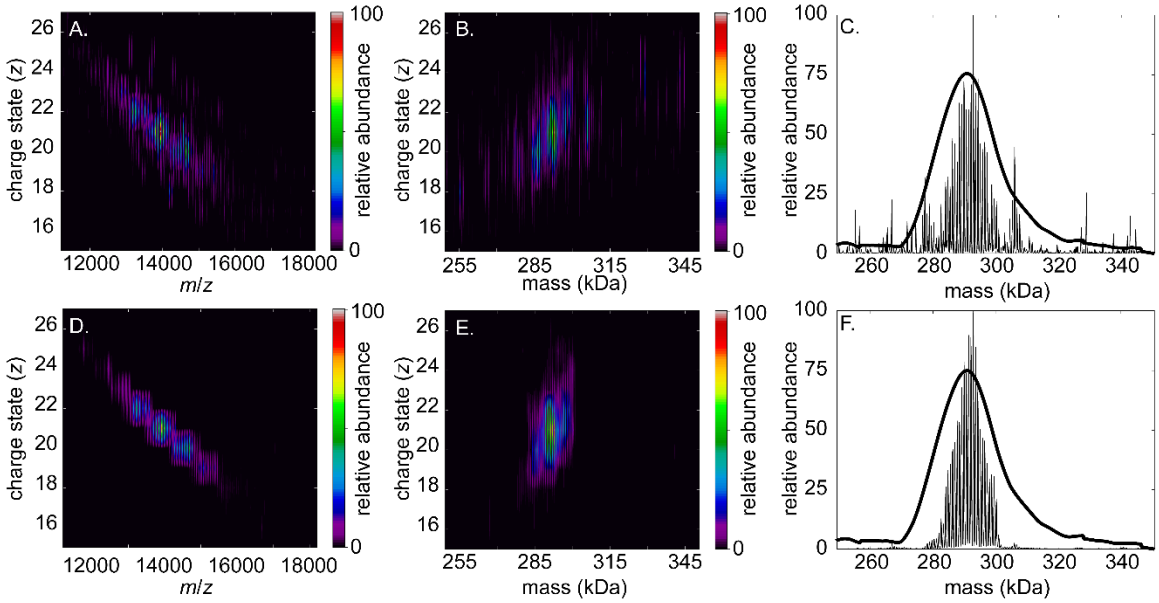


Figure 18. Deconvolved QTOF mass spectra vs. charge (left), mass vs. charge (center) and zero-charge mass spectra (right) of DPPC-MSP1E3D1 Nanodiscs determined using UniDec. Mass spectral data are the same as in Figure 13. A, B, and C result from using "naïve" input parameters for subunit mass, charge state range, peak width, and total mass range, whereas D, E, and F result from using values obtained from the FT-based method (see text). Smooth black trace in C and F represents zero-charge mass spectrum reconstructed using FT approach (see Figure 16D)

Shown in Figure 18 are two deconvolutions for the experimental DPPC-MSP1E3D1 Nanodisc mass spectrum from Figure 13 using UniDec's Bayesian deconvolution algorithm. The initial parameters for both analyses were chosen as follows: a charge state range of 15-27+ and a mass range for the entire complex of 250-

350 kDa. These values were chosen to simulate a typical scenario in native MS in which the approximate charge state and mass range of a population of assembly ions can be estimated, but the repeated sub-unit mass is unknown. Figures 18A, 18B, and 18C show results from using the Bayesian algorithm with no initial guess for the lipid mass and an initial peak FWHM, in *m*/*z*, of 8.9, which is obtained from the mass spectrum using the peak width tool in UniDec after performing an initial baseline subtraction. Overall, the results are similar, but not identical, to those from the FT-based approach (see Figure 15 and Table 1), though a number of high-intensity orphan peaks, presumably artifacts, are found in the zero-charge mass spectrum with these input parameters at masses below 270 kDa and above 300 kDa.

Figures 18D, 18E, and 18F show results from UniDec for the same mass spectrum, but using input mass spectra and parameters determined from the different FT-based approaches described above: The Fourier-filtered spectrum as the input spectrum, a subunit mass of 733 Da determined using the FT method and a peak FWHM of 13 determined from P(*k*). With these input parameters, the spurious peaks in the zero-charge mass spectrum are nearly eliminated and charge states and mass estimates for the ion population agree much more closely with results obtained using the FT analysis above. The zero-charge mass spectrum reconstructed with UniDec is somewhat narrower than the one reconstructed using the FT approach, consistent with the non-overlapping charge-state-specific mass spectra reconstructed in UniDec (Figure 18D). This demonstrates that information obtained from the Fourier analysis can be used as input parameters in Bayesian fitting algorithms to improve the quality of the results. Similar results were

found for UniDec processing of the Orbitrap and FT-ICR spectra, as shown in Figures B9 and B10

Conclusions

As the composition and structure of larger and more heterogeneous assembly ions probed with mass spectrometry continue to increase in complexity, analysis of their mass spectra demands more powerful approaches for assigning charge state, mass, and stoichiometry. Developing instruments with increased resolving power is one potential strategy, but access to these higher resolution instruments may not always be available for some laboratories, and it may be desirable to perform tandem experiments, such as ion mobility spectrometry and surface-induced dissociation, which are not widely available in high-resolution instruments. Furthermore, resolution limitations associated with current instruments will inevitably arise again for future instruments as the size and complexity of ions continues to increase, even with the advantages of charge-stripping or –reduction techniques.[42, 87] The results presented here illustrate how the Fourier Transform-based analysis strategy can allow one to work within current instrumental limitations when studying assembly ions with repeated subunits. Both experimental and computational results indicate that reliable and self-consistent charge state, subunit mass, and subunit stoichiometry determinations can be made using this strategy, especially if Fourier-domain peaks are well resolved (with inter-peak spacing at least ~1.5 times the sum of adjacent peak widths) and have signal-to-noise of at least 10:1. This deconvolution method can be used alone or in combination with Bayesian deconvolution techniques, for which it can provide input parameter values, such as charge state range, subunit mass, mass range and mass spectral peak width, to improve results. Furthermore,

information learned through FT-based analysis can help avoid potential errors in analyses associated with mass spectral domain deconvolution algorithms, including error attributed to curved baseline subtraction or parameter estimation. The general principles discussed here are valid for many other types of assembly ions with repeated subunits, including biomolecular assemblies, polymers, and inorganic cluster ions, due to the analogous form of their nESI mass spectra.

While higher harmonic frequencies can indeed be useful signal components, one challenge of the Fourier method is the potential of overlapping higher harmonic frequencies. When this overlap occurs, it can be difficult to uniquely identify charge states from the Fourier spectrum, and thus presents a serious limitation for ions with low charge states, a high number of different charge states, or mass spectra with sufficiently high resolution, such as polymer mass spectra.[4, 13] Thus, in order for the Fourier method to be useful for a variety of different molecules, it is necessary to expand the method such that this limitation can be overcome. Presented in the next chapter is one such expansion, where a 2-dimensional Fourier approach is developed that can separate overlapping higher harmonic frequencies.

CHAPTER IV

SHORT-TIME FOURIER TRANSFORMATION APPROACH FOR OVERLAPPING FREQUENCY DOMAIN SIGNALS

Includes co-authored material from:

> Cleary, S. P. and Prell, J. S. "Liberating Native Mass Spectrometry from Dependence on Volatile Salt Buffers by Use of Gábor Transform" *ChemPhysChem* **2019,** *20,* 519-523.

Introduction

Nanoelectrospray ionization mass spectrometry (nESI-MS) can be a powerful tool for analyzing the native structure and stoichiometry of large, multi-subunit complexes in the gas phase.[21, 29, 51, 60, 61, 63-69] However, for ions exhibiting a large degree of subunit stoichiometric polydispersity, such as long chain polymers[4, 5, 17] or lipoprotein Nanodiscs,[1, 13, 24, 25, 60, 78] analysis of the nESI mass spectrum can be very challenging due to the presence of tens to hundreds of closely spaced peaks. Analytes in native nESI-MS are almost universally characterized using buffers containing volatile salts such as ammonium formate or ammonium acetate[105-109] because the use of more common biochemical buffers with non-volatile salts, such as Tris/NaCl or HEPES/NaCl buffers, results in biological ions with many salt adducts as well as salt cluster ions that severely congest the mass spectrum.[35, 110-113] This reliance of native nESI-MS on volatile salt buffers can be problematic for native analytes that do not survive buffer exchange or may have different compositions or structures in different buffers.[20, 114, 115] Adduction of cations in positive ion mode (or anions in negative ion mode) can increase polydispersity of the biological ion of interest, making accurate mass measurement challenging, and overlapping signal from large, non-volatile salt cluster ions can further obfuscate analyte

signal. In a previous chapter, I demonstrated that Fourier transformation (FT) of the mass spectrum can greatly facilitate the analysis of samples exhibiting high degrees of subunit or adduct polydispersity.[13, 25] However, in some cases, analyte signals in congested mass spectra elude characterization even with the FT approach due to overlap in both the mass spectrum and frequency domain. This can occur when signal in the Fourier frequency domain from an ion of interest overlaps with signal from other analytes containing the same subunit, for example, a highly sodiated protein and a distribution of salt cluster ions with a signal at the same Fourier frequency. Here, I introduce a Gábor transform-based method that enables analysis of large biomolecular and polymer ions in extremely congested mass spectra dominated by signals from non-specific salt or small polymer cluster ions.

Methods

Sample Preparation. All water used in the data presented here was ultrapure (18.0 MΩ/cm impedance) water. Aqueous anthrax lethal factor N-terminal subunit (LFn) solution was prepared containing 10 μM LFn, 100 mM sodium chloride, 20 mM tris(hydroxymethyl)aminomethane (Tris), and 0.5 mM ethylenediaminetetraacetic acid (EDTA) at pH 7.4. β-lactoglobulin (β-Lac) was purchased from Sigma-Aldrich and was solubilized without further purification. Aqueous β-Lac solutions were prepared in a buffer containing 10 μM β-Lac, 100 mM sodium chloride, 20 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES), and 0.5 mM EDTA at pH 7.4. Polyethylene glycol with a manufacturer-estimated average molecular weight of 10,000 Da (PEG 10k, product number 309028) was purchased from Sigma-Aldrich (St. Louis, Missouri) and

used without further purification. PEG 10k samples were prepared as 0.5 mg/mL solutions in water for mass spectrometry experiments.

Mass Spectrometry. Native nanoelectrospray ionization (nESI) mass spectra were acquired on a Synapt G2-Si Quadrupole Time-of-Flight (QTOF; Waters-MS Technologies, Manchester, UK) mass spectrometer using a static nESI source. nESI emitters were prepared by pulling borosilicate capillaries (i.d. 0.78 mm, Sutter Instruments) to a tip i.d. of ~1 μm using a Flaming-Brown P-97 micropipet puller (Sutter Instruments, Novato, CA, USA). For each sample, ~3−5 μL of solution was loaded into an emitter, which was placed approximately 1-2 mm from the entrance of the mass spectrometer. A platinum wire inserted into the solution was used to apply an electrical potential of 0.6-1.0 kV relative to instrumental ground to initiate electrospray. The ion source was equilibrated to ambient temperature. Mass spectra were collected in Sensitivity mode to maximize signal-to-noise. Argon gas was introduced into the Trap at a flow rate of 5 mL/min in all experiments. Trap/Transfer collision energy was set to 10/5 V, respectively, unless otherwise stated. All mass spectral data were processed without smoothing using MassLynx v. 4.1 (Waters Corp., Milford, MA, USA).

Nuclear Magnetic Resonance Spectroscopy. A ^{13}C-NMR spectrum was acquired on a Bruker AV-III HD 500 MHz NMR spectrometer equipped with a 5 mm Prodigy BBO Cryprobe at 308 K. PEG 10k was solubilized in deuterated dimethyl sulfoxide to a concentration of 10 mM. All data processing of the ^{13}C-NMR spectrum, including integration of identified ^{13}C peaks, was performed using the software MestReNova version 12.0.4 from Mestrelab. The number-average degree of polymerization was determined as the ratio of the integrated internal monomer peak area

(at δ 70.3 ppm) to the average of the integrated terminal peak areas (at δ 72.8 and 60.8 ppm). The number-average molecular weight was determined as the molecular weight of a PEG polymer with this degree of polymerization, and the uncertainty (1 standard deviation) was estimated by propagating integration uncertainty and white noise.

Computational Work. All data processing involving a Gábor transformation was performed using the open-sourced Matlab package "The Large Time-Frequency Analysis Toolbox" (LTFAT) available for download here: http://ltfat.github.io/. All one-dimensional Fourier transform-based analysis was performed using the Prell group's home-built program, iFAMS (interactive Fourier Analysis for Mass Spectra) version 5.

Theory

Gábor transformation (GT) is a short-time Fourier transformation signal processing technique that can reveal where specific frequencies are localized in the mass spectral signal. In GT, the signal is windowed with a Gaussian prior to Fourier transformation, and this process is repeated with the window translated by equally-spaced intervals. The windowed Fourier-transformed data are then plotted as a 2d "spectrogram" with the horizontal axis indicating the position of the window in the *m*/*z* domain, the vertical axis being the frequency, *k* (= z/Δm), and the complex amplitude equal to the Gábor Transform coefficient as a function of these two variables. Thus, the GT spectrogram shows where in the mass spectrum signals of a particular frequency are localized. GT of a mass spectrum can be ideally represented mathematically:

$$G(\alpha,k) = \int s\left(\frac{m}{z}\right) g\left(\frac{m}{z} - \alpha\right) e^{-ik\left(\frac{m}{z}\right)} dk$$

where:

$$g\left(\frac{m}{z}\right) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{\left(\frac{m}{z}\right)^2}{(2\sigma)^2}}$$

α = the *m/z* center of the window

k = frequency

For an experimental (i.e. discretely sampled) mass spectrum, this is:

$$G(w,v) = \sum_{\frac{m}{z}} s\left(\frac{m}{z}\right)\ g\left(\frac{m}{z} - wdW\right)\ e^{-ik\left(\frac{m}{z}\right)v/V}$$

Where:

dW = spacing of window positions in *m/z*

V = frequency channels

The number of distinct window positions used in the GT limits the number of frequency samples that can be made without redundant representation of the signal in the spectrogram. Specifically, the number of points in the spectrogram, L, is equal to the number of window positions in the mass spectrum, W, times the number of distinct frequencies ("channels"), V at predefined intervals.

$$L = W \times V$$

In the information-limited case, i.e. where L is as large as possible without redundant information being contained in the spectrogram, L is equal to the number of data points in the mass spectrum. This means there is a trade-off between the information-limited resolution in frequency and window position (i.e., effective mass spectral resolution in GT).

Although V is adjustable, the bandwidth of frequencies remains constant because it is determined by the range of m/z values in the mass spectrum, and total number of data points in the mass spectrum:

$$\text{bandwidth} = (\text{number of data points})/(m/z \text{ span})$$

The Nyquist Theorem asserts that, in order to avoid aliasing, the sampling rate must be at least twice as high as the frequency of interest in the mass spectrum. Therefore, in order to resolve two adjacent bands in the GT spectrogram, V must be chosen such that the change in frequency one hopes to capture (i.e., the spacing between charge state peaks in the Fourier domain) is larger than the sampling rate. Sampling rate is determined by:

$$\text{sampling rate} = \text{bandwidth}/V$$

Because of this, oversampling, a scenario in which L is greater than M, is sometimes practically unavoidable. Since oversampling can lead to artifacts in the GT spectrogram, it is advised that the rate of oversampling be kept low, where:

$$\text{rate of oversampling} = L/W$$

Results

Limitations of a 1D Fourier transformation approach. As a demonstration of limitations associated with ordinary 1D Fourier analysis, shown in Figure 19 are the mass spectrum and corresponding Fourier spectrum of two disperse samples exhibiting the problem described in the introduction: Highly sodiated anthrax lethal factor N-terminal subunit (LFn, Figure 19A) and long-chain (~10 kDa average molecular weight, as reported by the manufacturer) polyethylene glycol ("PEG 10k") polymer (Figure 19B). While Fourier transformation of the mass spectrum facilitates sodium adduct and polymer subunit mass determination using previously described methods,[4, 13] it is difficult to uniquely identify charge states for both analytes (LFn and the long-chain polymer ions) in the Fourier spectrum due to the presence of many other peaks. Without confident

charge state determination, determining accurate mass information for these ions from either the mass spectrum or Fourier spectrum is very difficult.

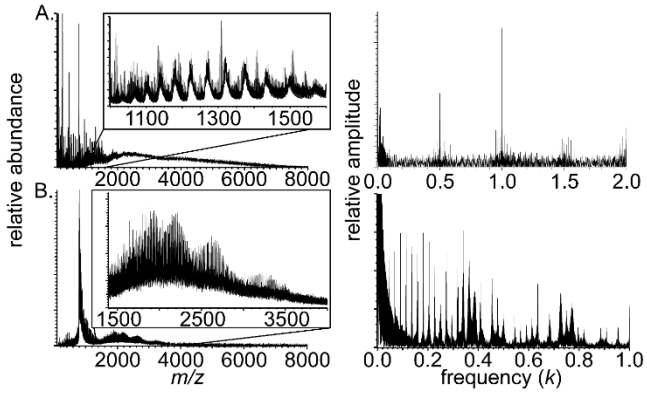


Figure 19. Native mass spectra (left) and FT spectra (right) of sodiated LFn ion (A) and polyethylene glycol polymer (manufacture-reported M_{avg} 10,000 Da) (B). Insets in mass spectra illustrate the adduction profile for the LFn ion, and the signal component for the long-chain polymer.

The LFn mass spectrum is complicated by multiple factors, including adduction of sodium to the protein, the presence of sodium chloride clusters, and a high baseline that is typical of spectra electrosprayed from a buffer containing a high concentration non-volatile salt (for comparison, a sample of the same protein sprayed from an ammonium acetate buffer, which is much more common in native ESI-MS, is shown in Appendix C, Figure C1). These three signal components each span the entire mass spectrum, making it difficult to uniquely assign peaks to specific components in the mass spectrum. This effect can be seen in inset of Figure 19A, which shows the partially unfolded charge states for LFn. (21-31+). Near-native (13-20+) and compact native (9-12+) charge states are not readily identified, however, as the high baseline due to salt clusters obscures the signal due to these charge states of the protein. Fourier transformation produces a spectrum with a prominent signal at integer multiples of frequency (k (= $z/\Delta m$) = 0.5). This signal arises from $Na^+(NaCl)_n$ clusters due to the distribution of chloride isotopes, which are spaced by ~2 Da within the mass spectral signal for each cluster. Although the signal at k = 0.5 can be assigned to chloride-

containing species that are singly-charged, signals at higher multiples of $k = 0.5$ cannot be uniquely assigned from the Fourier spectrum alone, because they can arise from multiply-charged salt clusters, harmonics of the $k = 0.5$ peak, or a mixture of both. Moreover, these intense salt cluster peaks strongly mask the weaker signals of other ion types. For example, LFn ions with multiple Na adducts are expected to contribute signals at integer multiples of $k \approx 1/22 = 0.045$ (the inverse of the mass shift due to the displacement of a proton by a sodium ion), but such signals are difficult to distinguish in the Fourier spectrum.

Similarly, the PEG 10k mass spectrum (Figure 19B) contains several apparent distributions of ion signal throughout the spectrum, including a highly abundant distribution around 1000 *m/z* and smaller overlapping distributions between 1500-3500 *m/z* (see inset). Each distribution contains tens to hundreds of closely spaced peaks, making it difficult to assign specific peaks to specific charge states and distributions. FT of this signal results in a spectrum with many peaks that appear equally spaced, an effect that has been described in detail elsewhere.[4, 13] Briefly, each peak is spaced by $\Delta k = 0.0227$, which is the frequency corresponding to the subunit mass of PEG (44.026 Da). A peak is present at nearly every multiple of this frequency up to the 18th harmonic. Due to the possible presence of overtones of the fundamental peak, which can overlap with the fundamental peak of all other charge states, it is difficult to determine the charge states present in the mass spectrum. This ambiguity frustrates further analysis of the mass distribution of these ions.

Gábor Transformation can Separate Overlapping Signals in the Fourier and Mass Spectrum. In contrast to ordinary FT, Gábor transformation (GT) is a "short-time"

Fourier transform that can reveal where specific frequencies are located in the mass spectral signal. As with other short-time Fourier transform methods, this is achieved by sliding a window with a user-defined shape (in the case of GT, a Gaussian) across the signal prior to Fourier transformation. The windowed, Fourier-transformed data are then plotted as a 2-dimensional "spectrogram" with the horizontal axis indicating the position of the center of the window in the *m*/*z* domain and the vertical axis being the frequency, *k*. The spectrogram shows where in the mass spectrum signals of a particular frequency are localized. Importantly, because GT spectrograms are 2-dimensional, frequency signals that overlap in the ordinary Fourier spectrum can often be separated in the GT spectrogram.

To illustrate how GT can reveal information in congested mass spectra, Figure 20 shows the mass spectrum, FT spectrum, and GT spectrogram of an idealized data set representing three different types of heterogeneous ion populations: a single-charge-state population of ions separated by constant mass, an ion population representing non-specific clusters that increase in average *m/z* as the charge state increases, and a multiple-charge-state population of ions that does not significantly change mass distribution as charge state increases (which represents a highly adducted protein or polymer population with a charge-state-independent mass distribution). For illustrative purposes, the mass spectrum is modeled such that the different mass spectral signal components are well separated in *m/z*, and the subunit mass is the same for each component. Similar to the FT spectrum in Figure 19B, the FT spectra in Figure 20 contains many peaks that are difficult to uniquely attribute to one specific mass spectral signal component. However, GT produces a spectrogram where the different signal components are spread out, and

appear as collections of ovoid bands in the GT spectrogram, and the 2-dimensional arrangement of these bands is different for the three types of ion population. Bands corresponding to the fundamental frequencies for each type of population are boxed in white, and each subsequent group of bands found at higher frequencies represents higher harmonic bands. Importantly, while these different frequency components overlap in the Fourier domain, the different series are separated and easily identified in the GT spectrogram. This result illustrates how GT can be used to identify different ion signals that overlap in both the mass spectrum and FT spectrum.

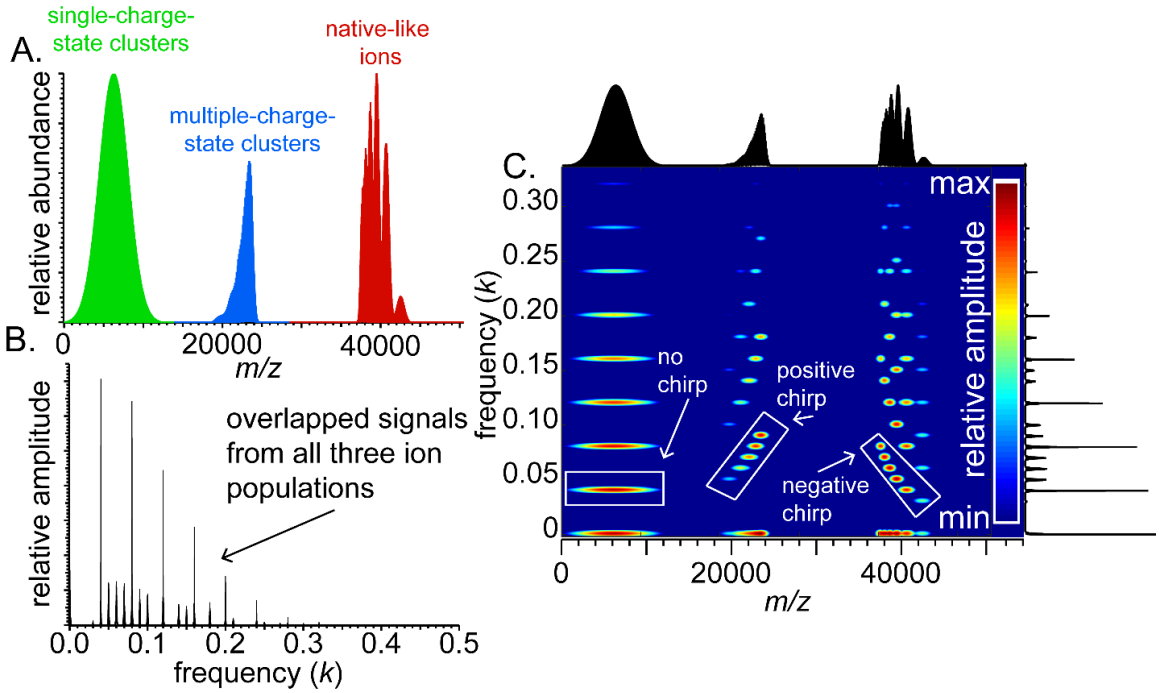


Figure 20. Gábor transformation (GT) of a modeled mass spectrum. Shown are the mass spectrum (A) and corresponding Fourier spectrum (B) and GT spectrogram (C). Mass spectrum was modeled to show three different mass populations, including a single-charge-state population (green), a multiple-charge-state population that increases in mass with increasing charge (blue), and a multiple-subunit population that does not significantly increase in mass (red, see text). Black spectra above and to the right of the GT spectrogram represent the mass spectrum (top) and Fourier spectrum (right) of the original data set.

The shape, or "chirp", of the different groups of bands (by analogy with the auditory phenomenon of time-varying frequencies in, e.g., a bird's chirp), can also reveal information about a family the type of ions they represent. That is, ion populations representing clusters of a single charge state and fixed subunit mass have no chirp, clusters with mass that tends to increase with charge state have positive chirp, and ion populations that do not change mass significantly with increasing charge state have negative chirp. (Indeed, native-like ions with charge-independent mass distributions will typically have GT signals with frequency inversely proportional to *m*/z, i.e., a hyperbolic negative chirp).

Gábor Transform can Separate Highly Sodiated Protein Signals from a High Baseline Associated with Salt Cluster Ions. Due to this ability to separate the different mass spectral and frequency signal components, GT was applied to the mass spectra shown in Figure 19. Figure 21 shows the GT spectrogram of the sodiated LFn using sampling parameters that were chosen to optimize subunit mass resolution while keeping the oversampling rate low (see methods). The spectrogram contains prominent features, including three high-intensity vertical strips seen at low *m/z*, horizontal strips (bands) that are seen at every multiple of $k = 0.5$, and a low-intensity pattern that has a negative chirp (shown in the inset of Figure 20). The highly abundant vertical strips seen at *m*/*z* ~122, 144, and 265 correspond to protonated Tris, sodiated Tris, and sodiated Tris dimers, respectively. The frequency sampling rate chosen for the spectrum in Figure 21 is too low to clearly resolve these frequencies, and the narrow width of the mass spectral signals for these ions results in the spreading out of their associated frequency information across the entire frequency domain. The horizontal bands corresponding to integer multiples of k =

0.5 are the same signals described previously for the Fourier spectrum in Figure 19A, which result from sodium chloride clusters.

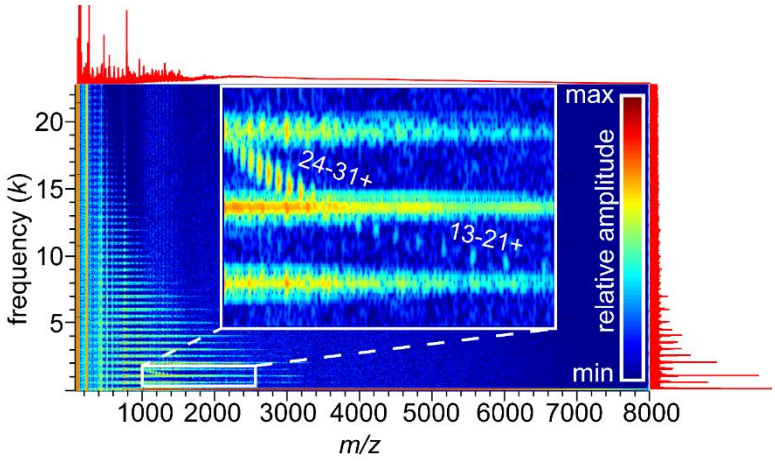


Figure 21. GT spectrogram of LFn sprayed from a NaCl/Tris buffer. Inset shows the different charge states identified for the sodiated protein. Red traces represent the mass spectrum (top) and Fourier spectrum (right).

The low-intensity negatively chirped group of bands found between frequencies 0.5 to 1.5 (see inset, Figure 21) is signal from the LFn ion population. The frequency spacing of these bands corresponds to the mass of sodium minus the mass of a proton and arises from the adduction of different numbers of sodium ions to the protein. Using the charge states identified from the frequencies of these bands, the measured accurate mass of LFn was determined to be 32,900 ± 200 Da, which is consistent with the measured accurate mass (32,723 ± 1 Da) found using the compact native LFn ions observed upon nESI out of ammonium acetate solution (Figure C1), albeit slightly larger due to more extensive adduction of sodium ions. Importantly, while these ions are difficult to characterize using either the mass spectrum or Fourier spectrum alone, they are much more easily identified in the GT spectrogram, especially the near-native charge states (13-21+). A reconstructed spectrum of the isolated, sodiated LFn peaks, as well as a comparison of this reconstruction to the mass spectrum acquired from ammonium acetate solution (Figure C1), are shown in Figures C2 and C3, respectively. Interestingly, the GT analysis is possible due to sodium ion adduction,

which is the origin of the observed frequencies in the frequency domain. Thus, while the adduction of cations is often thought of as a limitation in native mass spectrometry, here it is advantageous. This result suggests the enticing possibility of using GT to analyze native bimolecular samples electrosprayed from more common buffers in biochemistry, such as Tris/NaCl, HEPES/NaCl, and other buffers containing nonvolatile salts, especially for proteins or complexes that do not survive buffer exchange intact or adopt different structures in different buffers. (Shown in the Appendix C is a similar analysis for β-lactoglobulin in HEPES/NaCl buffer, including an ammonium acetate spectrum, Figure C4; HEPES/NaCl buffer spectra, Figure C5; and GT spectra, Figure C6).

Gábor Transform can Separate Overlapping Higher Harmonic Frequencies. Shown in Figure 22 is the GT spectrogram of the PEG 10k polymer from Figure 19B (The inset in Figure 22 shows the data between *m*/*z* 1,150 and 3,200). The low-intensity, broad bands that extend across most of the inset exhibit little chirp and represent short-chain polymer and polymer cluster contaminants with charge states of at least 3+. By contrast, the series of more intense bands with negative chirp (see Figure 22 inset) are readily assigned to charge states 4+ through 8+ for the fully-formed, long-chain polymer, which increases only slightly in mass with increasing charge state (m_{avg} = 12,920 and 13,610 Da for the 4+ and 8+ charge states, respectively). Signals representing the 1st through 4th harmonics for these long-chain polymer ions are observed, consistent with the relatively high resolution of individual peaks in the mass spectrum.

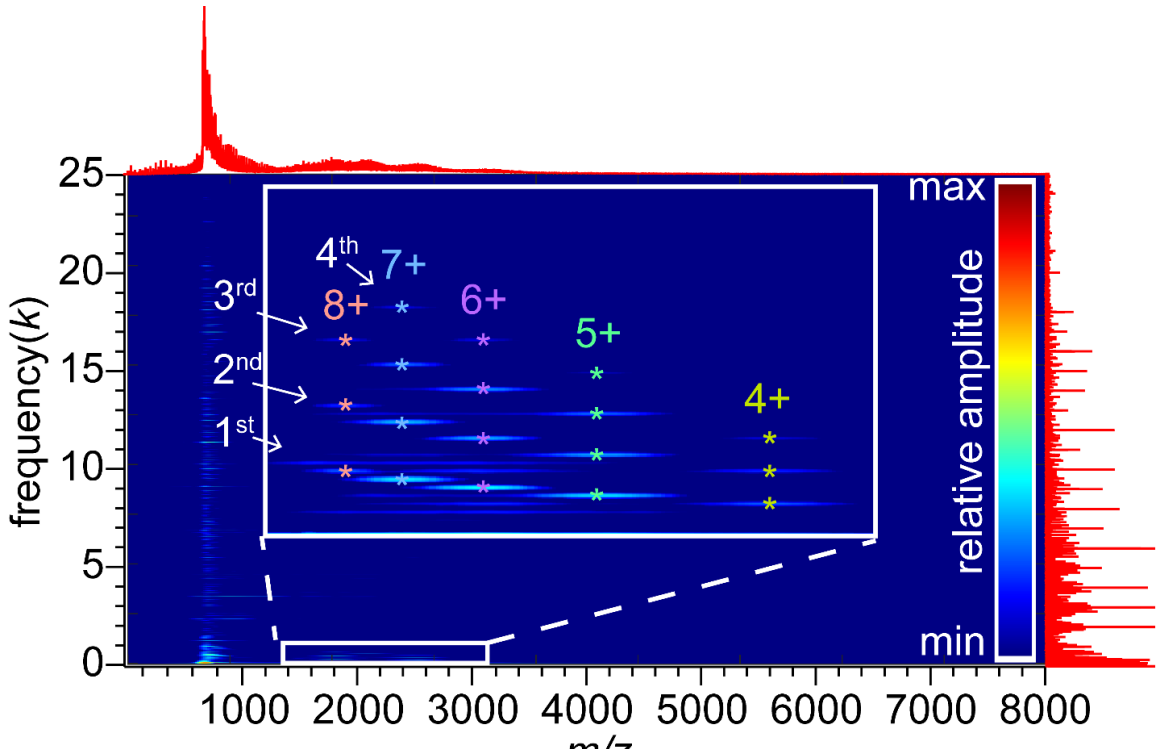


Figure 22. GT spectrogram of polyethylene glycol 10 kDa ions. Inset shows the different signal components for the long-chain polymer ions, which are labeled by color (★) for the specific charge state. Red traces represent the mass spectrum (top) and Fourier spectrum (right) of the original data set. Ordinal numbers in white indicate harmonics of long-chain polymer ion signal.

Ion signal from a GT spectrogram can be isolated and transferred back to the mass spectral domain by applying inverse GT to the isolated 2-dimensional data. This can be quite useful for polymers, as it allows one to characterize the average weight and polydispersity of the fully formed complex. A "zero-charge" spectrum using both the fundamental and higher-harmonic data for each charge state (Figure C7), as well as mass and stoichiometry statistics for all the charge states (Table S1), are reported in Appendix C using the GT of the PEG 10k spectrum. As in FT-based analysis,[25] using all available harmonics results in greater accuracy of the reconstructed signal. Interestingly, while the manufacture reported average mass is 10,000 Da, Gábor analysis demonstrates that the average mass of the long-chain polymer ion population is about ~30% larger (13,200 ±

300 Da). Furthermore, the charge-state-specific mass spectra for odd charge states are shifted by ~22 Da from those of even charge states. This observation is attributed to the addition of an extra sodium on average for each additional charge state, which is very close to half the mass of the PEG subunit mass (44.02 Da). It should be noted that the mass reported here is the number-average mass of the long-chain polymer ions, isolated from the large population of short-chain polymer contaminants. Conversely, methods such as gel permeation chromatography, which is used by the manufacturer to estimate the average mass, or other methods, such as nuclear magnetic resonance (NMR) spectroscopy, may include smaller, short-chain polymer contaminants in the average mass measurement. These methods can result in a somewhat lower value than that reported here for the isolated long-chain polymers. For example, a ^{13}C-NMR spectrum of the same PEG sample was acquired in deuterated dimethyl sulfoxide (Figure 2). Integration of the measured NMR peaks indicates a number-average polymer mass of ~10.0 ± 0.4 kDa, likely due to the inclusion of short-chain polymer contaminants in the NMR signal.

Conclusions

Analytes exhibiting high degrees of polydispersity, such as the highly sodium adducted protein and long-chain polymer samples presented here, are often considered challenging to analyze by conventional ESI-MS analysis. This observation has led researchers to use methods that can reduce the complexity of the spectra. Popular methods include desalting the protein prior to analysis, such as the standard practice in native ESI-MS of buffer-exchanging protein samples into volatile salt buffers such as

ammonium acetate,[105-109] or using smaller nESI emitter tips,[116] as recently demonstrated by Williams and coworkers. Other helpful methods include deconvolution algorithms[13, 14, 25, 47] or ion mobility separation, which has been previously shown to be effective for polymer mass spectrometry analysis.[5, 6, 17] However, many commercially available instruments, such as Orbitrap and Fourier-Transform Ion Cyclotron Resonance mass spectrometers, do not yet have ion mobility capabilities.

By contrast, the method presented here is suitable for a wide variety of common MS instrumentation and can be used to identify mass, charge state, and subunit mass for challenging ions such as those described above, even within an intense background of interferent ions. Because adduction of multiple adduct ions or additional polymer subunits encodes a frequency into the mass spectrum, this frequency can be found by Fourier transformation and further localized within the mass spectrum and extracted from background signal using GT. The chirp of different mass populations in the GT spectrogram can also be readily used to distinguish signals from non-specific (e.g., salt) clusters and native-like ions. This presents the attractive possibility to use GT to characterize the mass spectra of many types of samples that have been heretofore intractable with existing analysis methods, for example, proteins that do not survive buffer exchange into a volatile salt buffer such as ammonium acetate. Future directions of this research include determining the influence of buffer choice on the conformation and stoichiometry of biomolecules and biomolecular complexes in native MS.

Along with the previous two chapters, this chapter thus presents a variety of different approaches that can be used to probe the native stoichiometry of highly disperse molecules made up of a single subunit using native mass spectrometry. However, a

lingering question may be can the Fourier analysis method be used if a heterogeneous mass population contains dispersity from two different subunits, for example, a Nanodisc assembled with two lipids. An article[85] published in 2016 investigated this question, and found that for Nanodiscs made up of two lipids, the subunit mass as measured by Fourier transformation was representative of the average amount of each lipid contained within the Nanodiscs. While this was an enticing result, the explanation as to why this measurement was found was not discussed and thus it was assumed that the stoichiometry of a variety of different molecules with dispersity from two different subunits could be measured by this same method. In the next chapter, I explore the theoretical reasoning as to why this result was found, using molecules with a variety of different dispersity profiles.

CHAPTER V

THEORETICAL AND PRACTICAL APPROACHES TO PROBING MULTI-SUBUNIT STOICHIOMETRY WITH MASS SPECTROMETRY USING FOURIER TRANSFORMATION

Introduction

Multi-subunit, polydisperse complexes, including many protein complexes, protein complexes with bound ligands, and membrane protein-lipid complexes are essential for the function and structure of all organisms.[49, 117-120] Native electrospray ionization mass spectrometry (ESI-MS) is a powerful technique that can probe the native stoichiometry of these biomolecular complexes, because it allows many complexes to remain intact while being transferred into the gas phase, and the stoichiometry of the complex can be determined if the masses of the individual parts are known.[2, 7, 13, 21, 29, 66, 69, 70, 121] Developments in MS methodology and spectrum analysis tools have made it possible to characterize stoichiometry distributions even for polydisperse populations of complexes.[1, 13, 14, 25, 122] However, in cases where the polydispersity is exceptionally high or different for different charge states, or includes polydispersity arising from two different subunits, the mass spectrum can be exceedingly difficult to analyze, as the ESI source produces overlapping charge state distributions forming a superposition of tens to hundreds of peaks.

Multiple alternative methods have been developed to facilitate the analysis of highly polydisperse samples, including Fourier Transform (FT)-based algorithms,[4, 13, 25, 121, 122] *m/z*-domain deconvolution algorithms,[14, 47, 88] and macromolecular mass defect

analysis (akin to Kendrick mass defect analysis for polymers).[80, 123, 124] Commonly used mass spectral-domain deconvolution algorithms typically require user optimization of the algorithm and input of accurate guesses at several parameters describing the mass distribution. Macromolecular mass defect analysis can be used to characterize samples with multiple subunits with sufficiently high mass resolution. Conversely, FT-based mass deconvolution can often be achieved with little to no initial parameter guessing even when resolution is relatively low, as was recently shown for a sample of intact lipoprotein Nanodiscs sample containing over 300 lipids[25] and for bacterial toxin complexes embedded in nearly intact detergent micelles.[121] Furthermore, the Marty and co-workers recently demonstrated that FT methods can quantify the bulk subunit composition for Nanodisc ions containing two different lipid types.[85] In that study, an iterative FT approach was used to determine an average subunit mass from the entire Nanodisc mass spectrum, and it was revealed that this measurement reflected the bulk lipid stoichiometry of the mixture used to synthesize the Nanodiscs.

While it is perhaps intuitive that FT can be used to determine bulk composition of polydisperse ions containing two or more types of subunits, the appearance of the FT spectrum and the way it is to be interpreted depend strongly on the mathematical form of the composition distribution of the ions. That is, in general, one must know that an ion population follows a particular statistical description in order to interpret the corresponding FT spectrum correctly. Extreme examples that require very different interpretation include a mixture of two or more single-subunit populations and a population whose stoichiometry distribution can be described as a convolution of the stoichiometry distributions of the separate subunits. Here, we illustrate how such ion

distributions can be treated with FT deconvolution methods and demonstrate the theoretical utility and limitations of these methods. Finally, we show how these methods can be used to distinguish between a mixed-subunit population and a mixture of two single-subunit populations, even with the same bulk subunit composition, shedding light on the mechanisms by which the ions are assembled. These results should be readily generalizable to many complex samples, from copolymers to protein-ligand complexes.

Methods

Sample Preparation. Nanodiscs containing palmitoyloleoylphosphatidylcholine (POPC), dipalmitoylphosphatidylcholine (DPPC) or both were prepared according to a method adapted from that of Sligar and co-workers.[52, 55] Briefly, all lipids were purchased from Avanti Polar Lipids as 5 mg/mL solutions in chloroform, dried until opaque with dry nitrogen gas, and re-suspended to a final concentration of 50 mM in a pH 7.4 aqueous buffer containing 100 mM sodium cholate (Sigma-Aldrich), 20 mM Tris (Bio-Rad), 100 mM sodium chloride, and 0.5 mM ethylenediaminetetraacetic acid (EDTA). Membrane scaffold protein MSP1D1 (Sigma-Aldrich) was reconstituted in pH 7.4 aqueous buffer (20 mM Tris, 100 mM sodium chloride, 0.5 mM EDTA, 0.01% sodium azide) to a concentration of ~200 μM. Lipid suspensions were mixed with MSP1D1 solutions and additional buffer to a final concentration 50 μM in MSP1D1 and appropriate lipid concentrations. These concentrations were: 3.38 mM for pure POPC, 4.5 mM for pure DPPC, and 4.22 mM, 3.94 mM, and 3.66 mM for 75/25, 50/50, and 25/75 percent mixtures of DPPC/POPC respectively. The concentrations for mixtures were chosen based on a weighted measurement for pure lipid ratios (90-1 and 67.5-1 for DPPC and

POPC respectively to MSP1D1). The mixing of lipid stocks was performed after initial solubilization of pure lipid stocks by mixing samples in the appropriate ratios and sonicating the solution for an additional 30 minutes. Samples were incubated for 1 hr at 20 °C, room. Nanodisc self-assembly was initiated by cholate removal through both dialysis and the use of BioBeads SM-2 (Bio-Rad), where the BioBeads were placed in the dialysis buffer (20 mM Tris, 100 mM sodium chloride, and 0.5 mM EDTA) and constantly stirred through use of a stir bar. Samples were left overnight and the Nanodisc-containing supernatants were removed from dialysis and buffer-exchanged into 200 mM ammonium acetate (Sigma-Aldrich) using Micro Bio-Spin 6 columns (Bio-Rad) immediately before MS analysis.

Poly(ethylene glycol)-block-poly(propylene glycol)-block-poly(ethylene glycol) (PEG-PPG-PEG) was purchased from Sigma-Aldrich and used without further purification. The polymer was solubilized to a concentration of 0.5 mg/mL in ultra-pure (18.0 MΩ) water. Bovine ubiquitin (Ubq) was purchased from Sigma-Aldrich and used without further purification. Aqueous Ubq solutions were prepared containing 25 μM Ubq and 200 mM ammonium acetate.

Mass Spectrometry. All mass spectrometry analysis was performed with a Synapt G2-Si ion mobility mass spectrometer (Waters Corp.) using a static nanoelectrospray ionization (nanoESI) source. NanoESI emitters were prepared by pulling borosilicate capillaries (ID 0.78 mm, Sutter Instruments) to a tip ID of ~1 μm using a Flaming-Brown P-97 micropipette puller (Sutter Instruments). For each sample, ~3-5 μL of solution was loaded into an emitter, which was placed approximately 3-5 mm from the entrance of the mass spectrometer. A platinum wire inserted into the solution

was used to apply an electrical potential of 0.6-1.0 kV relative to instrumental ground to initiate electrospray. For Nanodisc mass spectra, samples were sprayed from a 200 mM ammonium acetate solution. Trap collisional voltage was set to 100 V while transfer was set to 5 V. Both sample cone and source offset were both set to 25. All spectra were averaged for 20 minutes. For polymer mass spectra, samples were sprayed from ultrapure (18.0 MΩ) water. Trap collisional voltage was set to 10 V while transfer was set to 5 V. Both sample cone and source offset were both set to 25. All spectra were averaged for 5 minutes. For Ubq spectrum, sample was sprayed from a 200 mM ammonium acetate solution. Trap collisional voltage was set to 10 V while transfer was set to 5 V. Both sample cone and source offset were both set to 25.

Computational Work. All FT-based analysis was performed using the Prell group's home-built program, iFAMS (interactive Fourier Analysis for Mass Spectra) v. 5.2.

Theory

Mathematical Descriptions of Ion Populations Containing Two or More Types of Repeated Subunit. The presence of a second polydisperse subunit population can significantly complicate the analysis of a mass spectrum, and it should be emphasized that there are a very large number of ways in which two repeated subunits can be distributed. A variety of these different ways are shown in Figure 23 in the form of modeled mass spectra, along with their corresponding Fourier spectra. Perhaps the simplest way is the mass spectrum arising from two non-interacting, independent single-

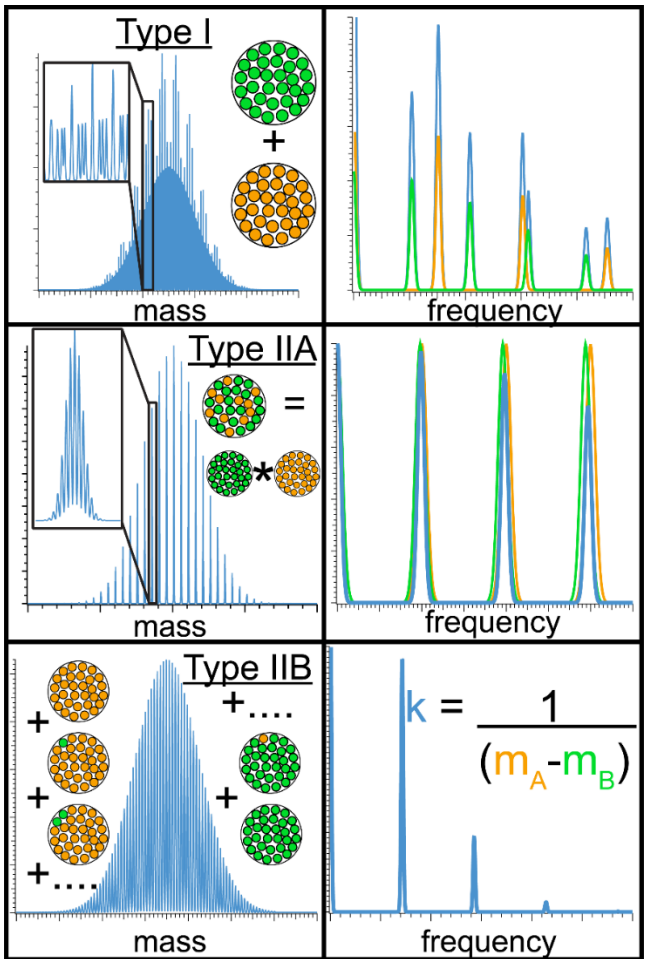


Figure 23. Mass spectra and the corresponding Fourier spectra for differing forms of multi-subunit polydispersity. Shown above are modeled spectra and the corresponding Fourier spectra for the different forms of multi-subunit polydispersity described in the theory section. Blue spectra in the Fourier spectra correspond to the Fourier transform of the multi-subunit spectra, while the other colored Fourier spectra (orange, green) correspond to relative single subunit populations.

subunit ion populations, which is shown in the upper spectrum of Figure 23, and is described by

$$s_{tot}(m/z) = s_1(m/z) + s_2(m/z)$$

$$S_{tot}(k) = S_1(k) + S_2(k)$$

where $S_{tot}(k) = FT[s_{tot}(m/z)]$, $S_1(k) = FT[s_1(m/z)]$, and $S_2(k) = FT[s_2(m/z)]$. That is, both the mass spectrum and its Fourier transform (due to the linearity of FT) can be simply described as a superposition of the individual single-subunit population. Such a population can arise simply from mixing together two purified samples that do not exchange subunits on the timescale of the experiment; we refer to this as a "Type I" population throughout the manuscript and note that it can be easily generalized for multiple subunits by similar reasoning.

In sharp contrast, many other common types of heterogeneous ion populations can arise from samples in which two or more types of subunit are assembled simultaneously

into complexes, such as random copolymers or phospholipid Nanodiscs. While there are a very large number of ways in which this can occur, two particularly relevant extremes are 1) the stoichiometry distribution of the resulting ion population can be described as the convolution of the underlying stoichiometry distributions of the individual subunit types, which is shown in the middle spectrum of Figure 23, the total number of subunits in the ions is essentially fixed, such that the number of subunits of each type always add up to the same number, which is shown in the bottom spectrum of Figure 23. We refer to these as "Type IIa" and "IIb" populations throughout the manuscript. A common example of a Type IIa population is a linear block copolymer formed by growing the second polymer block off one end of a pre-existing population of the first block. In this case, for each charge state *Z*,

$$s_{tot}(m/z) = s_1(m/z) * s_2(m/z)$$

and therefore,

$$S_{tot}(k) = S_1(k) \bullet S_2(k)$$

That is, the FT of the total ion spectrum is the product of the underlying mass spectra for each subunit type with charge state *Z*. For more than two subunits, the total mass spectrum is again the convolution of the underlying subunit mass spectra, and the corresponding total FT spectrum is the product of the underlying FT spectra.

Common examples of Type IIb populations are protein assemblies of fixed size with two or more substituents, such as hemoglobin tetramers, or isotopic distributions.

For each charge state and total ion size, the mass spectrum abundance follows a binomial distribution with respect to the average probability that an ion contains either subunit, i.e.

$$s_{tot}(m/z) \propto \binom{n_{tot}}{n_A} p_A^{n_A} (1 - p_A)^{n_{tot} - n_A}$$

for each m/z corresponding to an ion containing n_A of subunit A and n_B (= $n_{tot} - n_A$) of subunit B with probabilities p_A and p_B (= $1 - p_A$), respectively. This distribution is equivalent to an n_{tot}-fold autoconvolution of a distribution having just two peaks at m/z positions m_A/Z and m_B/Z with relatively abundances p_A and p_B, respectively. Thus the FT of s(m/z) is simply the n_{tot} power of the FT of this simple two-peak spectrum. (Incidentally, this idea can be generalized quite easily to compute the isotope distribution of any molecule from the natural abundances of its constituent atoms according to their individual stoichiometries.) *A priori*, one would expect that Type IIa mass distributions should often be much broader and have denser peaks spacing than Type IIb mass distributions, but it can be difficult to distinguish the two from one another or from Type I mass distributions simply by visual inspection of the mass spectrum, especially with poor resolution. However, Fourier transform can readily distinguish between these two possibilities, as the frequency of a Type IIb molecule is simply the inverse of the mass of the heavier subunit minus the mass of the lighter subunit.

$$k = \frac{1}{m_A - m_B} \; where \; m_A > m_B$$

Results and Discussion

Rationale for Selected Examples. The question arises how, without having to perform the time-consuming analysis of all the peaks in the mass spectrum, one can determine whether a multi-subunit ion population belongs to Type I, Type IIa or IIb, or some other type, and what statistics (such as average subunit composition, average ion size, and polydispersity) one can determine straightforwardly from the mass spectrum, even with poor resolution. To demonstrate how Fourier transform can discern between the different types of multi-subunit populations and what can be learned from the analysis, the following examples are used: Two single-lipid Nanodiscs mixed together (Type I), a mixed-lipid Nanodisc (Type IIA), a block copolymer (Type IIA), and an isotope distribution (Type IIB).

Type I Mixture of Single-Subunit Nanodiscs (Superposition). For samples that are a simple mixture of two non-interacting analytes (Type I mixtures), the distribution of peaks in the mass spectrum is a superposition of the underlying mass spectra belonging to each type of analyte. Due to the linearity of FT, the Fourier spectrum is therefore a superposition of the underlying Fourier spectra of each type of analyte, and the positions of peaks in the Fourier spectrum do not change for the mixture. An example of the mass spectrum of a Type I mixture and its corresponding Fourier spectrum, acquired under moderately activating conditions where ~20 lipids have been dissociated from the Nanodiscs,[13] is shown in Figure 24. Two Nanodisc samples were separately prepared using different lipids (POPC and DPPC), and mixed in approximately equal molar ratios immediately before analysis with nESI-MS. While it is difficult to see that the mass spectrum is a roughly 1:1 superposition of the mass spectra of pure POPC and DPPC

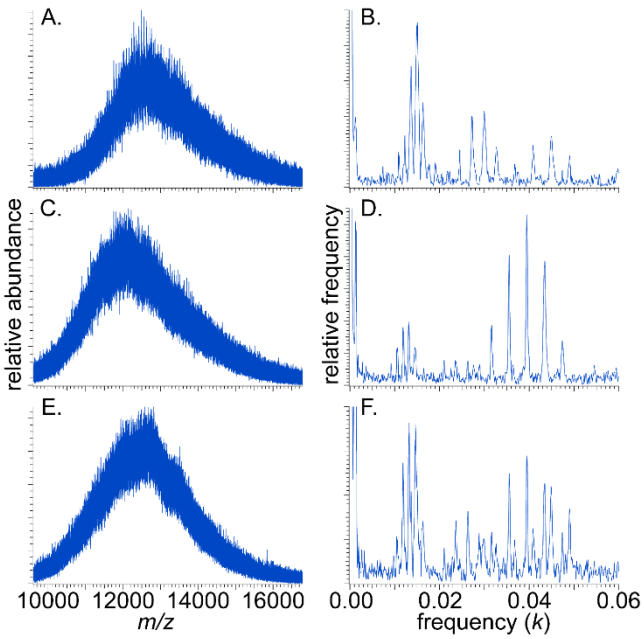


Figure 24. Mass spectra and corresponding Fourier spectra of single-lipid Nanodisc mixed together. Shown above are the mass spectra and corresponding Fourier spectra of Nanodiscs assembled with DPPC (A and B), POPC (C and D), and both samples mixed together (E and F) in equimolar ratios. The result of mixing together the two single-lipid Nanodiscs is a super position of both the mass and Fourier spectra, which is seen in E and F.

Nanodiscs, acquired under identical instrumental conditions, due to the relatively low resolution, it is clear that the corresponding Fourier spectrum is such a superposition. Notably, such an analysis is possible even in the absence of good peak resolution in the Fourier spectrum due to the presence of multiple harmonics for each Nanodisc type. This is particularly evident for the 3rd harmonic frequency peaks, where the same characteristic frequencies for the pure Nanodisc Fourier spectra are seen in the mixture Fourier spectrum. That is, they do *not* occur at the abundance-weighted average frequency of the POPC and DPPC Nanodisc peaks, in sharp contrast to results for Nanodiscs prepared from a 1:1 bulk mixture of POPC and DPPC lipids (see below) and to previous results for mixed-lipid (i.e., Type IIa) Nanodiscs.[85]

Type IIa Mixed-Lipid Nanodiscs (Convolution of Composition Distributions). Figure 25 shows the mass spectrum (A) and corresponding Fourier spectrum (B) for a Nanodisc sample prepared from a bulk mixture of POPC and DPPC in a molar ratio of 25 % to 75 % POPC:DPPC, which is expected to result in Nanodiscs that

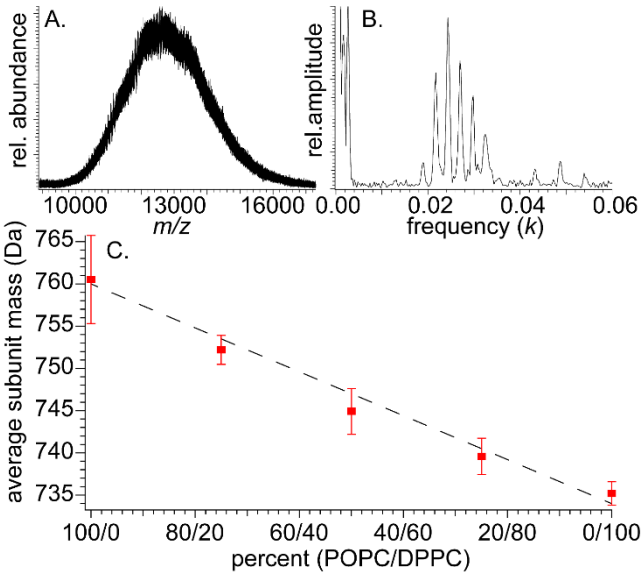


Figure 25. Mass spectra and corresponding Fourier spectra for Nanodiscs assembled with two different lipids. Shown above is the mass spectrum (A) and Fourier spectrum (B) of Nanodiscs assembled with 75% DPPC and 25% POPC. In contrast to the mixture of two single-lipid Nanodiscs mixed together, Nanodiscs assembled with 2 different lipids produces only a single series of peaks in the Fourier spectrum, whose spacing corresponds to the variance-weighted average subunit mass. This result is consistent across multiple different ratios, as is seen in C.

contain a mixture of both lipids. *A priori*, because the assembly mechanism for Nanodiscs is not well understood, it is not clear whether this should result in Type IIa or IIb (or another type) of heterogeneity. Type IIa heterogeneity would result from a Nanodisc assembly mechanism than can be approximate by two independent Poisson processes, i.e., in which the two different types of lipids add to the growing Nanodisc via pseudo-first-order kinetics with rate constants dependent on the bulk concentrations of the lipids. In this case, the distribution in the total number of lipids in the Nanodiscs should be a convolution of the (Poisson) distributions in the number of each lipid type contained in the Nanodisc. Because the convolution of two independent Poisson distributions is another Poisson distribution, and because the mean of a Poisson distributions is equal to its variance, this mechanism should result in Nanodiscs with approximately equal mean and variance in the total number of lipids. As has been previously observed in multiple studies of intact MSP1D1 Nanodiscs using native MS, the mean number of lipids is often ~120-180 with a standard deviation of ~10-15 (i.e., a

variance of ~100-225),[1, 13, 51] so this mechanism is consistent with these observations. Similarly, larger MSP1E3D1 Nanodiscs typically contain ~270-350 lipids with a variance of ~225-400,[25] these numbers increasing for smaller lipids. A somewhat tedious but straightforward derivation shows that Fourier-domain peaks for such a Type IIa sample (formed by convolution of two monomodal subunit distributions, e.g., Poisson or Gaussian distributions, with variance-proportional means) will be located at the bulk concentration-weighted average frequency of the two subunit types. Emphatically, if the variances of the subunit distributions are not simply proportional to their means, the frequency of peaks in the Fourier spectrum share *no simple relationship* to their bulk concentrations.

Figure 25C demonstrates this principle, in that frequencies in the Fourier spectra for all of the bulk concentration ratios tested for POPC and DPPC do indeed fall very close to the bulk concentration-weighted average frequencies expected for a sample with the Type IIa heterogeneity described above, i.e., with variance-proportional means. This result is consistent with previous observations by Marty and coworkers for a variety of different lipid mixtures.[85] We conjecture that this relationship therefore strongly supports a Nanodisc formation mechanism that proceeds by independent, pseudo-first-order addition of lipids, resulting in Nanodisc compositions that are largely kinetically trapped and not at equilibrium.

Type IIb mixed-lipid Nanodiscs (hypothetical equilibrated Nanodiscs limited by bilayer size). Type IIb heterogeneity, in stark contrast, would be associated with a mechanism whereby fully formed Nanodiscs equilibrate with one another by exchange of lipids before MS analysis, constrained by the number of lipids that can be accommodated

into the fully formed Nanodiscs. In this case, the Nanodisc composition distribution would be expected to be nearly binomial, and the variance in the distributions of each lipids would be nearly identical independent of the bulk lipid concentration ratios. For example, if Nanodiscs are tightly constrained to contain ~140 total POPC or DPPC lipids, any lipid within a Nanodisc that is not a POPC must be DPPC, thus the POPC and DPPC distributions within the Nanodiscs must have equal variances independent of the bulk lipid composition. As described in the Theory section, such Type IIb heterogeneity results in a Fourier spectrum with a fundamental frequency equal to the inverse of the difference between the two masses (26 Da for POPC and DPPC) *regardless* of bulk composition. The sharp contrast between this result and the experimental results shown in Figure 25 lends further support to the conclusion that these Nanodiscs are not at equilibrium, but form by pseudo-first-order kinetics and have a kinetically trapped composition distribution. Together with the Type I mixture results described above, these results indicate that the FT method can straightforwardly distinguish between Type I and Type IIa heterogeneity, even in cases where the bulk average subunit composition is the same.

Type IIa triblock copolymer (convolution of composition distributions).

Similar reasoning described above for Type IIa mixed-lipid Nanodiscs can be used to understand the mass spectra and corresponding Fourier spectra of many block copolymers formed by growing each successive block off pre-existing, polydisperse blocks. In many realistic cases, the exact size of each block will be uncorrelated with the exact size of the other component blocks, resulting in an overall copolymer length distribution that is the convolution of the block length distributions of each monomer

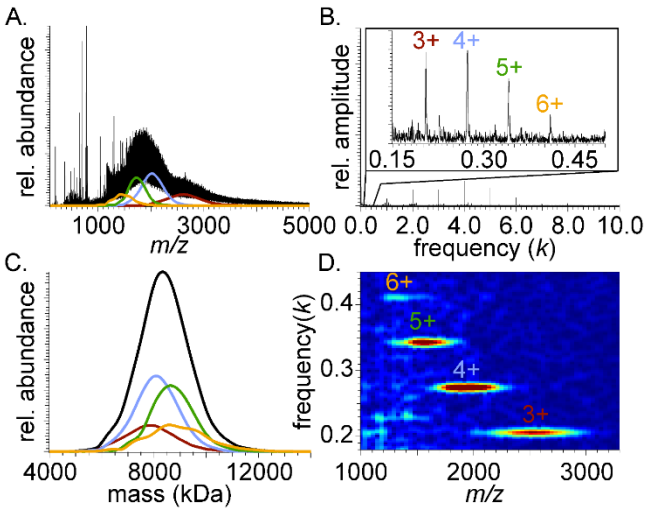


Figure 26. Deconvolved mass spectrum of poly(ethylene glycol)-block-poly(propylene glycol)-block-poly(ethylene glycol) (PEG-PPG-PEG, 8.4 kDa average molecular weight). The mass spectrum and reconstructed charge state specific distributions are shown in A. Colors correspond to the identified charge state peaks in the inset of the Fourier spectrum shown in B. The deconvolved zero-charge spectrum is shown in C, with the color of the trace corresponding to the appropriate charge states, and the black trace showing the summation of all the charge states. Shown in the D is the Gábor spectrogram, which is used to reconstruct the charge state specific envelope functions shown in A.

type. Again, peaks in the Fourier spectrum should appear at the variance-weighted average of the individual monomer frequencies. A nESI mass spectrum for a PEG:PPG:PEG linear triblock copolymer sample is shown in Figure 26. Intriguingly, the first visible peak at positive frequency in the corresponding Fourier spectrum occurs near the respective 3rd and 4th harmonic of the 3+ charge state corresponding to PEG and PPG, respectively. The reason that lower-frequency peaks, closer to the PEG and PPG fundamental frequencies, are not observed is that the polydispersity of the constituent PEG and PPG blocks is so large that the corresponding Fourier-domain peaks are too narrow to overlap for any lower-order harmonics and happen to nearly coincide for these harmonics (and multiples thereof). From the observed Fourier-domain frequencies, it can be concluded that the ratio of variances for the PEG and PPG monomer distributions is ~5:1. As described above, this value does not directly report the average composition of the triblock copolymer. This value is remarkably close to the manufacturer-reported composition of

3.8:1 PEG:PPG based on ^{1}H-NMR analysis. While a number of factors likely contribute to this discrepancy, purification of the PPG core before growth of the PEG blocks could

Table 4. Total polydispersity characterization of Poly(ethylene glycol)-block-poly(propylene glycol)-block-poly(ethylene glycol) 8.4 kDa using the Fourier transform method

adjusted subunit mass of PEG (Da)	adjusted subunit mass of PPG (Da)	avg. subunit mass measurement (Da)	% variance PEG	% variance PPG
14.675	14.510	14.646	82.38	17.62
mean of total population (Da)	mean of PEG population (Da)	mean of PPG population (Da)		
8370 ± 970	6620 ± 860	1740 ± 400		

be responsible for reducing the variance:mean ratio of the PPG block length distribution relative to that of the PEG blocks. This would result in a slightly higher PEG:PPG ratio being determined using the FT method with the mass spectrum based on the assumption of variance-proportional means. The results of this analysis is summarized in Table 2.

Type IIb isotope distributions (multinomial abundance distributions). As is well known, for an ion with chemical formula $A_{nA}B_{nB}C_{nC}\ldots$, where A, B, C, … are elements and nA, nB, nC, … are the stoichiometries of each element (assumed to follow their natural or other well-defined isotope abundances), the relative abundance of the isotopomer of the ion with total mass

$$m_{tot} = \left(n_{Ai}m_{Ai} + n_{Aj}m_{Aj} + n_{Ak}m_{Ak} + \cdots\right) + \left(n_{Bi}m_{Bi} + n_{Bj}m_{Bj} + n_{Bk}m_{Bk} + \cdots\right) + \left(n_{Ci}m_{Ci} + n_{Cj}m_{Cj} + n_{Ck}m_{Ck} + \cdots\right) + \cdots$$

where n_{Ai} and m_{Ai} are the respective stoichiometry and mass of the ith isotope of element A, etc.,

is

$$A = \binom{n_A}{n_{Ai}, n_{Aj}, n_{Ak} \ldots} p_{Ai}^{nAi} p_{Aj}^{nAj} p_{Ak}^{nAk} \ldots \binom{n_B}{n_{Bi}, n_{Bj}, n_{Bk} \ldots} p_{Bi}^{nBi} p_{Bj}^{nBj} p_{Bk}^{nBk} \ldots \binom{n_C}{n_{Ci}, n_{Cj}, n_{Ck} \ldots} p_{Ci}^{nCi} p_{Cj}^{nCj} p_{Ck}^{nCk} \ldots$$

where the terms in brackets are multinomial coefficients, and p_{Ai} is the relative bulk abundance of the ith isotope of element A, etc. The FT spectrum of this isotope distribution is simply the product of nA of the FT of element A's bulk isotope abundance spectrum times nB of that for element B, times nC of that for element C, etc. Because the total number of atoms in the ion is fixed, this is in fact an example of Type IIb heterogeneity for the ion in question. Figure 27 shows an isotope-resolved experimental mass spectrum of $(\text{ubiquitin} + 6\text{H})^{6+}$ as well as its corresponding predicted mass spectrum from the Fourier analysis. The Fourier spectrum is predicted in this manner using ubiquitin's chemical formula, $C_{378}H_{629}N_{105}O_{118}S_1$, and natural bulk isotope abundances.

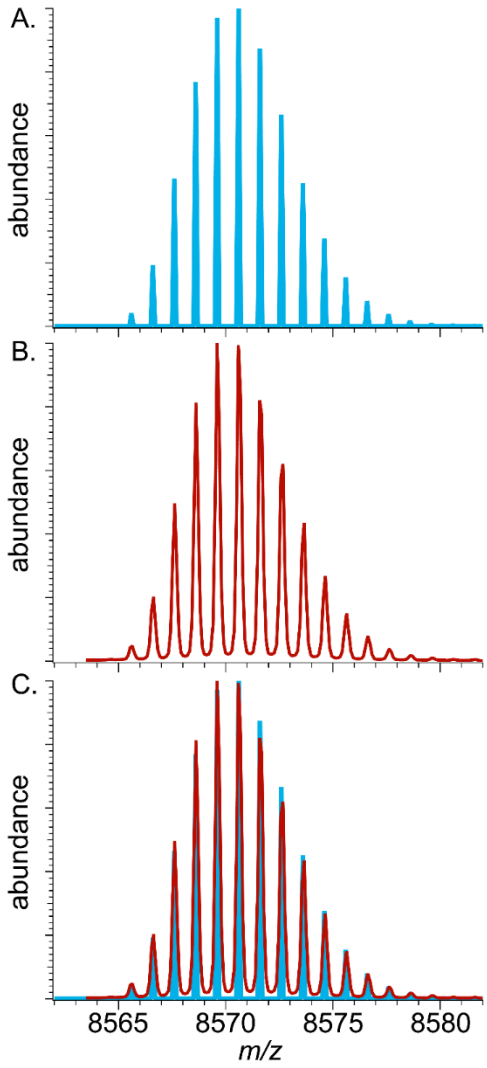


Figure 27. Estimated ubiquitin isotope distribution. Shown above is an estimated isotope distribution (A), a measured isotope distribution (B) and the two spectra overlayed (C).

The Fourier spectrum of even this relatively large ion can be computed extremely quickly (< 1 second) from pre-computed Fourier spectra for its constituent atoms (C, N, O, S, and H), and the Fourier spectrum can be inverted using Fast Fourier Transform to yield the predicted mass spectrum for the ion. Especially for proteins and other large ions, this can be extremely advantageous over conventional mass spectral-domain methods relying on convolution, because pointwise products are much faster to compute than convolutions. The resolution of the predicted mass spectrum is limited only by the maximum frequency included in the pre-computed Fourier spectra for the constituent atoms, thus it trivial to predict mass spectra in this manner to arbitrary resolution (including isotope fine structure) at minimal computational expense.

Conclusions

Analytes exhibiting dispersity from two different subunits can be considerably challenging to study by conventional mass spectrometry analysis methods. This

challenge arises not only from the overlap of 10's to 100's of peaks in the mass spectrum, which is a common difficulty seen with dispersity from a single subunit, but also from the variety of different statistical scenarios for which a multi-subunit complex can form. Recently, a publication introduced a novel method based on a Fourier transformation that was able to probe the lipid stoichiometry for Nanodiscs assembled with two different lipids.[85] While this is an enticing result for the challenge presented here, the theoretical reasoning behind this method was not present in the manuscript and had not been investigated. Due to the variety of different ways a multi-subunit complex can assemble, we concluded that without the theoretical background, it was difficult to say with certainty that the FT method could be applied to a variety of different analytes. Therefore, the theory behind the method was first explored in detail here.

By treating the mass spectrum of a mixed-lipid Nanodisc as the convolution of two underlying lipid stoichiometry distributions, the FT spectrum is shown here to simply be the product (or the overlap) of the FT of the underlying lipid distributions. Furthermore, since the variance of the underlying lipid distributions determines the shape of the underlying peaks in the FT spectra, the centroid of the overlapped peak is determined by the variances of the underlying lipid distributions. It is therefore shown that the variance of these distributions determines the average subunit mass, not the mean which is postulated in the original publication. Furthermore, it is demonstrated that a necessary condition to use the average subunit mass to measure the stoichiometric amounts of each subunit is a linear dependence between the mean and variance of the two convolved subunit populations, a condition that is found to be true for mixed-lipid Nanodiscs. Further insight into Nanodisc formation can also be found through the FT

analysis of mixed single-lipid Nanodiscs, where Nanodiscs are found to not be at equilibrium resulting from no exchange of lipids, and combined with results from the mixed-lipid Nanodisc analysis, formed by pseudo-first-order kinetics and have a kinetically trapped composition distribution. Future directions of this research can include investigating other molecules exhibiting non-equilibrium behavior, and investigating lipid preferences for mixed-lipid Nanodiscs.

OUTLOOK

The research presented in this dissertation is culmination of a variety of different studies dedicated to the development of a data analysis method for measuring the average mass and dispersity of heterogeneous mass populations. Native mass spectrometry was used as the primary analytical tool for measuring these different properties, but it was often infeasible to analyze the data by conventional mass spectrometry methods due to the time consuming process of uniquely identifying each mass peak. Furthermore, while programs were already developed that could deconvolve complex mass spectra, a common theme that was found amongst the programs was that prior knowledge of the sample and/or mass spectrum was needed in order for these methods to be successful. As the analytes measured by native mass spectrometry become larger and more complex, this information will undoubtedly become difficult to know, and thus it seemed that a method was needed that did not require assumptions about the data set. Examples of analytes like this are shown throughout this dissertation.

The Fourier transformation method presented here is different way to think about mass spectral data analysis, in that instead of iteratively fitting the measured data to modeled data sets, the data is instead plotted in a different space through linear transformation. This allows the Fourier method to make no assumptions about the mass spectrum, as the analysis of the Fourier spectrum is simply an analysis of the original data, just plotted in a different way. In Chapter 2, it was demonstrated that the subunit mass, charge states, and charge-state-specific mass distributions could be determined for a variety of different molecules using the Fourier transformation method. While this first result was very useful and perhaps was most pertinent for the mass spectral analysis of

heterogeneous mass populations, there is much more information that can be learned from a Fourier spectrum. Thus in Chapter 3, a thorough discussion was presented about all of this additional information, including what can be learned from the common observation of higher harmonic frequencies. When this method was introduced for polymer mass spectra in 2004, higher harmonic frequencies were considered a major limitation for uniquely identifying charge state information, and while the information discussed in Chapter 3 can be useful to know, the fact remained that higher harmonic overlap could prevent unique identification of charge states in the Fourier spectrum. Therefore, in Chapter 4, a newer short-time Fourier transformation method was presented that can overcome this limitation. Furthermore, this method provided a means to analyze biomolecular ions from more atypical biochemistry buffers, in contrast to the accepted method of using an ammonium acetate buffer. Finally, this method was explored for ions with dispersity from two different subunits by another research group in 2016, and while the result they found was enticing, the theoretical explanation for the result was not explored in great detail. The theoretical ability of Fourier transform to analyze heterogeneous mass populations with dispersity from two different subunits was thus explored in Chapter 5, including how Fourier transform could be used to decipher between to single subunit populations in a mass spectrum versus one population with two different subunits. Future directions of this research include uses of the phase information in the Fourier transform spectrum, which is not discussed in this dissertation, as well as determining the influence of buffer choice on the conformation and stoichiometry of biomolecular complexes when using native mass spectrometry, and development of the method as quantitation tool for biotherapeutics, such as monoclonal antibodies.

APPENDIX A

SUPPLEMENTAL INFORMATION FOR CHAPTER II

Computational Implementation of Analysis Method

All FFT analysis was performed using Mathematica 10.1 (Wolfram, Inc.), although many data analysis tools can be used to produce essentially identical results. Raw time-of-flight mass spectra acquired for these experiments consist of data points that are not equally spaced in *m*/*z*. Over the *m*/*z* region of interest, data were interpolated using a cubic interpolation to produce a data set that is equally spaced ("equispaced") in *m*/*z* and contains the same number of points as the raw mass spectrum. (Non-equispaced FFT algorithms were tested using raw mass spectral data for comparison but required significantly more computational time and produced essentially identical results.) The resulting equispaced mass spectrum was transformed with Fast Fourier Transform as Implemented in Mathematica 10.1. For polymer and Nanodisc experiments, centroids of *k*-domain peaks were then found by first computing the absolute value of the Fourier spectrum, and identifying *k*-domain peaks by eye, and a trial value for k_f was computed as the difference in the *k*-values for the two most intense adjacent peaks in the spectrum. The absolute-signal-weighted average value of *k* about a peak of interest within a window centered about the peak maximum and having width k_f was then computed. For all samples other than sodiated and potassiated Ubq, an updated value for k_f was subsequently determined from the (unweighted) average of the differences between adjacent peak centroids. A final value for k_f was computed by finding the absolute-signal-

weighted average value of k within a window of width k_f centered about each integer multiple of k_f.

For Nanodisc experiments, the data within each of these windows were returned to the mass spectral domain by Inverse Fast Fourier Transform, as implemented in Mathematica 10.1. In some cases, "ringing" was observed in the charge-state-specific mass spectra due to the sinc-function-shaped IFFT of the rectangular window. The charge-state-specific average mass was computed as the charge state (z) times the abundance-weighted average m/z value of the charge-state-specific mass spectrum with abundance at greater than the maximal depth of modulation of the ringing (typically ~1/10 the maximum abundance of the raw mass spectrum). The abundance-weighted standard deviation in the mass was computed for the same data subset. These values were converted to the average and standard deviation in the number of lipids contained in the Nanodiscs by assuming a mass m_{MSP} of 24661.9 u for MSP1D1 and a base mass of ($2*m_{MSP} + z*m_H + n*18$), where m_H is the mass of a proton, and n is the number of bound adducts of nominal mass 18 u (water or ammonium cation). This assumption has been previously used in native mass spectrometry studies of Nanodiscs,[1, 51] and I typically found best agreement between our raw mass spectra for Nanodiscs with $n \approx 7$, similar to previous results from literature.[51, 125] Although this assumption does not explicitly account for other possible adductions, such as sodium cations, the total mass of these adducts is not expected to affect our characterization of the distribution in the number of lipids in the Nanodiscs.

For sodiated and potassiated Ubq, the 2-dimensional version of the FFT algorithm was used instead due to the strong overlap between peaks belonging to the Fourier

spectrum comb for each charge state. Briefly, the interpolated, equispaced mass spectra were pre-multiplied by a 50 *m*/*z*-wide (= 2σ) Gaussian window, centered at one of 31 equally-spaced *m*/*z* values spanning the entire mass spectral range. The resulting windowed mass spectra were Fourier transformed, and 2-dimensional plots were prepared with the center *m*/*z* value for the Gaussian window along one axis and *k* along the other. Centroids for the clearly visible peaks in the resulting 2-dimensional plots were computed by finding the absolute Fourier signal-weighted average value of *k* for the three *m*/*z* values of the Gaussian window center where maximum total absolute Fourier signal was observed for the *k*-domain peak of interest.

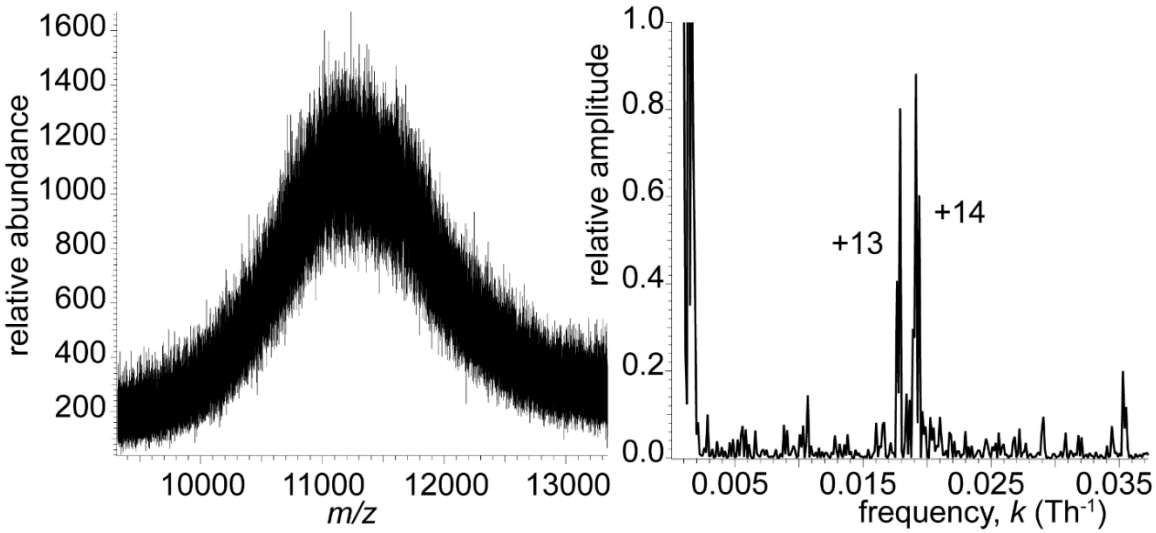


Figure A1. Mass spectrum of DPPC Nanodisc ions (left) and corresponding FFT spectrum (right) acquired under "mild" collisional activation conditions. The high-intensity peak near zero frequency in the FFT spectrum has been truncated for clarity.

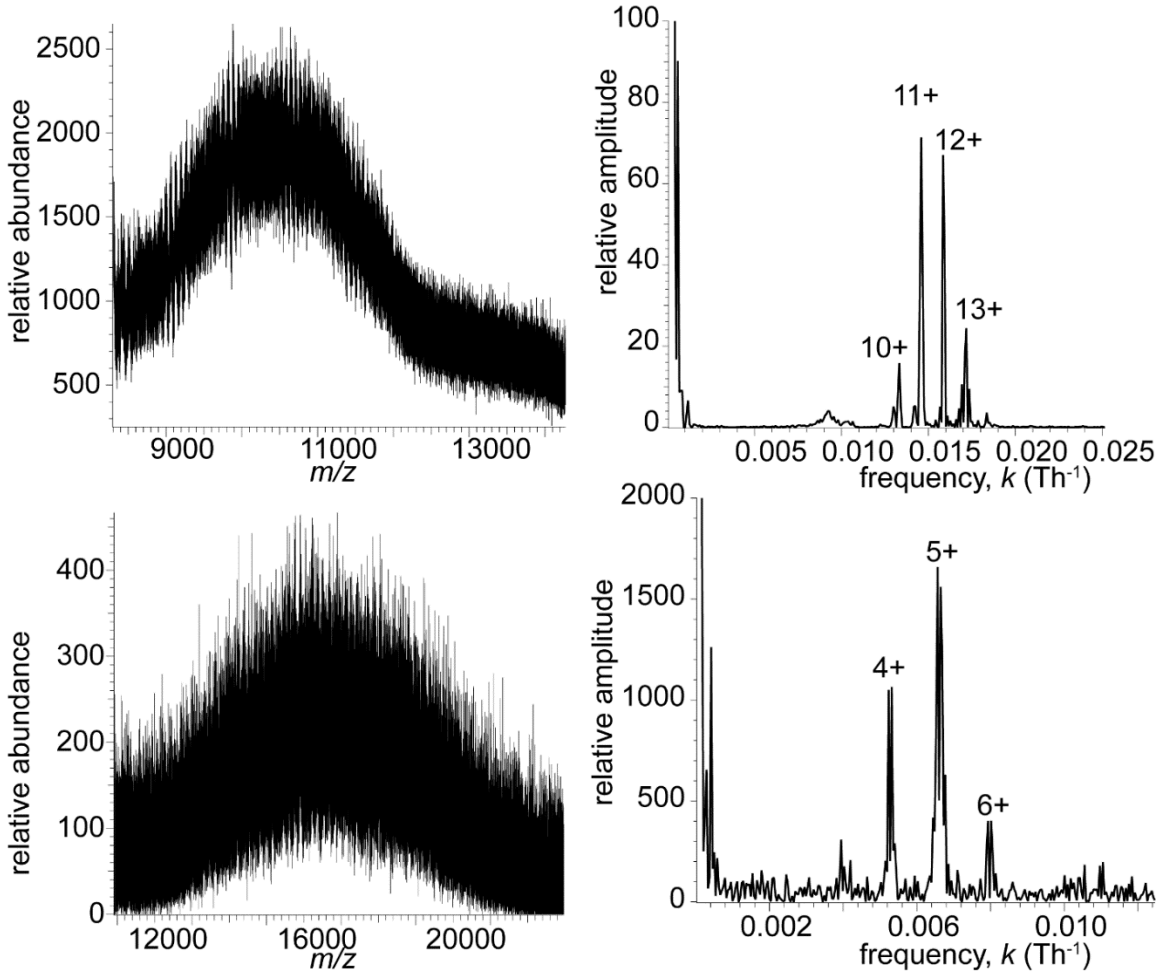


Figure A2. Mass spectra of POPC Nanodisc ions (left) and corresponding FFT spectra (right) acquired under "mild" (top) and "strong" (bottom) collisional activation conditions. The high-intensity peak near zero frequency in the FFT spectra has been truncated for clarity.

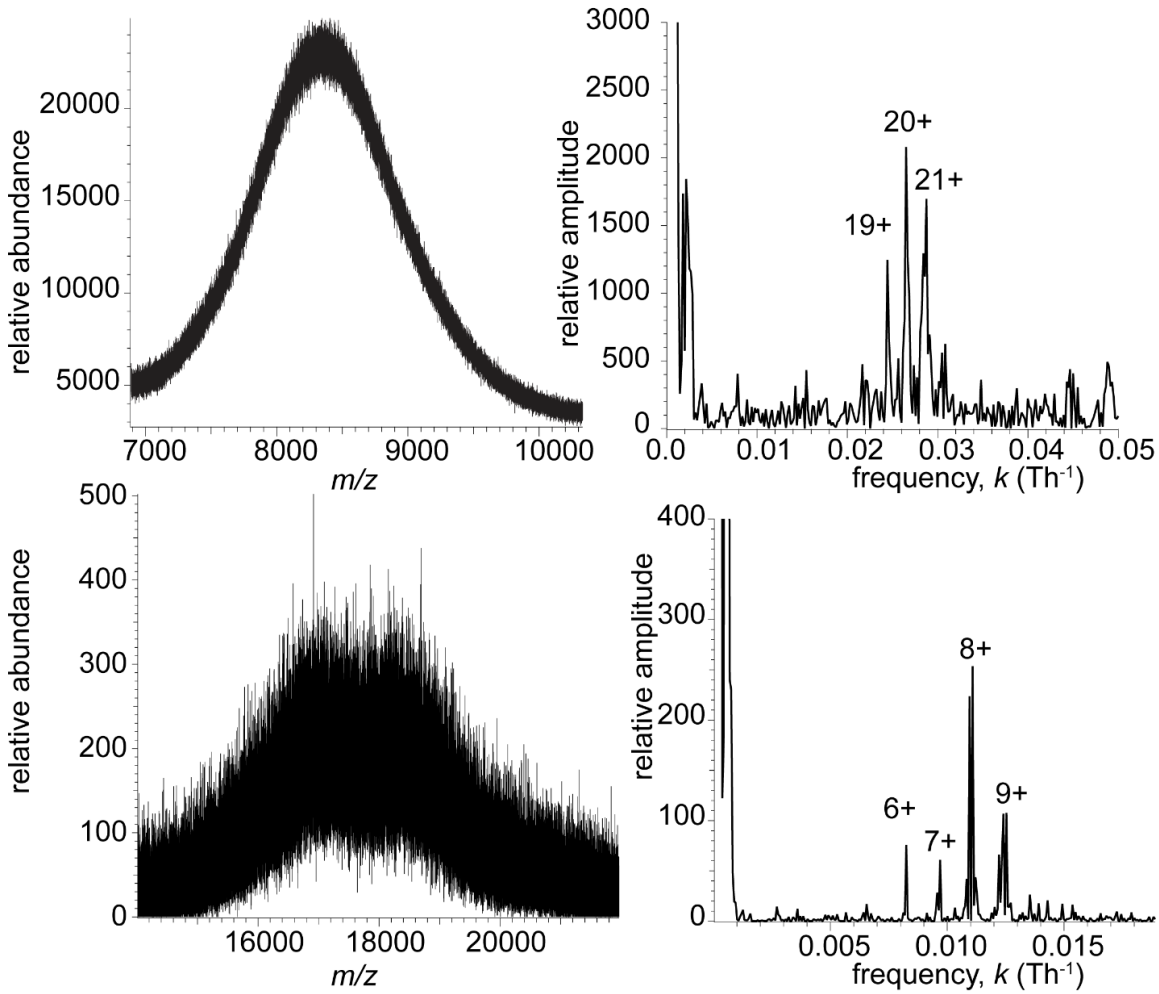


Figure A3. Mass spectra of 18:0-SM Nanodisc ions (left) and corresponding FFT spectra (right) acquired under "minimal" (top) and "strong" (bottom) collisional activation conditions. The high-intensity peak near zero frequency in the FFT spectra has been truncated for clarity.

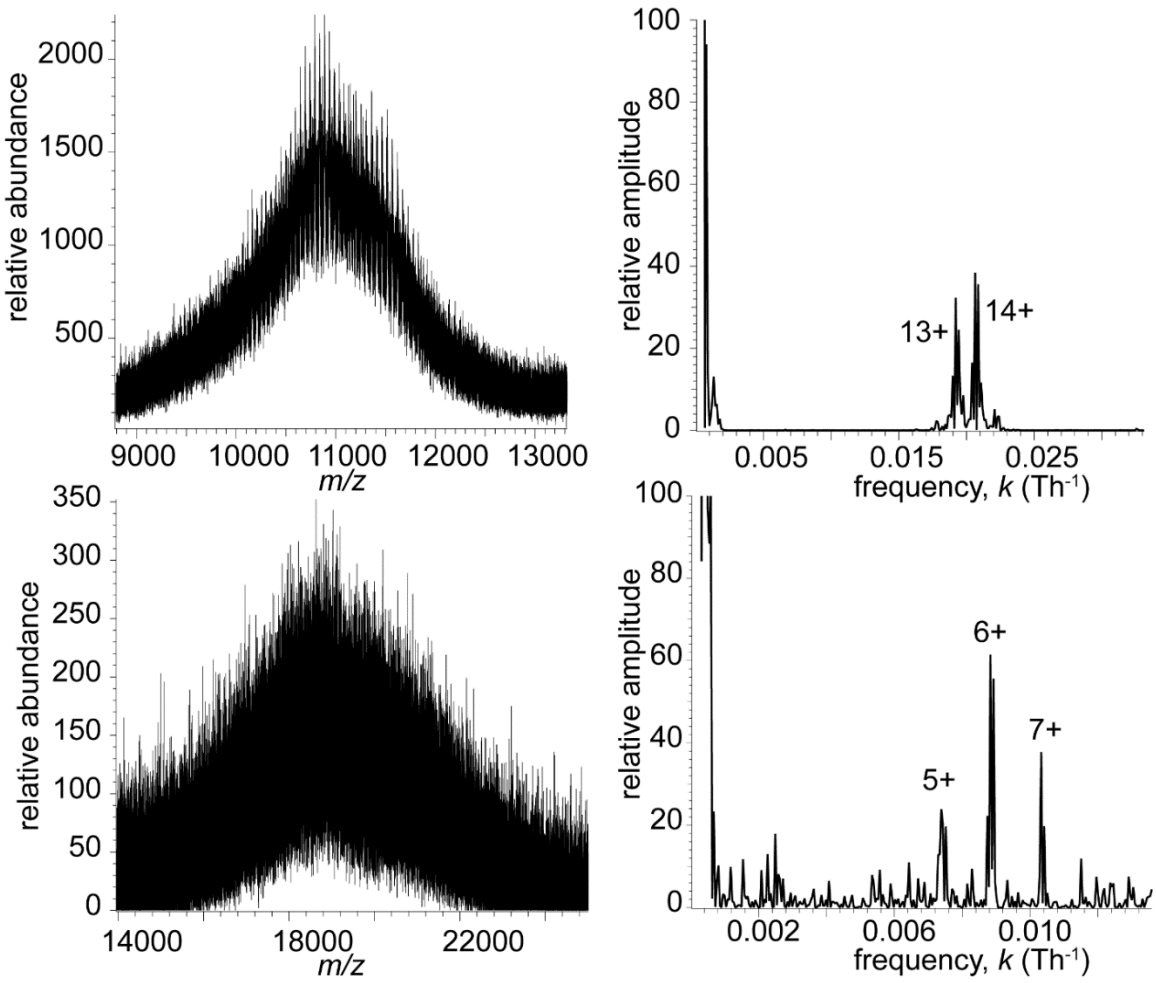


Figure A4. Mass spectra of DMPC Nanodisc ions (left) and corresponding FFT spectra (right) acquired under "mild" (top) and "strong" (bottom) collisional activation conditions. The high-intensity peak near zero frequency in the FFT spectra has been truncated for clarity.

APPENDIX B

SUPPLEMENTAL INFORMATION FOR CHAPTER III

Characterizing peak width and unresolved adductions in mass spectrum for non-Gaussian peak shapes. In the case that the mass spectral peak shape is unknown and not Gaussian, for example, when there is a long, asymmetric "tail" on each peak in the mass spectrum due to unresolved adductions of small ions or solvent molecules, the amplitude of the zero-frequency peak in the Fourier spectrum is not uniquely determine by the value of P(k) at the harmonic peaks. This can result in gross errors in estimating mass spectral peak width when the zero-frequency amplitude for a particular charge state is not known. This scenario is demonstrated for a single charge state in Supplementary Figures S7C and S7D, where a sawtooth wave function is used as a model of a highly asymmetric peak shape. Because the zero-frequency component is in general needed to uniquely reconstruct asymmetric peaks, the Gaussian fit to P(k) is unsurprisingly poor. This example illustrates the potential danger of determining peak shape characteristics when the mass spectral peaks have highly asymmetric or unknown shapes.

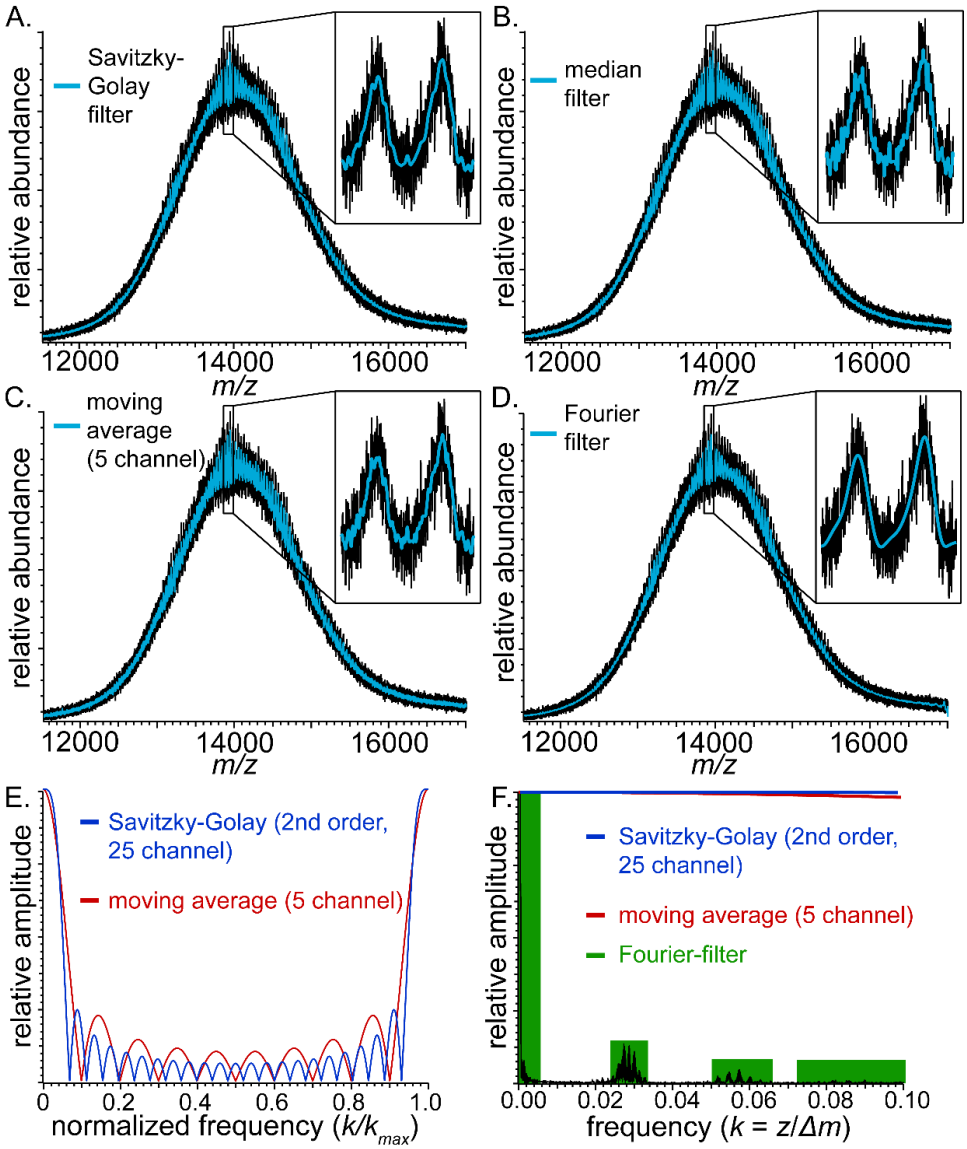


Figure B1. Comparison of DMPC-MSP1D1 Nanodisc mass spectrum (black traces) acquired using QTOF mass spectrometer and processed (blue traces) using different types of filtering/smoothing algorithms (A-D). Panel E shows the FT representation of Savitzky-Golay and moving-average filter functions over the entire Fourier domain. Panel F shows the FT representation of Savitzky-Golay and moving-average filters (colored lines) and Fourier-filtered data (highlighted in green) for the Nanodisc mass spectrum. The Savitzky-Golay and moving-average filters have the effect of damping very high-frequency noise and slightly altering relative Fourier-domain peak amplitudes. By contrast, the Fourier filter removes all low-frequency noise between each set of harmonics and all high-frequency noise beyond the highest-harmonic data used without altering the relative amplitudes of the preserved data.

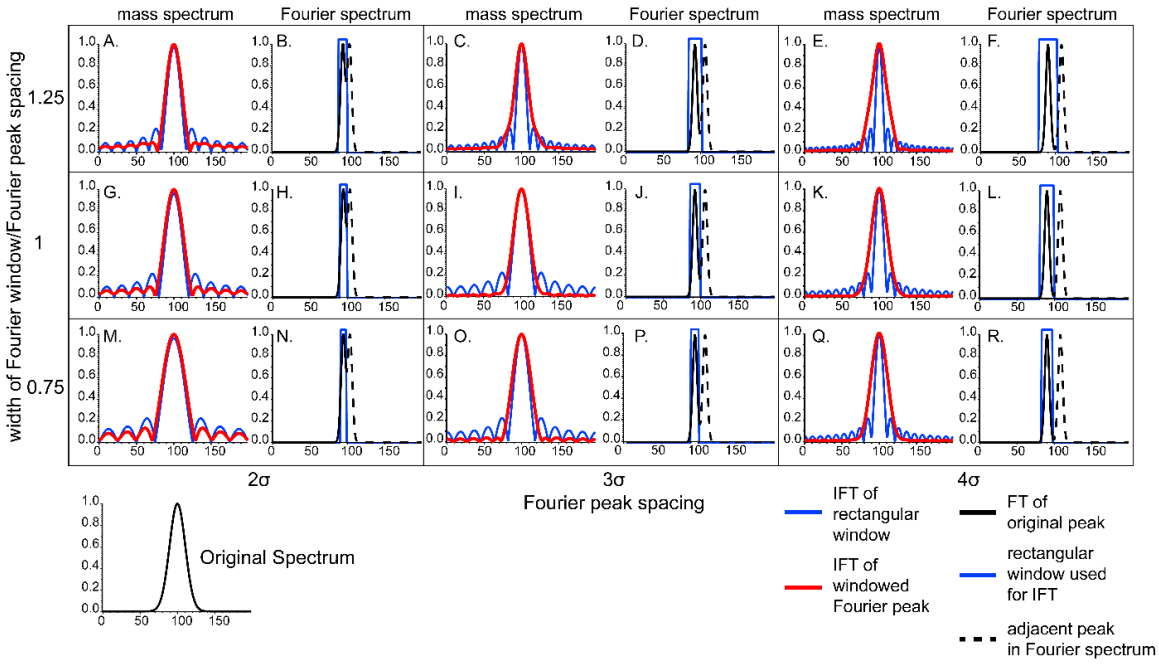


Figure B2. Model mass spectral envelopes (A, C, E, G, I, K, M, O, Q; black traces) and corresponding Fourier spectra (B, D, F, H, J, L, N, P, R; left peak of each peak pair), illustrating effects of overlap of a Fourier-domain peak with an adjacent peak and choice of window width used in IFT. Peaks in Fourier domain have fixed standard deviation (σ), but spacing between Fourier-domain peaks varies from left to right as indicated. Width of Fourier-domain window (blue boxes) used for IFT to reconstruct mass spectral envelopes (red traces) varies as indicated along vertical axis, and blue trace in mass spectra indicate the IFT of the rectangular window function itself. Faithfulness of reconstructed mass spectral envelope generally increases from bottom left to top right, as the spacing between Fourier-domain peaks and the width of the window used for IFT increases. Windowing and overlap artifacts are less than 5% of the mass spectral peak abundance when the spacing between Fourier-domain peaks is at least 3 times their width (measured as standard deviation, σ).

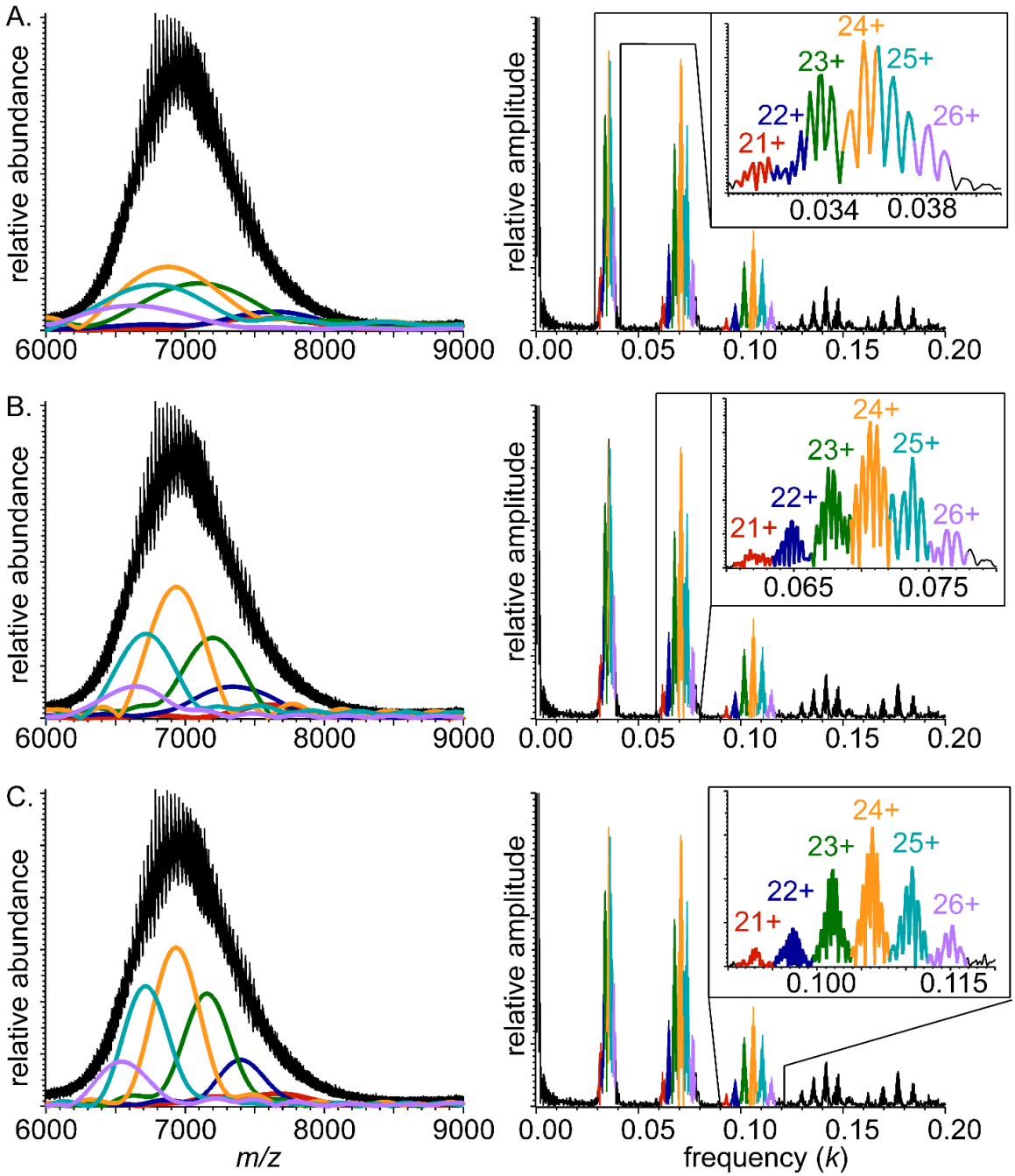


Figure B3. Mass spectra (left) and corresponding Fourier spectra (right) for DMPC-MSP1D1 Nanodiscs acquired using QTOF mass spectrometer, showing reconstructed charge-state-specific mass spectral envelopes found using Fourier-domain (A) fundamentals, (B) second harmonics, and (C) third harmonics. Insets show detail for Fourier-domain peaks, with reconstructed mass spectral envelopes corresponding to charge states labeled with the same color

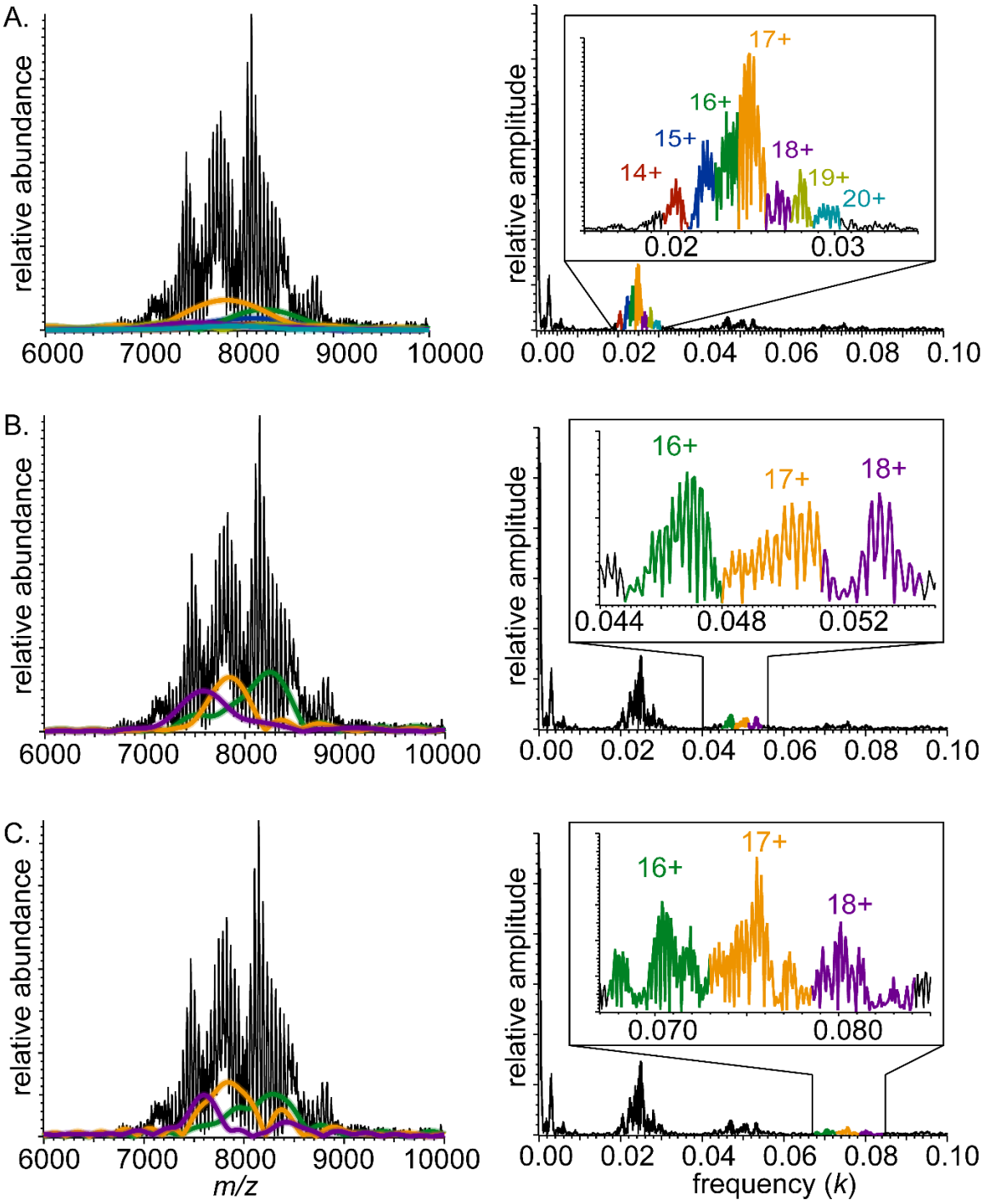


Figure B4. Mass spectra (left) and corresponding Fourier spectra (right) for DMPC-MSP1D1 Nanodiscs acquired using FT-ICR mass spectrometer, showing reconstructed charge-state-specific mass spectral envelopes found using Fourier-domain (A) fundamentals, (B) second harmonics, and (C) third harmonics. Insets show detail for Fourier-domain peaks, with reconstructed mass spectral envelopes corresponding to charge states labeled with the same color

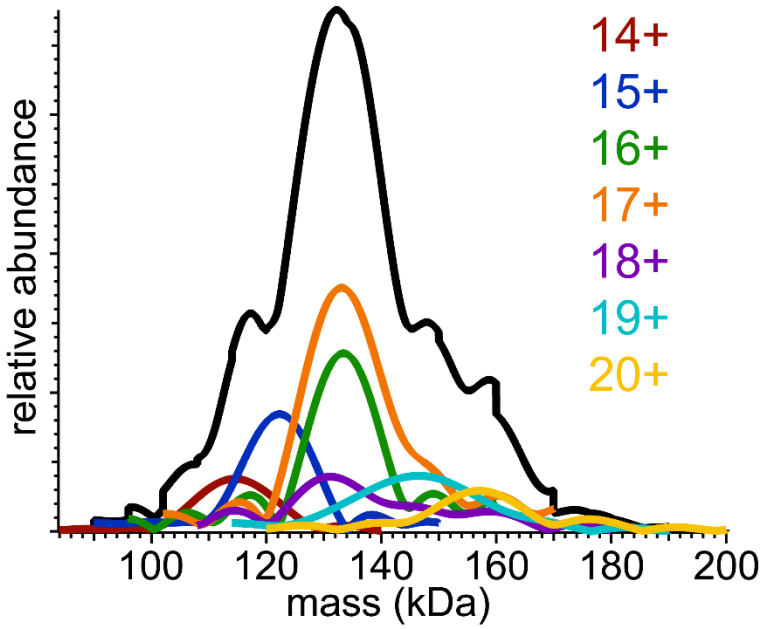


Figure B5. Reconstructed zero-charge mass spectrum of DMPC-MSP1D1 Nanodiscs acquired on an FT-ICR mass spectrometer using data from Fourier-domain fundamental peaks for individual charge states (colored traces) and for entire ion population (black trace). Higher harmonics were not used for full ion population reconstruction because charge states other than 16-18+ had poor 2nd and 3rd harmonic signal-to-noise in the Fourier spectrum. Very broad mass distributions are attributed in part to artifacts from overlapped fundamental peaks (see Supplementary Figure S4; Supplementary Figure S10C and S10F contain a zero-charge mass spectrum reconstruction using the fundamentals, 2nd, and 3rd harmonic peaks for the 16-18+ charge states)

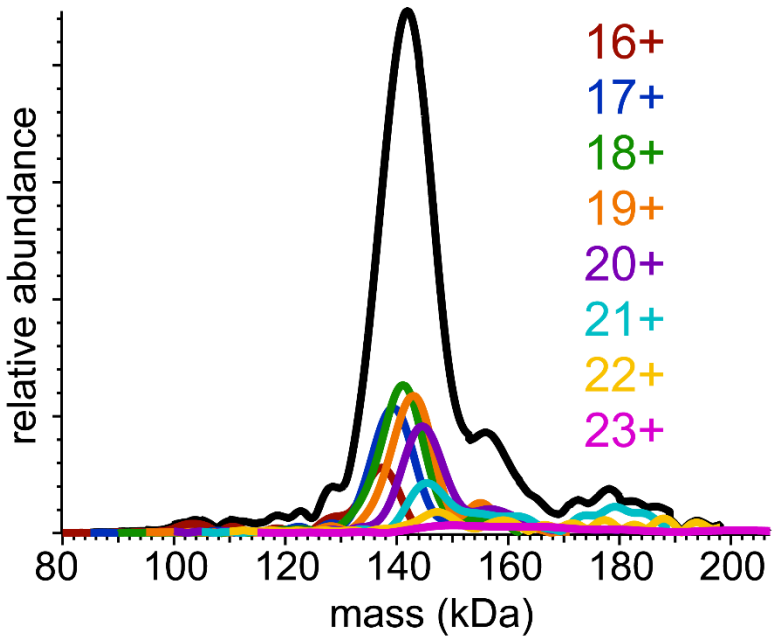


Figure B6. Reconstructed zero-charge mass spectrum of DMPC-MSP1D1 Nanodiscs acquired on an Orbitrap mass spectrometer using data from Fourier-domain 2nd harmonic peaks for individual charge states (colored traces) and for entire ion population (black trace). Narrower mass distribution than that shown in Supplementary Figure S5 is attributed to greater resolution of 2nd harmonic peaks for these data

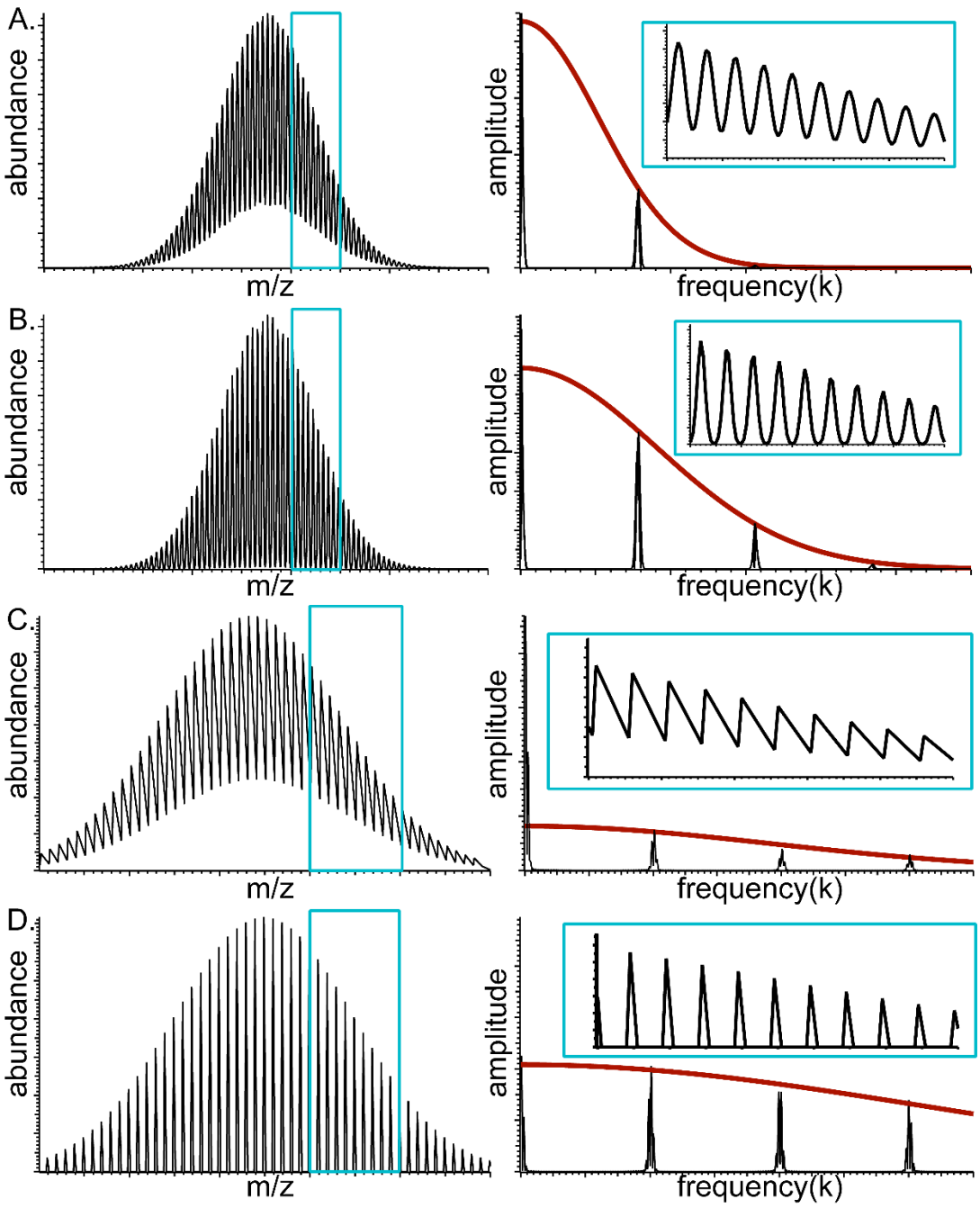


Figure B7. Model mass spectrum (left) for a single charge state of a population of assembly ions that varies in the number of subunits, and the corresponding Fourier spectrum (right) for different peak shapes: (A) overlapping Gaussian peaks with a large curved baseline, (B) baseline-resolved Gaussian peaks, (C) overlapping sawtooth-shaped peaks with a large curved baseline, and (D) baseline-resolved sawtooth-shaped peaks. Insets shows close-up view of mass spectra. Gaussian fits to Fourier-domain decay functions, P(k), shown in red, are better for Gaussian-shaped mass spectral peaks than for poorly resolved sawtooth-shaped mass spectral peaks

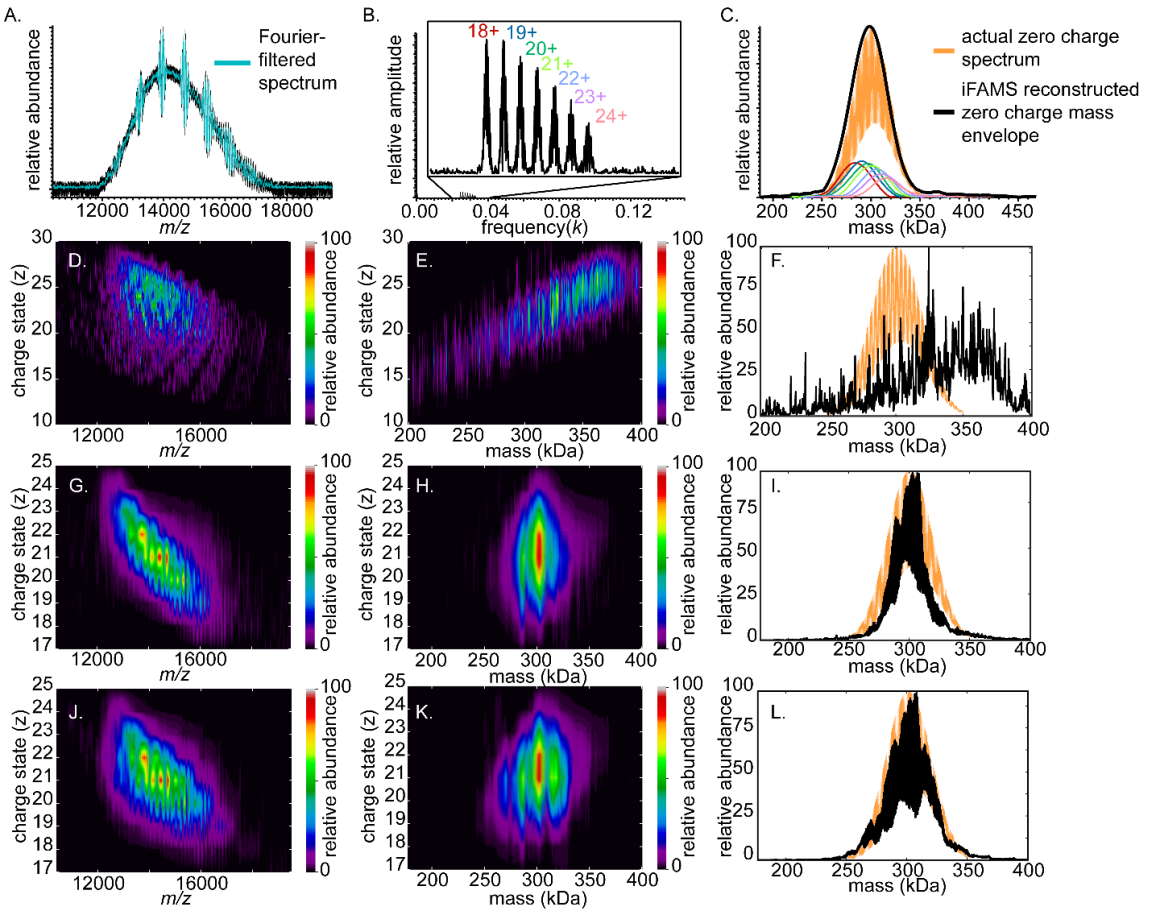


Figure B8. Simulated mass spectrum with 20:1 signal-to-white-noise (A, black trace) for DPPC-MSP1E3D1 Nanodiscs with charge states 18-24+ and ~300-340 lipids, with the average number of lipids increasing by 10 with each increasing charge state, similar to experimental data. The blue trace in panel A shows the Fourier-filtered spectrum reconstructed using the Fourier-Transformed data from panel B (inset: fundamental peaks). Panel C shows the reconstructed zero-charge mass spectral envelopes for each charge state (colors of traces correspond to the charge states shown in panel B) the total zero-charge mass spectral envelope (black trace), and the exact zero-charge mass spectrum (orange trace). The reconstructions are remarkably accurate despite the low signal-to-noise of the input mass spectrum. Panels D, E, and F show deconvolution results from UniDec with a wide input charge state range (10-30+). Panels G, H, and I show deconvolution results from UniDec with a narrow input charge state range (17-25+) based on charge states identified using the FT method. Panels J, K, and L show deconvolution results from UniDec using Fourier-filtered data from FT analysis and input charge state range 17-25+. Dramatic improvement of both the charge-state-specific and zero-charge mass spectra is observed when using output from FT analysis in UniDec.

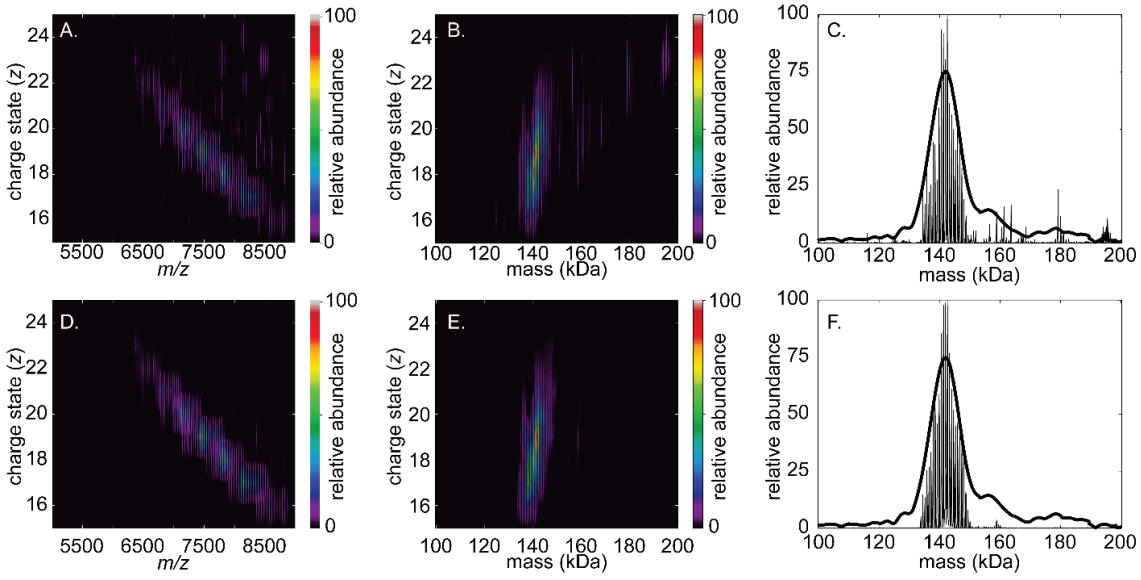


Figure B9. Charge vs. *m*/*z* (left), charge vs. mass (middle) and total zero-charge mass spectrum (right) for mass spectrum of DMPC-MSP1D1 Nanodiscs acquired using Orbitrap mass spectrometer and deconvolved using UniDec. A, B, and C result from using "naive" input parameters for subunit mass, charge state range, peak width, and total mass, whereas D, E, and F result from using input parameter values obtained using the FT-based method described in the text. Smooth black trace in C and F represents zero-charge mass spectrum reconstructed from Fourier spectrum 2nd harmonics using FT analysis.

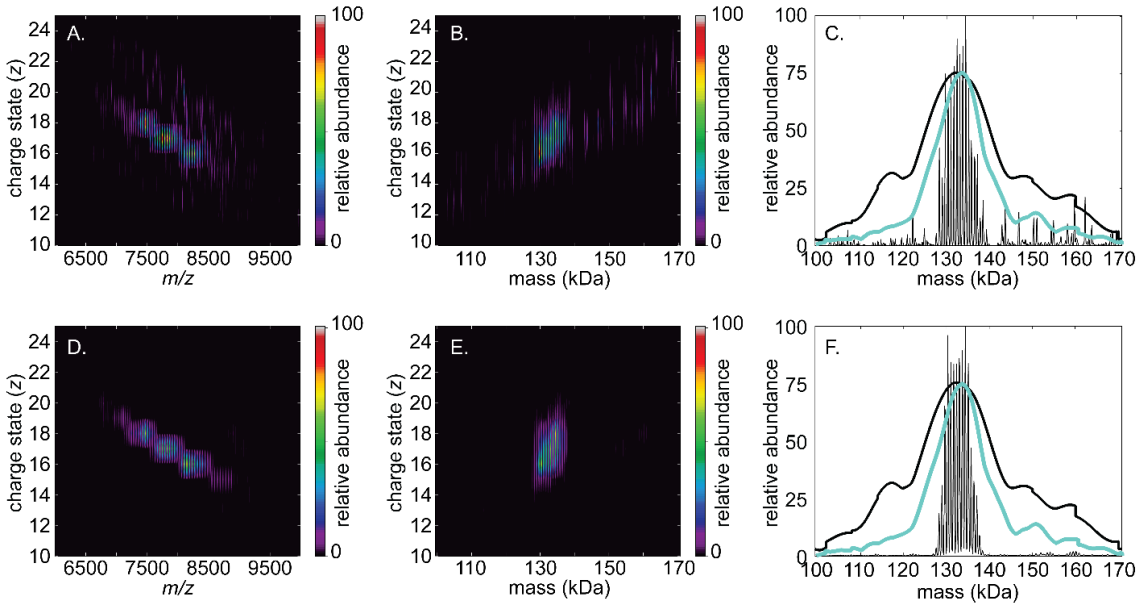


Figure B10. Charge vs. *m*/*z* (left), charge vs. mass (middle) and total zero-charge mass spectrum (right) for mass spectrum of DMPC-MSP1D1 Nanodiscs acquired using FT-ICR mass spectrometer and deconvolved using UniDec. A, B, and C result from using "naive" input parameters for subunit mass, charge state range, peak width, and total mass, whereas D, E, and F result from using input parameter values obtained using the FT-based method described in the text. Smooth black (resp., blue) trace in C and F represents zero-charge mass spectrum reconstructed from Fourier spectrum fundamentals (resp., fundamentals, 2nd, and 3rd harmonics for 16-18+ charge states) using FT analysis. (Charge states other than 16-18+ had poor 2nd and 3rd harmonic signal-to-noise in the Fourier spectrum.)

Table B1. Lipid Mass, Charge States, and Lipid Stoichiometry Statistics Determined for Native-Like Nanodisc Ions Using FT-Based Approach.

Analyte (Instrument)	**Harmonic**	**Subunit mass (Da.)**	***z***	**Lipid Stoichiometry**
DPPC-MSP1E3D1 (QTOF)	Fun	733. ± 2.	18	273 ± 36†
			19	290 ± 32†
			20	304 ± 28†
			21	316 ± 29†
			22	325 ± 30†
			23	340 ± 41†
			24	353 ± 47
	2nd	733.0 ± 0.8	18	275 ± 41*
			19	294 ± 29
			20	305 ± 26
			21	315 ± 26
			22	327 ± 33
			23	352 ± 46*
			24	383 ± 56*
	Average	N/A	18	274 ± 39
			19	292 ± 31
			20	304 ± 27
			21	316 ± 28
			22	326 ± 32
			23	346 ± 45
			24	372 ± 55
DMPC-MSP1D1 (Orbitrap)	Fun	678.5 ± 3.6	16	125 ± 15†
			17	129 ± 24†
			18	137 ± 19†
			19	143 ± 22†
			20	149 ± 20†
			21	158 ± 22†
			22	166 ± 23
			23	179 ± 23
	2nd	678.2 ± 1.0	16	128 ± 17
			17	136 ± 14
			18	143 ± 12

		19	146 ± 14
		20	150 ± 16
		21	154 ± 20
		22	162 ± 25
		23	178 ± 28
3rd	677.3 ± 1.2	16	127 ± 17
		17	137 ± 12
		18	141 ± 12
		19	145 ± 12
		20	150 ± 15

* indicates Fourier-domain peaks with signal-to-noise less than 10:1
† indicates Fourier-domain peaks spaced by less than 1.5 times the sum of their apparent standard deviations

APPENDIX C

SUPPLEMENTAL INFORMATION FOR CHAPTER IV

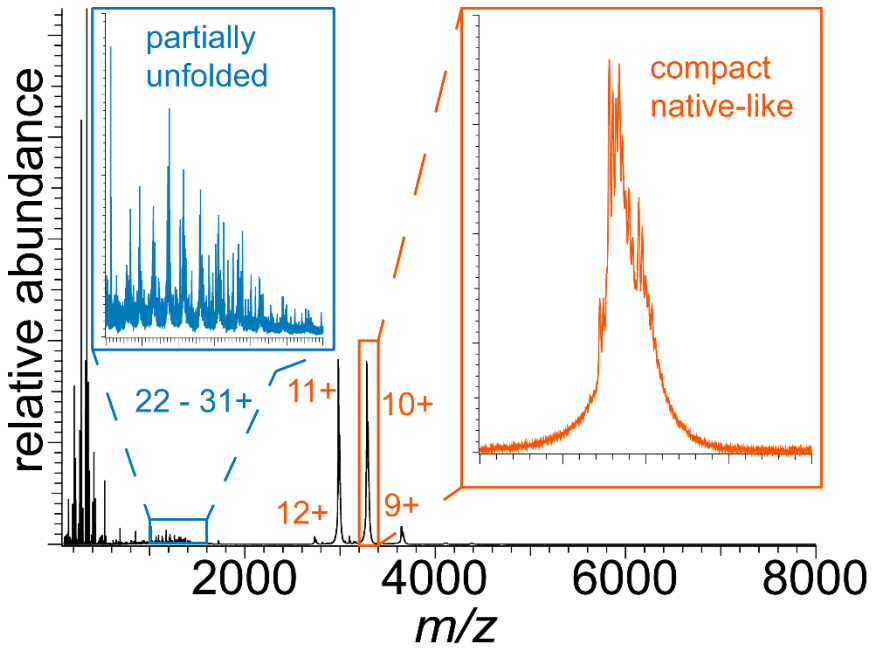


Figure C1. Mass spectrum of LFn electrosprayed from 200 mM ammonium acetate buffer. Insets show two different regions in the charge-state distribution for this protein, including unfolded states (blue, 22-31+) and compact native charge states (orange, 9-12+).

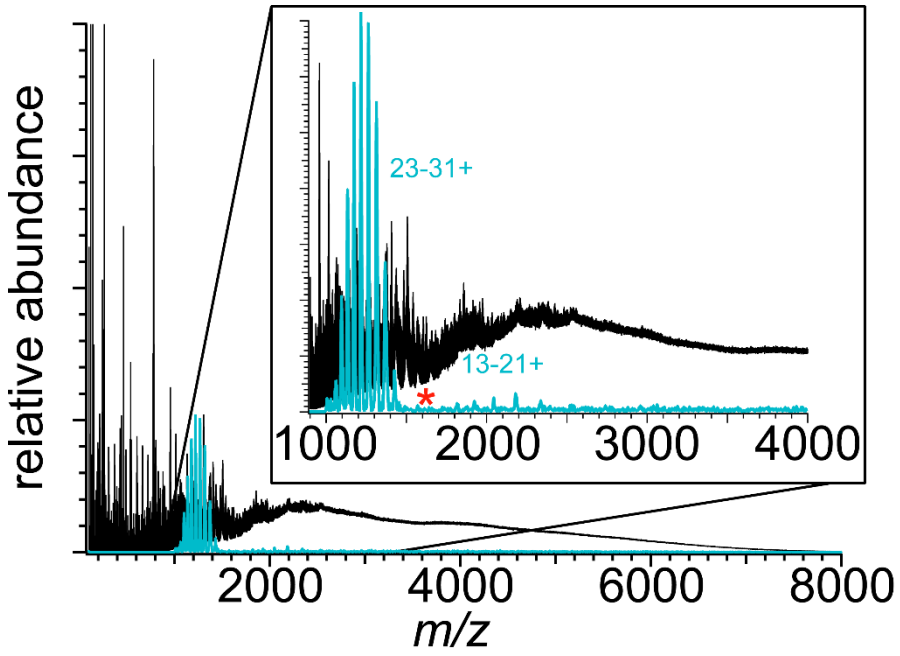


Figure C2. Inverse Gábor transform spectrum of LFn electrosprayed from buffer containing 20 mM Tris, 100 mM NaCl, and 0.5 mM EDTA. The inverse Gábor transform spectra (blue) from Figure 21 (Chapter 4) is overlaid with the original mass spectrum (black) from Figure 19 (Chapter 4). The red star indicates where one would expect the 22+ charge state to be. Because signal from this charge state is overlapped with the k = 1.0 frequency band in the Gábor spectrum (see Figure 21, Chapter 4) signal for this charge state of LFn is not resolved.

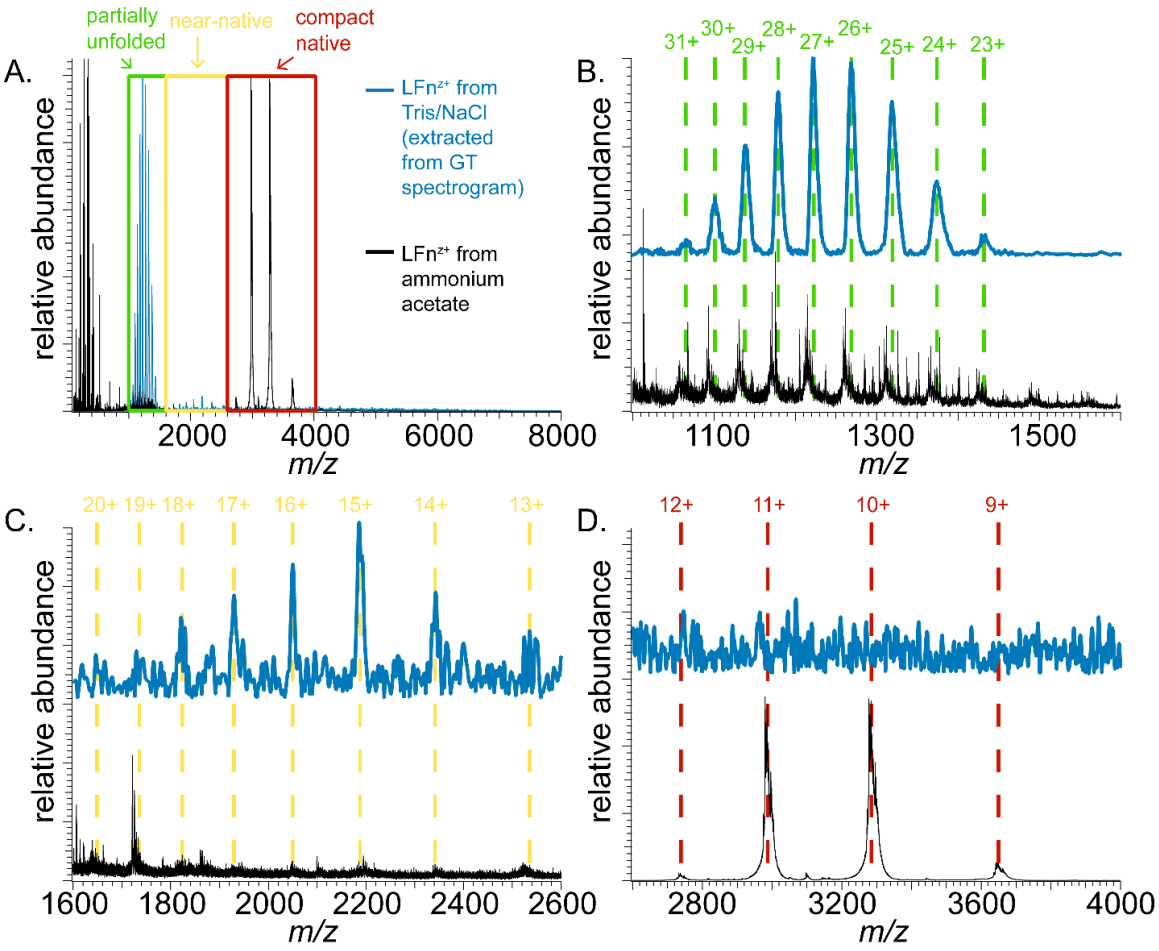


Figure C3. Comparison of the mass spectra of LFn ions electrosprayed from ammonium acetate and NaCl/Tris. (A) The ammonium acetate spectrum in Figure S1 (black) is overlaid with the inverse Gábor transformed (IGT) data from the NaCl/Tris spectrum in Figure S2 (blue). Different populations of charge states, including those corresponding to unfolded states (green), near-native states (yellow), and compact native states (red) are identified by color. (B-D) A comparison of the mass spectra from Tris/NaCl (found from the IGT data) and ammonium acetate solutions is shown for the different regions. Dotted lines in B and C imply line up with the centroid of the charge state peak in the inverse Gábor transform spectrum. Dotted lines in D line up with the centroid of the charge state peak for the ammonium acetate spectrum.

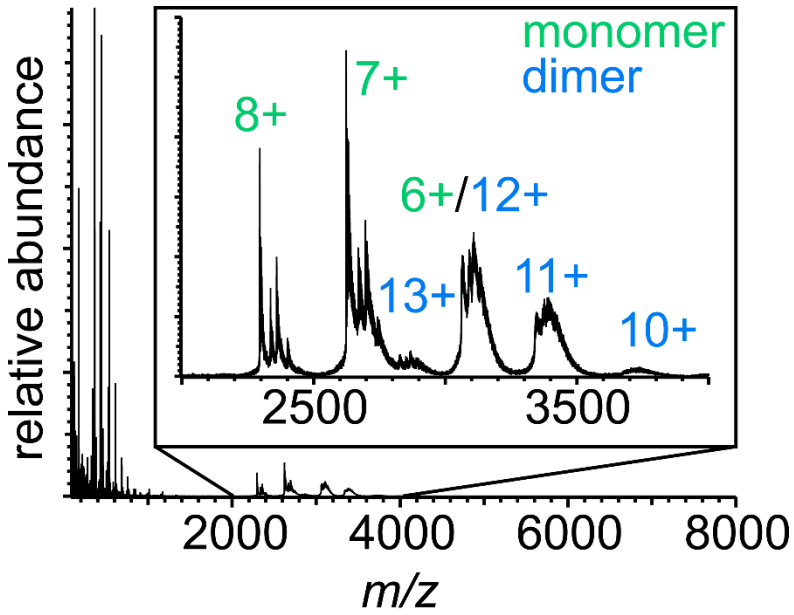


Figure C4. Mass spectrum of β-lactoglobulin electrosprayed from 200 mM ammonium acetate buffer. The color of the charge states indicates whether a peak is a monomer (green) or a dimer (blue). Multiple peaks at each charge state are due to different isoforms of the protein.

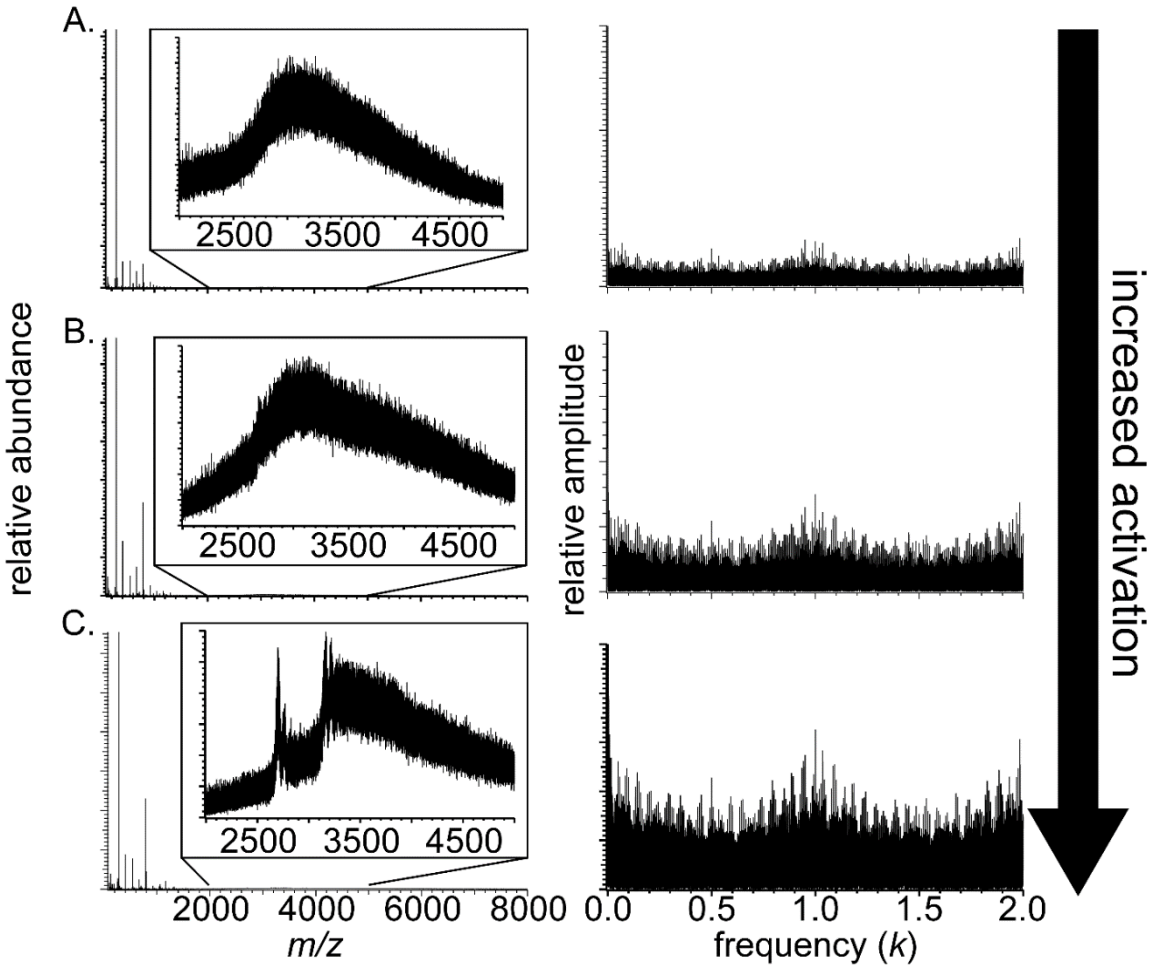


Figure C5. Mass spectrum of β-lactoglobulin electrosprayed from a 20 mM HEPES, 100 mM NaCl, and 0.5 mM EDTA. Trap collisional activation increases going downward, and is set to 50 (A), 70 (B), and 100 (C).

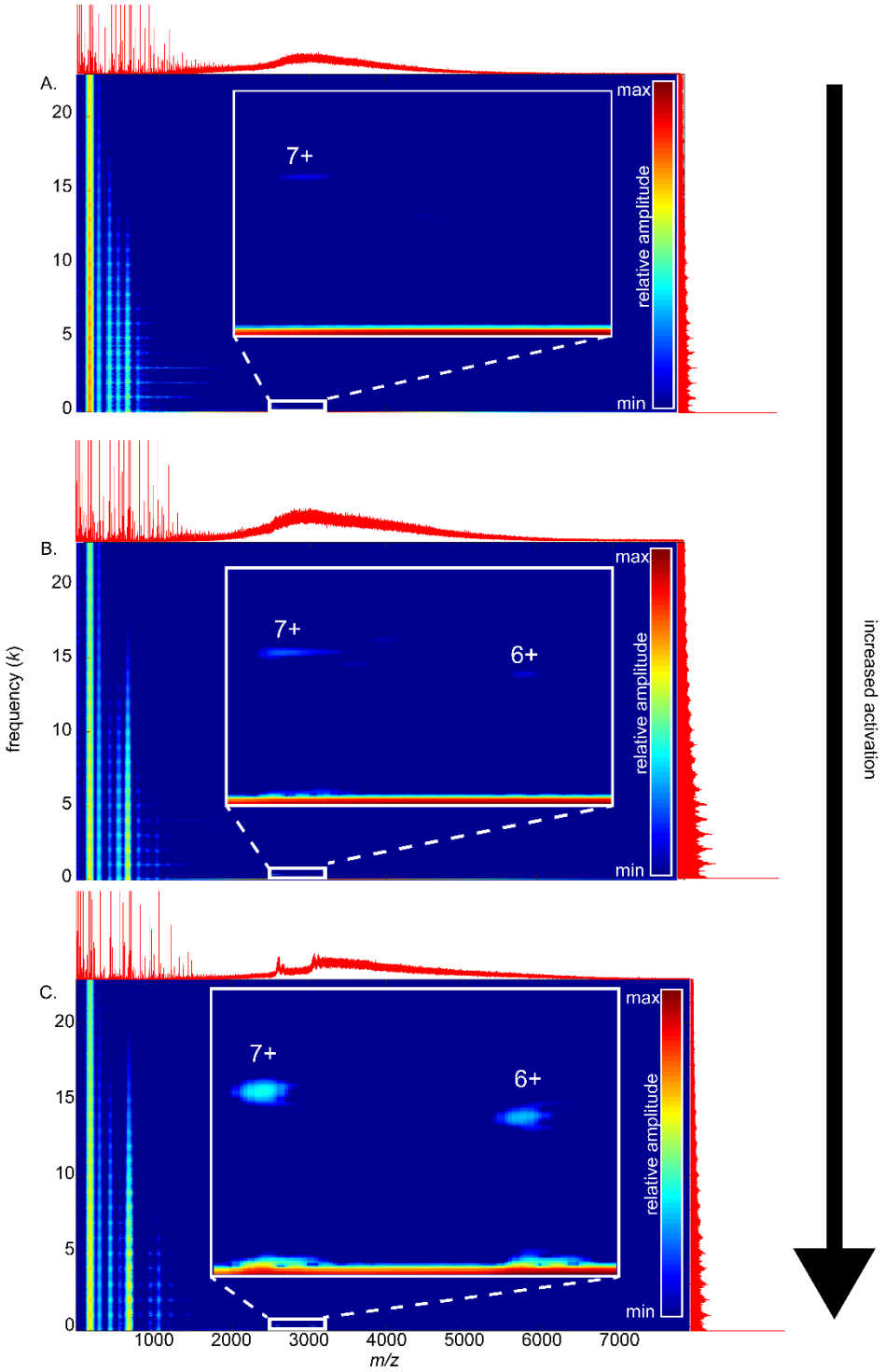


Figure C6. Gábor transformation of the β-lactoglobulin mass spectra in Figure S5, electrosprayed from a 20 mM HEPES, 100 mM NaCl, and 0.5 mM EDTA. The letter identifier corresponds to the same letter identifier for the spectra in Figure S5. Signals from the 6+ and 7+ native charge states are readily identifiable using GT analysis, even though their signals are strongly dominated by salt cluster signal in the mass spectra.

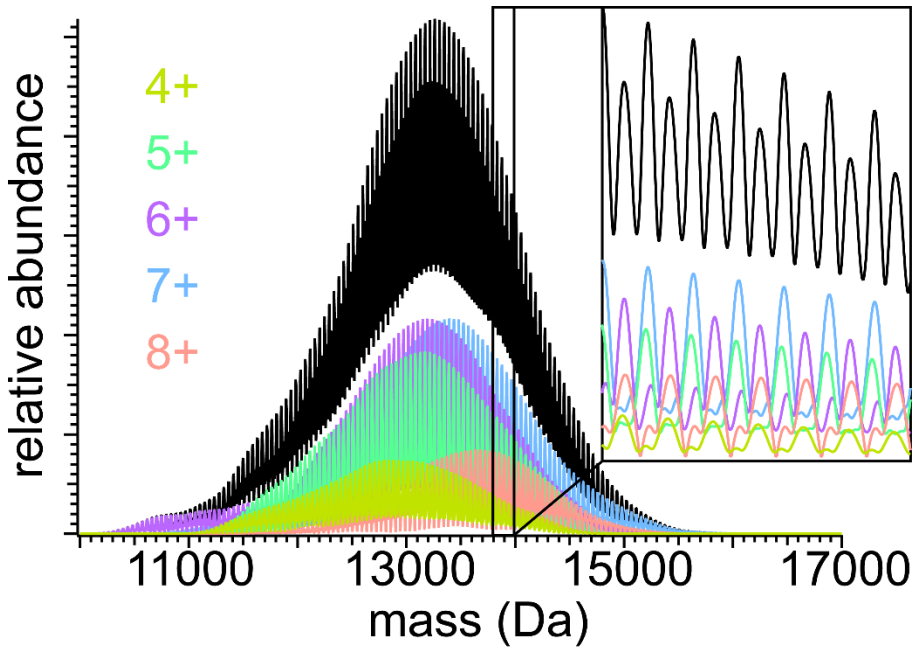


Figure C7. Zero-charge spectrum for the PEG 10k polymer. Zero-charge spectrum reconstructions are prepared from the isolated charge-state-specific mass spectra found by inverse Gabor transformation of data in the GT spectrogram. Individual charge state components are labeled by color, while the total mass spectrum is shown in black.

Table C1. Subunit statistics for the long-chain Polyethylene Glycol 10k Ion Population

charge state (z)	**polymer mass mean and standard deviation**	**equivalent number of ethylene glycol monomers**
"zero charge"	13240 ± 760	300 ± 17
4+	12920 ± 780	293 ± 18
5+	13100 ± 770	297 ± 17
6+	13160 ± 740	299 ± 17
7+	13410 ± 730	304 ± 17
8+	13610 ± 690	309 ± 16

REFERENCES CITED

1. Marty, M. T., et al. "Probing the Lipid Annular Belt by Gas-Phase Dissociation of Membrane Proteins in Nanodiscs" *Angew. Chem.-Int. Edit.* **2016,** *55*, 550-554.

2. Robinson, C. V., et al. "Probing the Nature of Noncovalent Interactions by Mass Spectrometry. A Study of Protein−CoA Ligand Binding and Assembly" *J. Am. Chem. Soc.* **1996,** *118*, 8646-8653.

3. Weidner, S. M. and Trimpin, S. "Mass Spectrometry of Synthetic Polymers" *Anal. Chem.* **2008,** *80*, 4349-4361.

4. Prebyl, B. S. and 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*, 127-136.

5. Trimpin, S. and Clemmer, D. E. "Ion Mobility Spectrometry/Mass Spectrometry Snapshots for Assessing the Molecular Compositions of Complex Polymeric Systems" *Anal. Chem.* **2008,** *80*, 9073-9083.

6. Trimpin, S., et al. "Resolving Oligomers from Fully Grown Polymers with IMS−MS" *Anal. Chem.* **2007,** *79*, 7965-7974.

7. Campuzano, I. D. G., et al. "Native-MS Analysis of Monoclonal Antibody Conjugates by Fourier Transform Ion Cyclotron Resonance Mass Spectrometry" *Anal. Chem.* **2018,** *90*, 745-751.

8. Tian, Y., et al. "Quantitative collision-induced unfolding differentiates model antibody–drug conjugates" *Protein Sci.* **2019,** *28*, 598-608.

9. Doncom, K. E. B., et al. "Dispersity effects in polymer self-assemblies: a matter of hierarchical control" *Chem. Soc. Rev.* **2017,** *46*, 4119-4134.

10. Bechara, C., et al. "A Subset of Annular Lipids is Linked to the Flippase Activity of an ABC Transporter" *Nat. Chem.* **2015,** *7*, 255-262.

11. Allison, T. M., et al. "Quantifying the stabilizing effects of protein–ligand interactions in the gas phase" *Nat. Commun.* **2015,** *6*, 8551.

12. Laganowsky, A., et al. "Membrane proteins bind lipids selectively to modulate their structure and function" *Nature* **2014,** *510*, 172-175.

13. Cleary, S. P., Thompson, A. M. and Prell, J. S. "Fourier Analysis Method for Analyzing Highly Congested Mass Spectra of Ion Populations with Repeated Subunits" *Anal. Chem.* **2016,** *88*, 6205-6213.

14. Marty, M. T., et al. "Bayesian Deconvolution of Mass and Ion Mobility Spectra: From Binary Interactions to Polydisperse Ensembles" *Anal. Chem.* **2015,** *87*, 4370-4376.

15. Gaborieau, M. and Castignolles, P. "Size-exclusion chromatography (SEC) of branched polymers and polysaccharides" *Anal. Bioanal. Chem.* **2011,** *399*, 1413-1423.

16. Izunobi, J. U. and Higginbotham, C. L. "Polymer Molecular Weight Analysis by 1H NMR Spectroscopy" *J. Chem. Educ.* **2011,** *88*, 1098-1104.

17. Larriba, C., de la Mora, J. F. and Clemmer, D. E. "Electrospray Ionization Mechanisms for Large Polyethylene Glycol Chains Studied Through Tandem Ion Mobility Spectrometry" *J. Am. Soc. Mass Spectrom.* **2014,** *25*, 1332-1345.

18. Akkaladevi, N., et al. "Following Natures Lead: On the Construction of Membrane-Inserted Toxins in Lipid Bilayer Nanodiscs" *J. Membr. Biol.* **2015,** *248*, 595-607.

19. Bush, M. F., et al. "Collision Cross Sections of Proteins and Their Complexes: A Calibration Framework and Database for Gas-Phase Structural Biology" *Anal. Chem.* **2010,** *82*, 9557-9565.

20. Freeke, J., Robinson, C. V. and Ruotolo, B. T. "Residual counter ions can stabilise a large protein complex in the gas phase" *Int. J. Mass Spectrom.* **2010,** *298*, 91-98.

21. Barrera, N. P., et al. "Mass Spectrometry of Membrane Transporters Reveals Subunit Stoichiometry and Interactions" *Nat. Methods* **2009,** *6*, 585-587.

22. Kintzer, A. F., et al. "The Protective Antigen Component of Anthrax Toxin Forms Functional Octameric Complexes" *J. Mol. Biol.* **2009,** *392*, 614-629.

23. Susa, A. C., et al. "Charging of Proteins in Native Mass Spectrometry" *J. Am. Soc. Mass Spectrom.* **2017,** *28*, 332-340.

24. Marty, M. T., et al. "Interpretation and Deconvolution of Nanodisc Native Mass Spectra" *J. Am. Soc. Mass Spectrom.* **2014,** *25*, 269-277.

25. Cleary, S. P., et al. "Extracting Charge and Mass Information from Highly Congested Mass Spectra Using Fourier-Domain Harmonics" *J. Am. Soc. Mass Spectrom.* **2018,** *29*, 2067-2080.

26. Mann, M., Meng, C. K. and Fenn, J. B. "Interpreting Mass-Spectra of Multiply Charged Ions" *Anal. Chem.* **1989,** *61*, 1702-1708.

27. Sharon, M. and Robinson, C. V. "The Role of Mass Spectrometry in Structure Elucidation of Dynamic Protein Complexes" *Annu. Rev. Biochem.* **2007,** *76*, 167-193.

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

29. Zhang, H., et al. "Native Mass Spectrometry of Photosynthetic Pigment-Protein Complexes" *FEBS Lett.* **2013,** *587*, 1012-1020.

30. Fenn, J., et al. "Electrospray ionization for mass spectrometry of large biomolecules" *Science* **1989,** *246*, 64-71.

31. Marcoux, J. and Robinson, Carol V. "Twenty Years of Gas Phase Structural Biology" *Structure* **2013,** *21*, 1541-1550.

32. Jurneczko, E. and Barran, P. E. "How useful is ion mobility mass spectrometry for structural biology? The relationship between protein crystal structures and their collision cross sections in the gas phase" *Analyst* **2011,** *136*, 20-28.

33. Bleiholder, C., et al. "Ion mobility spectrometry reveals the mechanism of amyloid formation of Abeta(25-35) and its modulation by inhibitors at the molecular level: epigallocatechin gallate and scyllo-inositol" *J Am Chem Soc* **2013,** *135*, 16926-16937.

34. Kintzer, A. F., et al. "The Protective Antigen Component of Anthrax Toxin Forms Functional Octameric Complexes" *J. Mol. Biol.* **2009,** *392*, 614-629.

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

36. Lössl, P., Snijder, J. and Heck, A. J. R. "Boundaries of Mass Resolution in Native Mass Spectrometry" *J. Am. Soc. Mass Spectrom.* **2014,** *25*, 906-917.

37. Bechara, C. and Robinson, C. V. "Different Modes of Lipid Binding to Membrane Proteins Probed by Mass Spectrometry" *J. Am. Chem. Soc.* **2015,** *137*, 5240-5247.

38. Struwe, W. B., et al. "Collision cross sections of high-mannose N-glycans in commonly observed adduct states--identification of gas-phase conformers unique to [M-H](-) ions" *Analyst* **2015,** *140*, 6799-6803.

39. Gotzke, H., et al. "Identification of putative substrates for the periplasmic chaperone YfgM in Escherichia coli using quantitative proteomics" *Mol Cell Proteomics* **2015,** *14*, 216-226.

40. Han, L. and Ruotolo, B. T. "Ion Mobility-Mass Spectrometry Differentiates Protein Quaternary Structures Formed in Solution and in Electrospray Droplets" *Anal Chem* **2015,** *87*, 6808-6813.

41. Quintyn, R. S., et al. "Surface-Induced Dissociation Mass Spectra as a Tool for Distinguishing Different Structural Forms of Gas-Phase Multimeric Protein Complexes" *Anal. Chem.* **2015,** *87*, 11879-11886.

42. Bagal, D., Zhang, H. and 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*, 2408-2418.

43. Lei, Q. P., et al. "Electrospray mass spectrometry studies of non-heme iron-containing proteins" *Anal Chem* **1998,** *70*, 1838-1846.

44. Landreh, M., et al. "Controlling release, unfolding and dissociation of membrane protein complexes in the gas phase through collisional cooling" *Chem Commun (Camb)* **2015,** *51*, 15582-15584.

45. Ruotolo, B. T. and Robinson, C. V. "Aspects of native proteins are retained in vacuum" *Curr Opin Chem Biol* **2006,** *10*, 402-408.

46. Zhang, Z. and Marshall, A. G. "A universal algorithm for fast and automated charge state deconvolution of electrospray mass-to-charge ratio spectra" *J Am Soc Mass Spectrom* **1998,** *9*, 225-233.

47. Morgner, N. and Robinson, C. V. "Massign: An Assignment Strategy for Maximizing Information from the Mass Spectra of Heterogeneous Protein Assemblies" *Anal. Chem.* **2012,** *84*, 2939-2948.

48. Marty, M. T., et al. "Nanodisc-solubilized membrane protein library reflects the membrane proteome" *Anal. Bioanal. Chem.* **2013,** *405*, 4009-4016.

49. Schuler, M. A., Denisov, I. G. and Sligar, S. G. "Nanodiscs as a new tool to examine lipid-protein interactions" *Methods Mol Biol* **2013,** *974*, 415-433.

50. Robb, D. B., et al. "Method of atmospheric pressure charge stripping for electrospray ionization mass spectrometry and its application for the analysis of large poly(ethylene glycol)s" *Anal Chem* **2014,** *86*, 9644-9652.

51. Marty, M. T., et al. "Native Mass Spectrometry Characterization of Intact Nanodisc Lipoprotein Complexes" *Anal. Chem.* **2012,** *84*, 8957-8960.

52. Bayburt, T. H. and Sligar, S. G. "Membrane Protein Assembly into Nanodiscs" *FEBS Lett.* **2010,** *584*, 1721-1727.

53. Fellgett, P. B. "On the ultimate sensitivity and practical performance of radiation detectors" *J Opt Soc Am* **1949,** *39*, 970-976.

54. Chiu, S. W., et al. "Structure of sphingomyelin bilayers: a simulation study" *Biophys J* **2003,** *85*, 3624-3635.

55. Denisov, I. G., et al. "Directed Self-Assembly of Monodisperse Phospholipid Bilayer Nanodiscs with Controlled Size" *J. Am. Chem. Soc.* **2004,** *126*, 3477-3487.

56. Laszlo, K. J. and Bush, M. F. "Analysis of Native-Like Proteins and Protein Complexes Using Cation to Anion Proton Transfer Reactions (CAPTR)" *J. Am. Soc. Mass Spectrom.* **2015,** *26*, 2152-2161.

57. Bagal, D., et al. "Gas Phase Stabilization of Noncovalent Protein Complexes Formed by Electrospray Ionization" *Analytical Chemistry* **2009,** *81*, 7801-7806.

58. Catalina, M. I., et al. "Decharging of Globular Proteins and Protein Complexes in Electrospray" *Chem. - Eur. J.* **2005,** *11*, 960-968.

59. Lermyte, F., et al. "Extensive Charge Reduction and Dissociation of Intact Protein Complexes Following Electron Transfer on a Quadrupole-Ion Mobility-Time-of-Flight MS" *Journal of The American Society for Mass Spectrometry* **2015,** *26*, 1068-1076.

60. Campuzano, I. D. G., et al. "Native MS Analysis of Bacteriorhodopsin and an Empty Nanodisc by Orthogonal Acceleration Time-of-Flight, Orbitrap and Ion Cyclotron Resonance" *Anal. Chem.* **2016,** *88*, 12427-12436.

61. Benesch, J. L. P. and Ruotolo, B. T. "Mass Spectrometry: Come of Age for Structural and Dynamical Biology" *Curr. Opin. Struct. Biol.* **2011,** *21*, 641-649.

62. Ewing, M. A., Glover, M. S. and Clemmer, D. E. "Hybrid Ion Mobility and Mass Spectrometry as a Separation Tool" *J. Chromatogr. A* **2016,** *1439*, 3-25.

63. Ewing, S. A., et al. "Collidoscope: An Improved Tool for Computing Collisional Cross-Sections with the Trajectory Method" *J. Am. Soc. Mass Spectrom.* **2017,** *28*, 587-596.

64. Donor, M. T., et al. "Extended Protein Ions Are Formed by the Chain Ejection Model in Chemical Supercharging Electrospray Ionization" *Anal. Chem.* **2017,** *89*, 5107-5114.

65. Li, H. L., et al. "Native Top-Down Electrospray Ionization-Mass Spectrometry of 158 kDa Protein Complex by High-Resolution Fourier Transform Ion Cyclotron Resonance Mass Spectrometry" *Anal. Chem.* **2014,** *86*, 317-320.

66. Laganowsky, A., et al. "Mass Spectrometry of Intact Membrane Protein Complexes" *Nat. Protoc.* **2013,** *8*, 639-651.

67. Zhou, M., Dagan, S. and Wysocki, V. H. "Protein Subunits Released by Surface Collisions of Noncovalent Complexes: Nativelike Compact Structures Revealed by Ion Mobility Mass Spectrometry" *Angew. Chem., Int. Ed.* **2012,** *51*, 4336-4339.

68. Sterling, H. J., et al. "Supercharging Protein Complexes from Aqueous Solution Disrupts their Native Conformations" *J. Am. Soc. Mass Spectrom.* **2012,** *23*, 191-200.

69. Heck, A. J. R. and van den Heuvel, R. H. H. "Investigation of Intact Protein Complexes by Mass Spectrometry" *Mass Spectrom. Rev.* **2004,** *23*, 368-389.

70. Loo, J. A. "Electrospray Ionization Mass Spectrometry: a Technology for Studying Noncovalent Macromolecular Complexes" *Int. J. Mass Spectrom.* **2000,** *200*, 175-186.

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

72. Konijnenberg, A., Butterer, A. and Sobott, F. "Native Ion mobility-Mass Spectrometry and Related Methods in Structural Biology" *Biochim. Biophys. Acta, Proteins Proteomics* **2013,** *1834*, 1239-1256.

73. Abzalimov, R. R. and Kaltashov, I. A. "Electrospray Ionization Mass Spectrometry of Highly Heterogeneous Protein Systems: Protein Ion Charge State Assignment via Incomplete Charge Reduction" *Anal. Chem.* **2010,** *82*, 7523-7526.

74. Zhou, M. and Wysocki, V. H. "Surface Induced Dissociation: Dissecting Noncovalent Protein Complexes in the Gas phase" *Acc. Chem. Res.* **2014,** *47*, 1010-1018.

75. Salbo, R., et al. "Traveling-Wave Ion Mobility Mass Spectrometry of Protein Complexes: Accurate Calibrated Collision Cross-Sections of Human Insulin Oligomers" *Rapid Commun. Mass Spectrom.* **2012,** *26*, 1181-1193.

76. Pan, J., et al. "Solution-Phase Chelators for Suppressing Nonspecific Protein−Metal Interactions in Electrospray Mass Spectrometry" *Anal. Chem.* **2009,** *81*, 5008-5015.

77. McKay, A. R., et al. "Mass Measurements of Increased Accuracy Resolve Heterogeneous Populations of Intact Ribosomes" *J. Am. Chem. Soc.* **2006,** *128*, 11433-11442.

78. Hopper, J. T. S., et al. "Detergent-Free Mass Spectrometry of Membrane Protein Complexes" *Nat. Methods* **2013,** *10*, 1206-1208.

79. Zhang, Y. X., et al. "Protein-Glycosphingolipid Interactions Revealed Using Catch-and-Release Mass Spectrometry" *Anal. Chem.* **2012,** *84*, 7618-7621.

80. Fouquet, T. and 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*, 2682-2686.

81. Causon, T. J. and Hann, S. "Theoretical Evaluation of Peak Capacity Improvements by use of Liquid Chromatography Combined with Drift Tube Ion Mobility-Mass Spectrometry" *J. Chromatogr. A* **2015,** *1416*, 47-56.

82. Arthur, K. L., et al. "Increasing Peak Capacity in Nontargeted Omics Applications by Combining Full Scan Field Asymmetric Waveform Ion Mobility Spectrometry with Liquid Chromatography - Mass Spectrometry" *Anal. Chem.* **2017,** *89*, 3452-3459.

83. Ruotolo, B. T., et al. "Ion Mobility-Mass Spectrometry Analysis of Large Protein Complexes" *Nat. Protoc.* **2008,** *3*, 1139-1152.

84. Clemmer, D. E. and Jarrold, M. F. "Ion Mobility Measurements and Their Applications to Clusters and Biomolecules" *J. Mass Spectrom.* **1997,** *32*, 577-592.

85. Hoi, K. K., Robinson, C. V. and Marty, M. T. "Unraveling the Composition and Behavior of Heterogeneous Lipid Nanodiscs by Mass Spectrometry" *Anal. Chem.* **2016,** *88*, 6199-6204.

86. Dwivedi, P., et al. "Gas-Phase Chiral Separations by Ion Mobility Spectrometry" *Anal. Chem.* **2006,** *78*, 8200-8206.

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

88. Zheng, H., et al. "Heuristic Charge Assignment for Deconvolution of Electrospray Ionization Mass Spectra" *Rapid Commun. Mass Spectrom.* **2003,** *17*, 429-436.

89. van Breukelen, B., et al. "Resolving Stoichiometries and Oligomeric States of Glutamate Synthase Protein Complexes with Curve Fitting and Simulation of Electrospray Mass Spectra" *Rapid Commun. Mass Spectrom.* **2006,** *20*, 2490-2496.

90. Stengel, F., et al. "Dissecting Heterogeneous Molecular Chaperone Complexes Using a Mass Spectrum Deconvolution Approach" *Chem. Biol.* **2012,** *19*, 599-607.

91. Danis, P. O. and Huby, F. J. "The Computer-Assisted Interpretation of Copolymer Mass Spectra" *J. Am. Soc. Mass Spectrom.* **1995,** *6*, 1112-1118.

92. Denisov, I. G. and Sligar, S. G. "Nanodiscs in Membrane Biochemistry and Biophysics" *Chem. Rev.* **2017,** *117*, 4669-4713.

93. Ritchie, T. K., et al. "Reconstitution of Membrane Proteins in Phospholipid Bilayer Nanodiscs" *Methods Enzymol.* **2009,** *464*, 211-231.

94. Cong, X., et al. "Determining Membrane Protein–Lipid Binding Thermodynamics Using Native Mass Spectrometry" *J. Am. Chem. Soc.* **2016,** *138*, 4346-4349.

95. Uetrecht, C., et al. "Interrogating Viral Capsid Assembly with Ion Mobility–Mass Spectrometry" *Nat. Chem.* **2011,** *3*, 126-132.

96. Lange, O., et al. "Enhanced Fourier transform for Orbitrap mass spectrometry" *Int. J. Mass Spectrom.* **2014,** *369*, 16-22.

97. Lu, J., et al. "Improved Peak Detection and Deconvolution of Native Electrospray Mass Spectra from Large Protein Complexes" *J. Am. Soc. Mass Spectrom.* **2015,** *26*, 2141-2151.

98. Dörr, J. M., et al. "Detergent-Free Isolation, Characterization, and Functional Reconstitution of a Tetrameric K^+ Channel: The Power of Native Nanodiscs" *Proc. Natl. Acad. Sci. U. S. A.* **2014,** *111*, 18607-18612.

99. Gao, Y., et al. "TRPV1 Structures in Nanodiscs Reveal Mechanisms of Ligand and Lipid Action" *Nature* **2016,** *doi: 10.1038/nature17964*.

100. Schultze, M., et al. "Attosecond Band-Gap Dynamics in Silicon" *Science* **2014,** *346*, 1348-1352.

101. Forbes, A. M. G. "Fourier Transform Filtering - a Cautionary Note" *J. Geophys. Res.-Oceans* **1988,** *93*, 6958-6962.

102. Mosierboss, P. A., Lieberman, S. H. and Newbery, R. "Fluorescence Rejection in Raman-Spectroscopy by Shifted-Spectra, Edge-Detection, and FFT Filtering Techniques" *Appl. Spectrosc.* **1995,** *49*, 630-638.

103. Pandey, P. R. and Roy, S. "Headgroup Mediated Water Insertion into the DPPC Bilayer: A Molecular Dynamics Study" *J. Phys. Chem. B* **2011,** *115*, 3155-3163.

104. Reid, D. J., et al. "MetaUniDec: High-Throughput Deconvolution of Native Mass Spectra" *J. Am. Soc. Mass Spectrom.* **2018**.

105. Light-Wahl, K. J., Schwartz, B. L. and Smith, R. D. "Observation of the Noncovalent Quaternary Associations of Proteins by Electrospray Ionization Mass Spectrometry" *J. Am. Chem. Soc.* **1994,** *116*, 5271-5278.

106. Kebarle, P. and Verkerk, U. H. "Electrospray: From ions in solution to ions in the gas phase, what we know now" *Mass Spectrom. Rev.* **2009,** *28*, 898-917.

107. Yen, T.-Y., Charles, M. J. and Voyksner, R. D. "Processes that affect electrospray ionization-mass spectrometry of nucleobases and nucleosides" *J. Am. Soc. Mass Spectrom.* **1996,** *7*, 1106-1108.

108. Felitsyn, N., Peschke, M. and Kebarle, P. "Origin and number of charges observed on multiply-protonated native proteins produced by ESI" *Int. J. Mass Spectrom.* **2002,** *219*, 39-62.

109. Iavarone, A. T., Udekwu, O. A. and Williams, E. R. "Buffer loading for counteracting metal salt-induced signal suppression in electrospray ionization" *Anal. Chem.* **2004,** *76*, 3944-3950.

110. Wang, G. and Cole, R. B. "Effect of Solution Ionic Strength on Analyte Charge State Distributions in Positive and Negative Ion Electrospray Mass Spectrometry" *Anal. Chem.* **1994,** *66*, 3702-3708.

111. Pan, P. and McLuckey, S. A. "The Effect of Small Cations on the Positive Electrospray Responses of Proteins at Low pH" *Anal. Chem.* **2003,** *75*, 5468-5474.

112. Mirza, U. A. and Chait, B. T. "Effects of Anions on the Positive Ion Electrospray Ionization Mass Spectra of Peptides and Proteins" *Anal. Chem.* **1994,** *66*, 2898-2904.

113. Liu, Y., et al. "Signal and Charge Enhancement for Protein Analysis by Liquid Chromatography-Mass Spectrometry with Desorption Electrospray Ionization" *Int. J. Mass Spectrom.* **2012,** *325-327*, 161-166.

114. Konermann, L. "Addressing a Common Misconception: Ammonium Acetate as Neutral pH “Buffer” for Native Electrospray Mass Spectrometry" *J. Am. Soc. Mass Spectrom.* **2017,** *28*, 1827-1835.

115. Streicher, W. W., Lopez, M. M. and Makhatadze, G. I. "Modulation of quaternary structure of S100 proteins by calcium ions" *Biophys. Chem.* **2010,** *151*, 181-186.

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

117. Sarkar, D., Singh, Y. and Kalia, J. "Protein–Lipid Interfaces Can Drive the Functions of Membrane-Embedded Protein–Protein Complexes" *ACS Chemical Biology* **2018,** *13*, 2689-2698.

118. Clifton, L. A., et al. Structural Investigations of Protein–Lipid Complexes Using Neutron Scattering. In *Lipid-Protein Interactions: Methods and Protocols*, J. H. Kleinschmidt, Springer New York: New York, NY, 2019; 201-251

119. Gupta, K., et al. "Identifying key membrane protein lipid interactions using mass spectrometry" *Nat. Protoc.* **2018,** *13*, 1106-1120.

120. Montenegro, F. A., Cantero, J. R. and Barrera, N. P. "Combining Mass Spectrometry and X-Ray Crystallography for Analyzing Native-Like Membrane Protein Lipid Complexes" *Front Physiol* **2017,** *8*, 892-892.

121. Wilson, J. W., et al. "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*, 10204-10211.

122. Cleary, S. P. and Prell, J. S. "Liberating Native Mass Spectrometry from Dependence on Volatile Salt Buffers by Use of Gábor Transform" *ChemPhysChem* **2019,** *20*, 519-523.

123. Cody, R. B. and Fouquet, T. "Paper spray and Kendrick mass defect analysis of block and random ethylene oxide/propylene oxide copolymers" *Anal. Chim. Acta* **2017,** *989*, 38-44.

124. Walker, L. R., et al. "Native Mass Spectrometry of Antimicrobial Peptides in Lipid Nanodiscs Elucidates Complex Assembly" *Anal. Chem.* **2019,** *91*, 9284-9291.

125. Marty, M. T., et al. "Native mass spectrometry characterization of intact nanodisc lipoprotein complexes" *Anal Chem* **2012,** *84*, 8957-8960.