Fourier-Transform Approach for Reconstructing Macromolecular Mass Defect Profiles

Andrew K. Swansiger,[1] Michael T. Marty,[2] and James S. Prell*[1,3]

Submitted to *Journal of the American Society for Mass Spectrometry*

21 October 2021

Revised 30 November 2021

1 Department of Chemistry and Biochemistry, 1253 University of Oregon, Eugene, OR 97403-1253, USA

2 Department of Chemistry and Biochemistry, 1306 University of Arizona, Tucson, AZ 85721, USA

3 Materials Science Institute, 1252 University of Oregon, Eugene, OR 97403-1252, USA

*Correspondence should be addressed to jprell@uoregon.edu

Tel: (541) 346-2597

Fax: (541) 346-4643

Abstract

State-of-the-art native mass spectrometry (MS) methods have been developed for analysis of highly heterogeneous intact complexes and have provided much insight into the structure and properties of non-covalent assemblies that can be difficult to study using denatured proteins. These native MS methods can often be used to study even highly polydisperse membrane proteins embedded in detergent micelles, Nanodiscs, and other membrane mimics. However, characterizing highly polydisperse native complexes which are also heterogeneous presents additional challenges for native MS. Macromolecular mass defect (MMD) analysis aims to characterize heterogeneous ion populations obfuscated by adduct polydispersity and revealing the distribution of "base" masses, and was recently implemented in the Bayesian analysis software UniDec. Here, we illustrate an alternative, orthogonal MMD analysis method implemented in the deconvolution program iFAMS, which takes advantage of Fourier Transform (FT) to deconvolve low-resolution data with few user-input parameters and which can provide high quality results even for mass spectra with a signal-to-noise ratio of ~5:1. Agreement between this method, which is based on frequency-domain data, and the mass-domain algorithm of UniDec provides strong evidence that both methods can accurately characterize highly polydisperse and heterogeneous ion populations. The FT algorithm is expected to be very useful in characterizing many types of analytes ranging from membrane proteins to polymer-conjugated proteins, branched polymers, and other large analytes, as well as for reconstructing isotope profiles for highly complex, but still isotope-resolved, mass spectra.

Introduction

Structural and compositional heterogeneity is important to the function of many types of natural and therapeutic biomolecules and biomolecular complexes, including protein complexes, protein-nucleic acid complexes, membrane protein-lipid assemblies, and antibodies.[1-7] With the growing importance of biotherapeutic proteins such as monoclonal antibodies, the need to accurately characterize intact protein complexes and their constituent isoforms for biomolecule characterization and biopharmaceutical design is rapidly growing. Native electrospray ionization mass spectrometry (ESI-MS) and ion mobility-mass spectrometry (IM-MS) have established themselves as highly sensitive tools for analysis of intact non-covalent complexes,[8-10] from antibody-drug conjugates and glycosylation states,[1, 11-13] to studies of membrane protein assembly structure and lipid binding,[14-20] and have been used to both update results from and inform sample preparation for more traditional crystallographic and microscopic structural analysis methods.[13, 21-22] The development of several MS deconvolution tools has even allowed for compositional studies of analytes with exceptionally high mass polydispersity (due to, e.g., variation in the number of a repeated subunits, such as glycosylation states of glycoproteins, lipids in a membrane or micelle, or polymer chain length).[9] These types of analytes often produce native-MS spectra with several overlapped charge states and many tens or even hundreds of overlapped peaks.[21, 23-25] Fourier Transform (FT)-based signal processing methods take advantage of sample polydispersity as a source of periodic signals that are used to determine ion charge states, subunit masses, composition, and total masses from the frequency domain.[21, 24, 26-29] Conversely, Bayesian statistical analysis methods as used in the software packages UniDec from the Marty group and PMI Intact from Protein Metrics, Inc., deconvolve mass spectra directly from the *m*/*z* domain by iteratively optimizing a model spectrum based on an initial set of user-input parameters.[30-34]

While deconvolution methods like those mentioned above can often deconvolve a mass spectrum to a "zero-charge" mass spectrum with many fewer peaks, it can still be very challenging to interpret the zero-charge mass spectrum for an ion population that is highly polydisperse in the stoichiometry of a particular adduct, especially when there is significant heterogeneity in the underlying "base" composition, i.e., the portion of each ion other than the polydisperse adducts. Recently, Marty and co-workers have reported on the preferred lipid

bilayer-incorporation stoichiometries of several antimicrobial peptides by macromolecular mass defect (MMD) analysis of UniDec reconstructed mass spectra,[35-38] using modular arithmetic to filter out polydisperse signal in a manner similar to how Kendrick mass defect analysis is used to characterize end groups of polymers.[39-41] Their studies of melittin incorporation into Nanodiscs, self-assembled membrane mimics consisting of a phospholipid bilayer encircled by two amphipathic membrane proteins (MSPs),[42-44] revealed no strong preferences for specific stoichiometries—contrasting with results from earlier studies which suggested a preference for melittin tetramers in lipid bilayers. These results provide an excellent basis from which to explore orthogonal MMD analysis methods both to validate the Bayesian results and to provide alternative means of generating MMD profiles, possibly with unique advantages.

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

Methods

Sample Preparation. All phospholipid Nanodiscs were prepared using a method adapted from that of Sligar and coworkers.[43-46] Nanodiscs containing only scaffold proteins and lipids were prepared at the University of Oregon, using water purified to 18 MΩ•cm resistivity. Briefly, dimyristoylphosphatidylcholine (DMPC) lipids purchased from Avanti Polar Lipids as 5 mg/mL suspensions in chloroform were dried with nitrogen gas until opaque, then re-suspended to a concentration of 50 mM in a pH 7.4 aqueous buffer composed of 100 mM sodium cholate

(Sigma-Aldrich, St. Louis, MO), 20 mM Tris (Bio-Rad, Hercules, CA), 100 mM sodium chloride, and 0.5 mM ethylenediaminetetraacetic acid. Histidine-tagged (MSP1D1His) and non-histidine-tagged (MSP1D1(−)) membrane scaffold proteins were reconstituted separately in identical pH 7.4 aqueous buffer without sodium cholate. For single-MSP experiments, these solutions were used directly. For mixed-MSP Nanodisc samples, equimolar mixtures of the two MSP1D1 variants were prepared at room temperature (~25 °C) in microcentrifuge tubes (Sorenson, Salt Lake City, UT), with a final protein concentration of ~200 μM. Each of the above MSP1D1 solutions were then mixed with DMPC-detergent suspensions and additional buffer for an expected Nanodisc concentration of 50 μM and incubated at room temperature for 15 minutes. Nanodisc self-assembly was started by 1000:1 dialysis by volume into buffer in which Bio-Beads SM-2 (Bio-Rad, Hercules, CA) were suspended to assist in removing cholate. Nanodisc samples were removed from dialysis after 24 hours and buffer-swapped twice via Micro Bio-Spin 6 columns (Bio-Rad, Hercules, CA) into 200 mM aqueous ammonium acetate (pH 7.0).

Data for melittin-embedded Nanodiscs prepared at the University of Arizona were taken from a previous publication, where their preparation is described in detail.[36] Briefly, dimyristoylphosphatidylglycerol (DMPG) Nanodisc samples with MSP1D1(−) scaffold proteins were prepared at room temperature similarly to the DMPC Nanodiscs described above, to a final concentration of 2.5 μM in 200 mM ammonium acetate (pH 6.8). Various dilutions of melittin (GenScript, Piscataway, NJ) in methanol were directly added to stock solutions of preformed Nanodiscs to final Nanodisc concentrations of 2 μM in 13% methanol by volume, with molar ratios of melittin to Nanodiscs ranging from 0 to 24:1. Finally, a stock solution of 400 mM imidazole was added to the melittin-Nanodisc solutions as a charge-reducing agent, to a final concentration of 25 mM.

Native Mass Spectrometry. Mixed-MSP1D1 DMPC Nanodisc native mass spectra were acquired at the University of Oregon with a Synapt G2-S*i* ion mobility-mass spectrometer (Waters Corp., Milford, MA) using a static nanoelectrospray ionization (nanoESI) source at a capillary voltage of ~0.8 kV relative to instrumental ground and Trap/Transfer collisional energies of 75-100/5 V. This relatively high Trap potential was used to dissociate enough salt adducts from the native Nanodisc ions to resolve different scaffold protein compositions in the

resulting mass spectra. NanoESI capillaries were prepared from 0.78 mm i.d. borosilicate capillary tubes on a Flaming-Brown P-97 micropipette puller (Sutter Instrument, Novato, CA) to a final i.d. of ~1 μm. Native mass spectra of melittin-embedded DMPG Nanodiscs were acquired at the University of Arizona in positive ion mode on a Q-Exactive HF quadrupole-Orbitrap mass spectrometer with Ultra-High Mass Range (UHMR) research modifications (Thermo Fisher Scientific, Waltham, MA) and a nanoESI source. Further experimental details for these melittin-embedded Nanodiscs are described in the original paper.[36]

Deconvolution and MMD Analysis. FT deconvolutions were performed via the Prell group's home-built program iFAMS v. 5.3. Lipid masses of 666.9 Da and 677.9 Da were used for DMPG and DMPC Nanodisc data, respectively, to determine the charge states present from FT spectra. The lipid polydispersity inherent to Nanodiscs results in multiple overlapped charge states in the mass spectrum, each with a broad distribution of lipid stoichiometries. However, the corresponding FT spectrum contains well-resolved peaks corresponding to fundamentals and harmonics for each charge state present, allowing for charge state deconvolution by inverse-FT of charge state-specific data.[24, 26, 28] MMD profiles for simulated data made use of 14 FT harmonics, while profiles of Synapt G2-S*i* data made use of 4 harmonics due to their lower resolution as compared to the Orbitrap UHMR data. All profiles for melittin-embedded Nanodiscs used 14 harmonics except for samples with melittin:Nanodisc ratios of 0 (10 harmonics), 1.5 (4 harmonics), and 24 (12 harmonics) due to differences in the density of data points in their mass spectra and the corresponding frequency range of their FT spectra.

Bayesian deconvolution was performed via UniDec v. 4.2.1, which generates a delta-function matrix to model the spectral data and iteratively adjusts charge state probabilities and ion abundances by comparing *m*/*z* peak intensities in the model and the spectral data.[31, 36] Input charge-state ranges were determined using iFAMS and lipid masses of 666.9 Da and 677.9 Da for DMPG and DMPC Nanodiscs, respectively, then entered in UniDec with two additional charge states above and below the found range. Output mass ranges were determined from the lowest *m*/*z* in the mass spectra multiplied by the lowest charge state and the highest *m*/*z* multiplied by the highest charge state. Melittin-embedded Nanodisc data were batch processed by MetaUniDec.[47] More detailed UniDec deconvolution parameters can be found in Tables S1 and S2.

MMD analysis was performed via both UniDec's built-in method, which uses the UniDec-deconvolved mass spectrum as input,[35-36] and a new FT-based method in iFAMS (see Theory section below) that extracts spectral information from the complex phase of the Fourier Transform. In both cases, peaks in the reconstructed MMD profiles were centroided by integration between local minima to assign mass defects and calculate root-mean-squared-deviations (RMSDs) from the expected base masses (based on the theoretical exact masses of the scaffold proteins, lipids, and melittin). Scaffold protein mass contributions were expected to be 24661.6 Da for MSP1D1His and 22043.9 Da for MSP1D1(−), as reported by the manufacturer. MMD profile mass accuracy was calculated only for peaks present in both the FT-based and Bayesian MMD profiles. Richardson-Lucy peak sharpening was used to explore melittin-Nanodisc stoichiometry conclusions made in the original publication of the data,[36] using UniDec's built-in "Raw/Centroid mode" for the Bayesian profiles and an analogous Richardson-Lucy treatment (50 iterations using a gaussian point-spread function with a standard deviation of 30 *m*/*z*) for the FT-based MMD profiles.

Theory

Fourier Phase Analysis for Macromolecular Mass Defect Profile Reconstruction. Fourier MMD treatment assumes that the population of ions for a given charge state consist of one or more non-isobaric "base compositions" (e.g., protein isoforms, different polymer end groups, or heterogeneous protein oligomer masses) to which a polydisperse distribution of repeated subunits (non-covalent adducts like lipids/detergents, or covalent adducts like glycans in a glycoprotein or monomers within a polymer) are adducted, and we call the masses of these base compositions "base masses". Let the *n* different base compositions be labeled B_1, B_2, …, B_n, and the repeated subunits have mass L, and finally assume that the net charge on each ion is supplied by charge carriers of mass X. Then, for charge state Z, we have a population of ions with stoichiometries $B_1L_jX_Z$, $B_2L_jX_Z$, …, $B_nL_jX_Z$, where j ranges over the (polydisperse) number of repeated subunits incorporated. It is assumed that, for each value of Z, the polydispersity in the number of lipids is sufficient to result in well-resolved peaks in the corresponding Fourier spectrum.[26] The goal of the present analysis is to determine the masses of B_1, B_2, and B_n *modulo* the mass of L as well as their relative abundances in the ion population. This procedure is

referred to as Macromolecular Mass Defect (MMD) analysis,[35-37] and the goal of this paper is to demonstrate how it can be achieved using Fourier/Gábor Transform (FT/GT)-based methods as well as compare its efficacy to mass-domain MMD analysis such as those implemented in UniDec.[30-32, 35-36, 47-48]

In the following, we use similar notation to that in our previous publications on Fourier and Gábor Transform analysis of mass spectra.[24, 26-28] The mass spectrum for a particular charge state is represented as $s\left(\frac{m}{z}\right) = [c\left(\frac{m}{z}\right) * p\left(\frac{m}{z}\right)] \cdot e(\frac{m}{z})$, where $c\left(\frac{m}{z}\right)$ is a comb function with peaks spaced by $\frac{L}{z}$; $p\left(\frac{m}{z}\right)$ is the "peak shape" representing the shape of each peak in the comb; and $e(\frac{m}{z})$ is the envelope function that describes the abundance distribution of each adduct stoichiometry, L_j. $[c\left(\frac{m}{z}\right) * p\left(\frac{m}{z}\right)]$ itself is a periodic function with period $\frac{L}{z}$ which inherently encodes the *m*/*z* spacing as well as the relative abundance of B_1, B_2, …, B_n and mass spectral peak widths associated with each of them; and $e(\frac{m}{z})$ is typically much broader in *m*/*z* than $\frac{L}{z}$ for Nanodiscs, long-chain polymers, and many similar highly polydisperse ion populations.[26] The Fourier Transform of $s\left(\frac{m}{z}\right)$ is $S(k) = [C(k) \cdot P(k)] * E(k)$, with $[C(k) \cdot P(k)]$ a periodic function of period $\frac{2\pi Z}{L}$, and $E(k)$ now the shape of each peak in the frequency spectrum $S(k)$ (typically much narrower than the frequency spacing, $\frac{2\pi Z}{L}$). Our goal is to extract $[C(k) \cdot P(k)]$ from $S(k)$ and invert the FT to recover the *m*/*z* spacing as well as the relative abundance of B_1, B_2, …, B_n (*mod* L) and mass spectral peak widths associated with each of them.

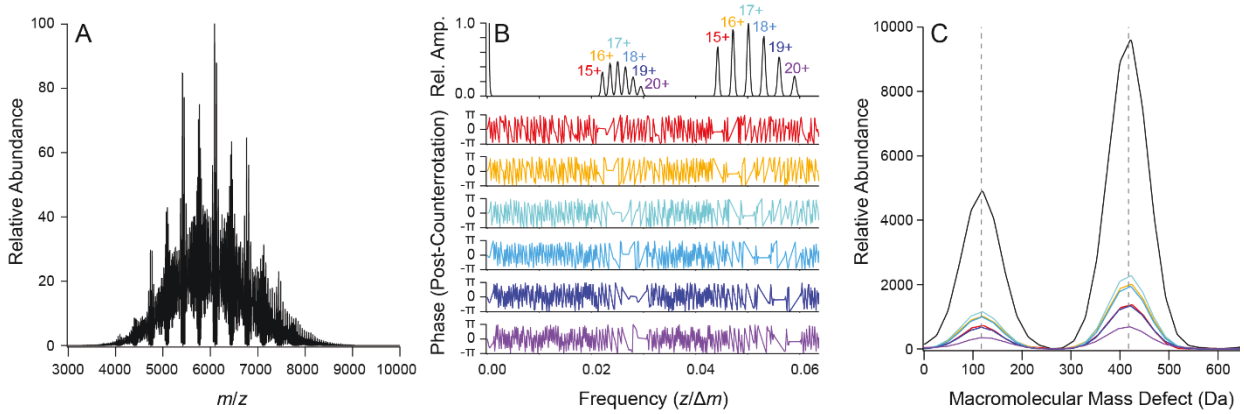


Figure 1 Illustration of MMD analysis on a simulated mass spectrum (A), containing charge states 15+ through 20+ of a polydisperse ion population containing multiple lipids with a mass of 678 Da and base masses with mass defects of 417 and 117 Da in a ratio of 2:1. (B) FT of the mass spectrum (top) with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

At first glance, this might appear challenging, as the phase of $S(k)$ can oscillate very rapidly in the frequency domain, depending on where the maximum of $e(\frac{m}{z})$ is in the mass spectrum. (In fact, in the common case that the mass spectrum of the polydisperse ion population looks relatively "well-centered" within the *m*/*z* window being viewed, $S(k)$ oscillates in phase as rapidly as possible, i.e., by approximately π radians at each successive point in the frequency domain!) However, as long as the curvature of $e(\frac{m}{z})$ near its maximum is sufficiently low (i.e., the polydisperse ion population is in fact not very monodisperse), the Stationary Phase Condition[49] makes it possible to counter-rotate the phase of $S(k)$ by multiplying it pointwise by $\exp\left(-ik\left(\frac{m}{z}\right)_0\right)$, where $\left(\frac{m}{z}\right)_0$ is an *m*/*z* value at or near the maximum of $e(\frac{m}{z})$. This is mathematically equivalent to "sliding over" the original mass spectrum to the left by $\left(\frac{m}{z}\right)_0$, such that the top of $e(\frac{m}{z})$ is now very close to 0 *m*/*z* (an explicit proof of this mathematical transformation can be found in the ESI). The phase of the resulting function $S'(k) = S(k) \cdot \exp\left(-ik\left(\frac{m}{z}\right)_0\right)$ will tend to be nearly flat at each peak in $[C(k) \cdot P(k)]$ as shown in the colored phase plots of Figure 1B, making it trivial to visually "read off" and/or numerically determine the phase and magnitude of $[C(k) \cdot P(k)]$ at each of its peaks, except possibly at $k = 0$, but it is

a theorem that the phase there must be 0 because the mass spectrum is real-valued. One needs to simply invert the recovered comb spectrum $[C(k) \cdot P(k)]$ to recover an image, averaged over all L_j states, of the *m*/*z* spacings and relative abundance of B_1, B_2, …, B_n (*mod* L) and their mass spectral peak widths. This spectrum is the MMD profile of the ion population. Unique advantages of this approach over "direct" approaches within in the *m*/*z* domain are that information from different charge states can in many cases be easily isolated and used to determine charge-state-specific MMD profiles, white noise at almost every frequency is eliminated, and a deconvolved "zero-charge" mass spectrum (in this case, *mod* L) can be readily reconstructed as desired from the individual charge-state-specific reconstructions.

Mathematically and empirically, two challenges that sometimes arise in this approach are identifying appropriate $(\frac{m}{z})_0$ values to use for each charge state and handling the ambiguity of frequency-domain charge state magnitudes at $k = 0$, where signal from all charge states overlap. In solving the first challenge, it is often useful to first analyze the FT spectrum of the initial mass spectral data (i.e., without counter-rotation) to reconstruct *m*/*z* envelopes for each charge state, as described in our previous publications.[24, 26, 28] These *m*/*z* envelopes can then be straightforwardly used to guess $(\frac{m}{z})_0$ needed for the phase reconstruction described above by computing the centroid of each envelope. This is generally easy to do, as long as $e(\frac{m}{z})$ has a clear maximum. The second challenge is circumvented by assuming both the phase and magnitude at k = 0 are zero. Because mass spectra are real valued by definition, the phase at $k = 0$ must be zero. We can also justify the zero-magnitude assumption for well-resolved ion populations with low baseline in the *m*/*z*-domain, where the integral over the mass spectrum and thus the FT magnitude at k = 0 for each charge state is relatively small and setting it equal to zero for each charge state before inverse FT merely adjusts the baseline of the reconstructed MMD profile by a small amount. For poorly resolved ion populations and/or those with a large noise baseline, the magnitude of the FT at k = 0 can be significant, and the apparent zero of abundance in the MMD profile will be systematically low by an indeterminate constant, although peak positions and shapes in the MMD are well-preserved. This baseline discrepancy also occurs in mass-domain MMD, where noise sums up across the spectra to yield a significant baseline, and in both cases can be compensated for by a linear baseline correction.

Results and Discussion

To explore the limits and robustness of the FT-based MMD analysis, ranging from signal-to-noise levels (Figure 2) to peak resolution (Table S3 and Figure S1), we start our discussion of the method with simulated spectra for which we know exactly the mass defects and their relative abundances. From there, we benchmark the method against the mass-domain MMD analysis integrated in UniDec using published, high-resolution Orbitrap spectra of melittin-embedded Nanodiscs (Figure 3). Finally, to illustrate the ability of the algorithm to handle lower-resolution data, as on currently available commercial Q-IM-MS instruments, we use it to investigate incorporation of Nanodisc MSPs with different masses using mass spectra acquired on a Synapt G2-S*i* Q-IM-TOF mass spectrometer (Figure 4).

Simulated Data. As a proof of concept, a model spectrum was constructed to contain charge states 15+ through 20+ of a polydisperse population of DMPC lipids (mass approximated to 678 Da) with a 2:1 ratio of base masses with MMDs (*mod* 678 Da) of 417 Da and 117 Da, respectively. The resultant total MMD profile peaks were of a similar full-width half-maximum to the peaks in the original model spectrum (~47.1 *m*/*z*) and also preserved the 2:1 ratio of mass defect relative abundances. The analysis was repeated with several levels of white noise relative to the model spectrum's global maximum, returning nearly identical results to those from a mass spectrum with a signal-to-noise ratio (SNR) of 5:1 (Figure 2, black), well below the commonly accepted limit of quantitation (SNR of 10:1).[50] The ability of the FT-based MMD algorithm to perform well at such a low SNR arises from elimination of white noise contributions at almost all frequencies other than at the harmonic peaks where the phase is sampled. However, the fewer harmonics that are sampled, the fewer total data points the MMD profile contains (reducing profile resolution) and the more likely it is, in principle, that artifacts may dominate the MMD profile. Due to the ambiguity of charge-state-specific Fourier-domain amplitude at 0 frequency, as described above, a baseline correction was applied to the reconstructed MMD profiles to account for relative abundance distortion from the summation of charge-state-specific MMD reconstructions. The difference between applying baseline correction to the individual charge state mass defect profiles and applying the correction to the summation of the charge state mass defect profiles was found to be negligible (see Figure S2). These results indicate that this FT-based MMD algorithm can be exceptionally useful for MMD analysis of low SNR data due to its high noise tolerance.

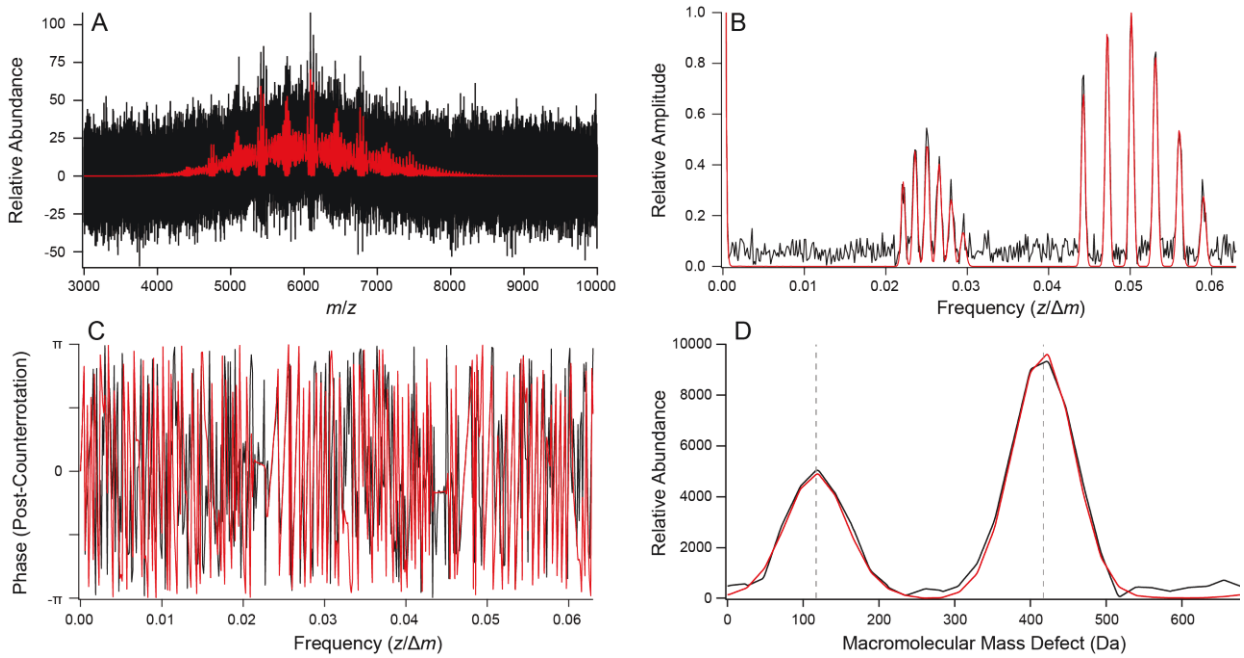


Figure 2 MMD analysis of simulated data containing charge states 15+ through 20+ of a polydisperse ion population containing multiple 678 Da lipids and base masses with mass defects of 417 Da and 117 Da in a 2:1 ratio. Red/black traces correspond to infinite/5:1 SNR. Simulated mass spectra (A), corresponding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

Melittin-Embedded Nanodiscs. To compare the performance of the FT-based method with mass-domain MMD analysis, previously published experimental data collected by the Marty lab for melittin-embedded DMPG Nanodiscs were analyzed via both iFAMS and UniDec, using 666.9 Da as the molecular weight of DMPG. In the original publication,[36] the MMD profile information was distilled in plots tracking changes in the relative abundance of incorporated melittin stoichiometries as a function of bulk solution melittin:Nanodisc ratios, which we have recreated here for both UniDec and iFAMS MMD profile series. Features in each were assigned to the same predicted mass defects for the sake of comparison. In these data, most of the MMD information has a resolution (defined as peak separation divided by peak fwhm) of 1.5 or greater, indicating that good results are obtained at this resolution despite the overlap of many charge states and individual peaks in the corresponding polydisperse mass spectra, where adjacent peaks are often strongly overlapped. (An upper bound to the lowest tractable resolution in MMD profiles is identified in Figure S1, where a distinct shoulder peak can still be discerned

at resolutions as low as 1.06 despite the 2:1 difference in peak height.) The RMSDs and melittin-incorporation plots generated from MMD peak integration reflect the remarkable consistency with which the FT phase algorithm mirrors the results from UniDec for all 7 mass spectra, representing bulk melittin:Nanodisc molar ratios ranging from 0:1 to 24:1. Mass defects and relative abundances using both methods (see Table S4) were within the uncertainties of the previously published UniDec results.[36]

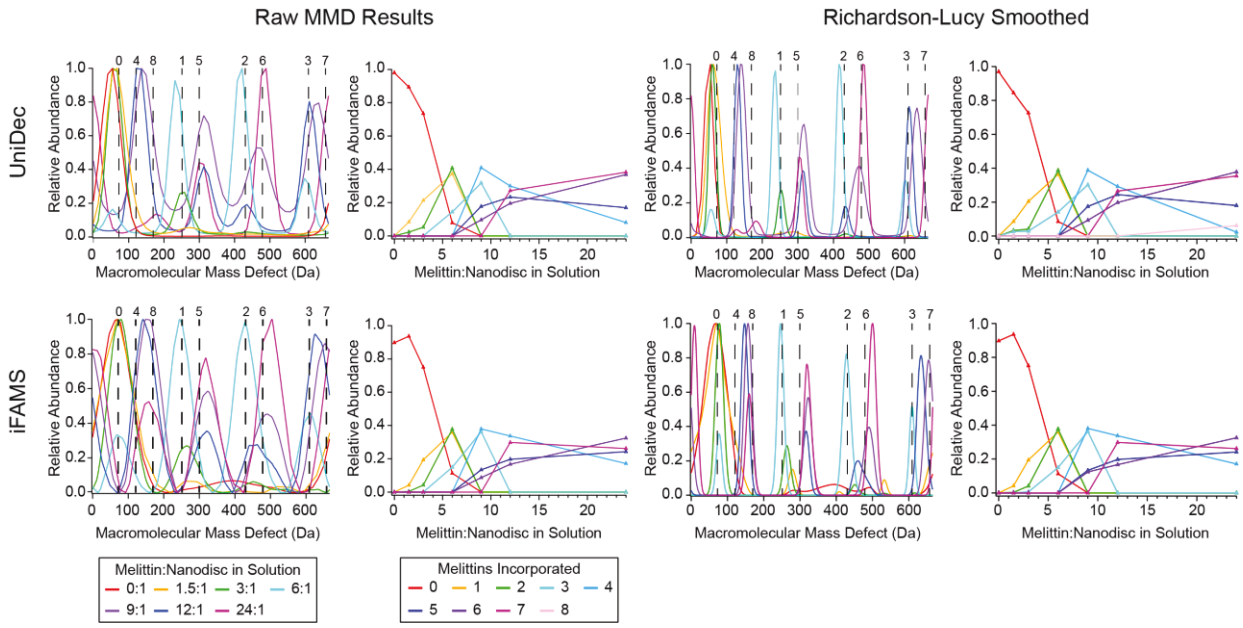


Figure 3 Comparison of melittin-embedded DMPG Nanodisc data deconvolved via UniDec (top) and iFAMS (bottom). MMD profiles are normalized to their global maximum. Dashed lines in MMD profiles indicate the expected mass defects for various numbers of incorporationed melittin. MMD profile color legend (left) and x-axis of the melittin-incorporation plots indicate the bulk solution melittin-to-Nanodisc ratios prepared, while the melittin-incorporation plot color legend (right) signifies the melittin-incorporation ratio being tracked as a function of bulk concentrations

Minor differences between UniDec and iFAMS MMD profiles include the appearance of a broad, low-abundance peak around 400 Da in the iFAMS profiles for samples containing no melittin that is much larger than its counterpart in the UniDec MMD profile. This minor peak was not assigned in the original publication and is most likely contributed by a small population of cholate (expected mass of 407.56 Da) in the Nanodiscs which could be removed with further dialysis. A second minor difference between the UniDec and iFAMS MMD profiles shown in Figure 3 is the presence of a single peak at 180 Da in the iFAMS and UniDec MMD profiles for

the solution with a bulk concentration ratio of 24 melittin monomers per Nanodisc, which appears as two lower abundance peaks in the published UniDec MMD profile. These discrepancies are attributed to artifactual splitting of broad peaks that occurs when using a much narrower target peak shape, a well-known occurrence when using Richardson-Lucy sharpening algorithms.[48, 51]

Comparison of the MMD profiles presented in Figure 3 after Richardson-Lucy peak sharpening reveals another minor difference between results from the two MMD analysis methods, namely, that the peak at ~180 Da in the UniDec reconstruction assigned to incorporation of 8 melittins for the 24:1 bulk melittin:Nanodisc sample is absent in the iFAMS MMD reconstruction. This is attributed to slight differences in the breadth of the un-sharpened peak, which splits into two peaks after Richardson-Lucy sharpening for the UniDec data but remains a single, narrow feature for the iFAMS reconstruction. This result indicates that care must be exercised in interpreting results from Richardson-Lucy sharpening for initially broad peaks. From the results both with and without Richardson-Lucy sharpening, we find generally excellent agreement between both the MMD profiles and melittin incorporation profiles as a function of bulk melittin:Nanodisc concentration for both the FT phase reconstruction method in iFAMS and the mass-domain results from UniDec. These results confirm the previously published conclusion that melittin incorporation is essentially non-specific in these samples.

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

present in solution (see Figures S3 and S4), we addressed the broader question of what Nanodisc MSP compositions are obtained when His-tagged and non-His-tagged MSP's are used to synthesize Nanodiscs in solution.

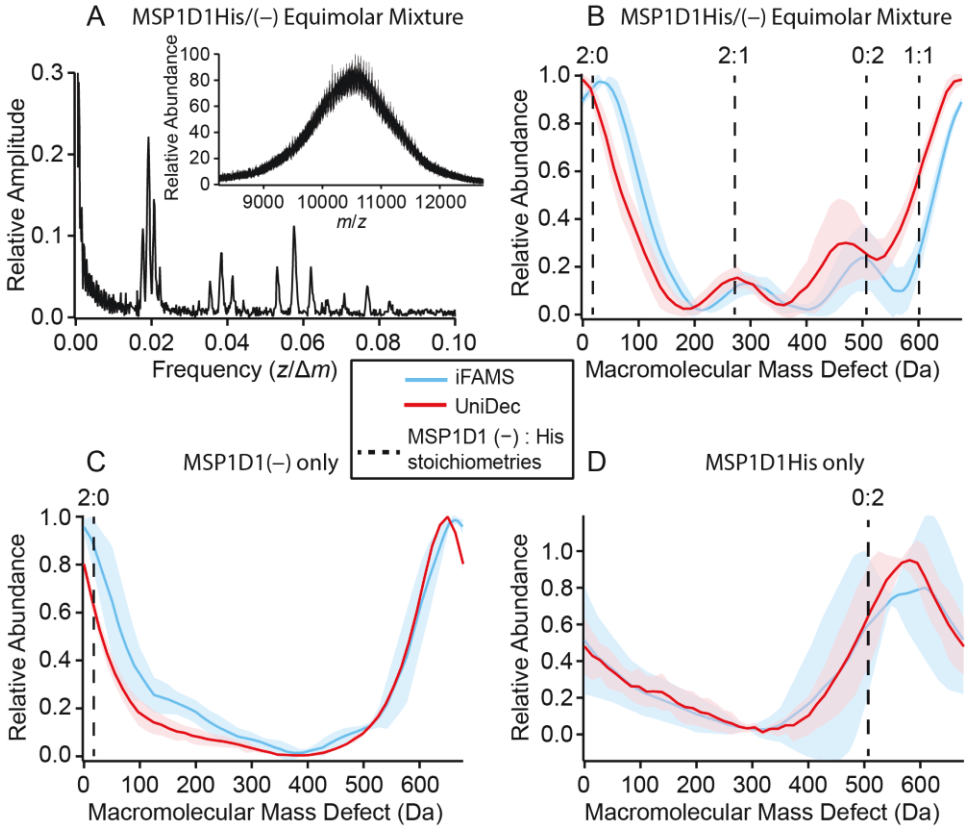


Figure 4 MMD analysis of "empty" MSP1D1 Nanodiscs with a variety of MSP incorporation states. (A)FT of Nanodiscs samples formed from a mixture of MSP1D1(−) and MSP1D1His. The inset shows the original mass spectrum truncated to just the Nanodisc distribution, serving as a reminder of the degree of charge state overlap from which the well-resolved charge state information in the FT is produced. Average MMD profiles of triplicate data deconvolved via both UniDec (red) and iFAMS (blue) are shown for MSP1D1 Nanodiscs prepared with both variants of MSP1D1 (B) and with only MSP1D1(−) (C) or MSP1D1His (D). Shaded regions denote 1 standard deviation in the abundance of the triplicate data and dashed lines indicate expected mass defects from various stoichiometries of MSP1D1(−) and MSP1D1His.

DMPC Nanodisc solutions prepared from a mixture of MSP1D1His and MSP1D1(−) were analyzed to determine whether the resulting Nanodiscs demonstrate a preference in MSP incorporation (see average MMD profiles in Figure 4). Despite the degree of charge state overlap observed in the mixed MSP Nanodisc mass distribution (inset of Figure 4A) the charge states are well-resolved in the frequency domain, allowing for identification of multiple base compositions from the MMD profiles. The dominant peaks for all samples correspond to MSP homodimers

(expected *mod* 677.9 Da mass defects of 24.3 Da and 514.4 Da for MSP1D1(−) homodimer and MSP1D1His homodimer, respectively), and there is little to no evidence heterodimers at the expected mass defect of 601.5 Da. The peak around 281.5 Da in the mixed MSP profiles indicates the presence of Nanodiscs containing three MSP monomers, a phenomenon which has been reported previously[52] and is believed to have physiological analogs (i.e., in mature lipoprotein particles containing three monomers of apolipoprotein A-1 rather than MSP1D1).[53] Although relative abundances for the triplicate experiments exhibit enough variability to make quantitation of the different protein stoichiometries difficult, peaks corresponding to at least three and possibly four different protein stoichiometries are clearly visible among the MMD profiles. Overall, the agreement of the measured MMD peak positions with their theoretical values is somewhat worse for these lower-resolution data than for the high-resolution melittin-Nanodisc data described above, presumably due to salt or other co-solute adduction in combination with the poorer resolving power of the Q-IM-TOF instrument. However, agreement of the FT-based algorithm results with theory is similar to or, in one case (the 514.4 Da peak), slightly better than the mass-domain algorithm (see Figure 4B).

Conclusions

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

The FT phase-based method presented here has potential applications in extracting isoform profiles from a variety of heterogeneous samples exhibiting polydispersity, from branched polymers and polymer-conjugated protein therapeutics used to increase drug longevity *in vivo*[54] to membrane protein-lipid/detergent complexes, as well as reconstructing isotope profiles for highly complex mass spectra which still exhibit isotope resolution. The new algorithm also provides a straightforward means of confirming analyses obtained from other orthogonal approaches, such as the mass-domain methods in UniDec, PMI Intact, and other deconvolution tools.[9, 55] The FT MMD reconstruction algorithm described above has been incorporated into iFAMS v. 6.1 with both FT and GT-based implementations, which is publicly available as an open-source Python program at https://github.com/prellgroup/iFAMS/releases. Recent versions of UniDec are publicly available as an open-source Python program at https://github.com/michaelmarty/UniDec/releases.

Associated Content

Supporting Information: Mathematical proof of global phase elimination; tables of detailed deconvolution parameters and statistical data for MMD profile peak identification; Figure S1, demonstration of the effects of mass spectral peak superposition on MMD peak resolution; Figure S2, comparison of baseline correction procedures; Figures S3 and S4, ESI-MS spectra of MSP1D1(−) only and MSP1D1His only (PDF)

Conflict of Interest

Authors declare no conflicting financial interests.

Acknowledgements

Research reported in this publication was supported by the National Science Foundation under award number CHE-1752994 to J.S.P. and award number CHE-1845230 to M.T.M. and the American Society for Mass Spectrometry Research Award (to J.S.P.).

References

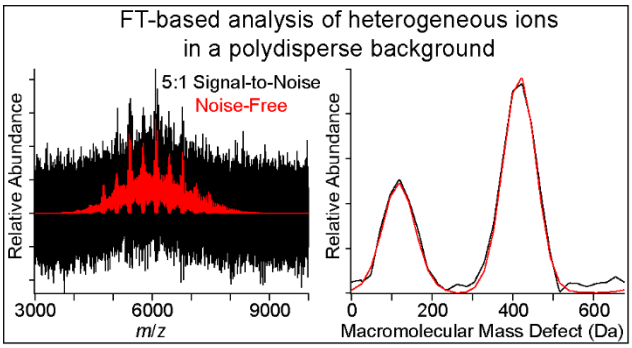
1. Struwe, W. B.; Robinson, C. V., Relating glycoprotein structural heterogeneity to function – insights from native mass spectrometry. *Curr. Opin. Struct. Biol.* **2019,** *58*, 241-248.
2. Wu, D.; Struwe, W. B.; Harvey, D. J.; Ferguson, M. A. J.; Robinson, C. V., N-glycan microheterogeneity regulates interactions of plasma proteins. *Proc. Natl. Acad. Sci. U. S. A.* **2018,** *115* (35), 8763-8768.
3. Wohlschlager, T.; Scheffler, K.; Forstenlehner, I. C.; Skala, W.; Senn, S.; Damoc, E.; Holzmann, J.; Huber, C. G., Native mass spectrometry combined with enzymatic dissection unravels glycoform heterogeneity of biopharmaceuticals. *Nat. Commun.* **2018,** *9* (1), 1713.
4. Gupta, K.; Donlan, J. A. C.; Hopper, J. T. S.; Uzdavinys, P.; Landreh, M.; Struwe, W. B.; Drew, D.; Baldwin, A. J.; Stansfeld, P. J.; Robinson, C. V., The role of interfacial lipids in stabilizing membrane protein oligomers. *Nature* **2017,** *541* (7637), 421-424.
5. Laganowsky, A.; Reading, E.; Allison, T. M.; Ulmschneider, M. B.; Degiacomi, M. T.; Baldwin, A. J.; Robinson, C. V., Membrane proteins bind lipids selectively to modulate their structure and function. *Nature* **2014,** *510* (7503), 172-175.
6. van de Waterbeemd, M.; Snijder, J.; Tsvetkova, I. B.; Dragnea, B. G.; Cornelissen, J. J.; Heck, A. J. R., Examining the Heterogeneous Genome Content of Multipartite Viruses BMV and CCMV by Native Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **2016,** *27* (6), 1000-1009.
7. Agasid, M. T.; Sørensen, L.; Urner, L. H.; Yan, J.; Robinson, C. V., The Effects of Sodium Ions on Ligand Binding and Conformational States of G Protein-Coupled Receptors—Insights from Mass Spectrometry. *J. Am. Chem. Soc.* **2021,** *143* (11), 4085-4089.
8. Boeri Erba, E.; Petosa, C., The emerging role of native mass spectrometry in characterizing the structure and dynamics of macromolecular complexes. *Protein Sci.* **2015,** *24* (8), 1176-1192.
9. Rolland, A. D.; Prell, J. S., Approaches to Heterogeneity in Native Mass Spectrometry. *Chem. Rev.* **2021**.
10. Ruotolo, B. T.; Marty, M. T., Native mass spectrometry for structural biology: A perspective. *Int. J. Mass Spectrom.* **2021,** *468*, 116655.
11. Thompson, N. J.; Rosati, S.; Rose, R. J.; Heck, A. J. R., The impact of mass spectrometry on the study of intact antibodies: from post-translational modifications to structural analysis. *Chem. Commun.* **2013,** *49* (6), 538-548.
12. Campuzano, I. D. G.; Netirojjanakul, C.; Nshanian, M.; Lippens, J. L.; Kilgour, D. P. A.; Van Orden, S.; Loo, J. A., Native-MS Analysis of Monoclonal Antibody Conjugates by Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. *Anal. Chem.* **2018,** *90* (1), 745-751.
13. Marculescu, C.; Lakshminarayanan, A.; Gault, J.; Knight, J. C.; Folkes, L. K.; Spink, T.; Robinson, C. V.; Vallis, K.; Davis, B. G.; Cornelissen, B., Probing the limits of Q-tag bioconjugation of antibodies. *Chem. Commun.* **2019,** *55* (76), 11342-11345.
14. Rolland, A. D.; Prell, J. S., Computational insights into compaction of gas-phase protein and protein complex ions in native ion mobility-mass spectrometry. *Trends Anal. Chem.* **2019,** *116*, 282-291.
15. Landreh, M.; Costeira-Paulo, J.; Gault, J.; Marklund, E. G.; Robinson, C. V., Effects of Detergent Micelles on Lipid Binding to Proteins in Electrospray Ionization Mass Spectrometry. *Anal. Chem.* **2017,** *89* (14), 7425-7430.

16. Landreh, M.; Marty, M. T.; Gault, J.; Robinson, C. V., A sliding selectivity scale for lipid binding to membrane proteins. *Curr. Opin. Struct. Biol.* **2016,** *39*, 54-60.
17. Gupta, K.; Li, J.; Liko, I.; Gault, J.; Bechara, C.; Wu, D.; Hopper, J. T. S.; Giles, K.; Benesch, J. L. P.; Robinson, C. V., Identifying key membrane protein lipid interactions using mass spectrometry. *Nat. Protoc.* **2018,** *13* (5), 1106-1120.
18. Patrick, J. W.; Boone, C. D.; Liu, W.; Conover, G. M.; Liu, Y.; Cong, X.; Laganowsky, A., Allostery revealed within lipid binding events to membrane proteins. *Proc. Natl. Acad. Sci. U. S. A.* **2018,** *115* (12), 2976-2981.
19. Lippens, J. L.; Nshanian, M.; Spahr, C.; Egea, P. F.; Loo, J. A.; Campuzano, I. D. G., Fourier Transform-Ion Cyclotron Resonance Mass Spectrometry as a Platform for Characterizing Multimeric Membrane Protein Complexes. *J. Am. Soc. Mass Spectrom.* **2018,** *29* (1), 183-193.
20. Wörner, T. P.; Snijder, J.; Bennett, A.; Agbandje-McKenna, M.; Makarov, A. A.; Heck, A. J. R., Resolving heterogeneous macromolecular assemblies by Orbitrap-based single-particle charge detection mass spectrometry. *Nat. Methods* **2020,** *17* (4), 395-398.
21. Wilson, J. W.; Rolland, A. D.; Klausen, G. M.; Prell, J. S., Ion Mobility-Mass Spectrometry Reveals That α-Hemolysin from *Staphylococcus aureus* Simultaneously Forms Hexameric and Heptameric Complexes in Detergent Micelle Solutions. *Anal. Chem.* **2019,** *91* (15), 10204-10211.
22. Olinares, P. D. B.; Kang, J. Y.; Llewellyn, E.; Chiu, C.; Chen, J.; Malone, B.; Saecker, R. M.; Campbell, E. A.; Darst, S. A.; Chait, B. T., Native Mass Spectrometry-Based Screening for Optimal Sample Preparation in Single-Particle Cryo-EM. *Structure* **2021,** *29* (2), 186-195.e6.
23. Marty, M. T.; Zhang, H.; Cui, W. D.; Blankenship, R. E.; Gross, M. L.; Sligar, S. G., Native Mass Spectrometry Characterization of Intact Nanodisc Lipoprotein Complexes. *Anal. Chem.* **2012,** *84* (21), 8957-8960.
24. Cleary, S. P.; Thompson, A. M.; Prell, J. S., Fourier Analysis Method for Analyzing Highly Congested Mass Spectra of Ion Populations with Repeated Subunits. *Anal. Chem.* **2016,** *88* (12), 6205-6213.
25. Campuzano, I. D. G.; Robinson, J. H.; Hui, J. O.; Shi, S. D. H.; Netirojjanakul, C.; Nshanian, M.; Egea, P. F.; Lippens, J. L.; Bagal, D.; Loo, J. A.; Bern, M., Native and Denaturing MS Protein Deconvolution for Biopharma: Monoclonal Antibodies and Antibody–Drug Conjugates to Polydisperse Membrane Proteins and Beyond. *Anal. Chem.* **2019,** *91* (15), 9472-9480.
26. Cleary, S. P.; Li, H.; Bagal, D.; Loo, J. A.; Campuzano, I. D. G.; Prell, J. S., Extracting Charge and Mass Information from Highly Congested Mass Spectra Using Fourier-Domain Harmonics. *J. Am. Soc. Mass Spectrom.* **2018,** *29* (10), 2067-2080.
27. Cleary, S. P.; Prell, J. S., Liberating Native Mass Spectrometry from Dependence on Volatile Salt Buffers by Use of Gábor Transform. *ChemPhysChem* **2019,** *20* (4), 519-523.
28. Cleary, S. P.; Prell, J. S., Distinct classes of multi-subunit heterogeneity: analysis using Fourier Transform methods and native mass spectrometry. *Analyst* **2020,** *145* (13), 4688-4697.
29. Prebyl, B. S.; Cook, K. D., Use of Fourier Transform for Deconvolution of the Unresolved Envelope Observed in Electrospray Ionization Mass Spectrometry of Strongly Ionic Synthetic Polymers. *Anal. Chem.* **2004,** *76* (1), 127-136.
30. Kostelic, M. M.; Marty, M. T., Deconvolving Native and Intact Protein Mass Spectra with UniDec. *ChemRxiv* **2020**, DOI: 10.26434/chemrxiv.13417118.v1.

31. Marty, M. T.; Baldwin, A. J.; Marklund, E. G.; Hochberg, G. K. A.; Benesch, J. L. P.; Robinson, C. V., Bayesian Deconvolution of Mass and Ion Mobility Spectra: From Binary Interactions to Polydisperse Ensembles. *Anal. Chem.* **2015,** *87* (8), 4370-4376.
32. Marty, M. T.; Zhang, H.; Cui, W. D.; Gross, M. L.; Sligar, S. G., Interpretation and Deconvolution of Nanodisc Native Mass Spectra. *J. Am. Soc. Mass Spectrom.* **2014,** *25* (2), 269-277.
33. Hoi, K. K.; Robinson, C. V.; Marty, M. T., Unraveling the Composition and Behavior of Heterogeneous Lipid Nanodiscs by Mass Spectrometry. *Anal. Chem.* **2016,** *88* (12), 6199-6204.
34. Bern, M.; Caval, T.; Kil, Y. J.; Tang, W.; Becker, C.; Carlson, E.; Kletter, D.; Sen, K. I.; Galy, N.; Hagemans, D.; Franc, V.; Heck, A. J. R., Parsimonious Charge Deconvolution for Native Mass Spectrometry. *J. Proteome Res.* **2018,** *17* (3), 1216-1226.
35. Marty, M. T.; Hoi, K. K.; Gault, J.; Robinson, C. V., Probing the Lipid Annular Belt by Gas-Phase Dissociation of Membrane Proteins in Nanodiscs. *Angew. Chem. Int. Edit.* **2016,** *55* (2), 550-554.
36. Walker, L. R.; Marzluff, E. M.; Townsend, J. A.; Resager, W. C.; Marty, M. T., Native Mass Spectrometry of Antimicrobial Peptides in Lipid Nanodiscs Elucidates Complex Assembly. *Anal. Chem.* **2019,** *91* (14), 9284-9291.
37. Keener, J. E.; Zambrano, D. E.; Zhang, G.; Zak, C. K.; Reid, D. J.; Deodhar, B. S.; Pemberton, J. E.; Prell, J. S.; Marty, M. T., Chemical Additives Enable Native Mass Spectrometry Measurement of Membrane Protein Oligomeric State within Intact Nanodiscs. *J. Am. Chem. Soc.* **2019,** *141* (2), 1054-1061.
38. Townsend, J. A.; Keener, J. E.; Miller, Z. M.; Prell, J. S.; Marty, M. T., Imidazole Derivatives Improve Charge Reduction and Stabilization for Native Mass Spectrometry. *Anal. Chem.* **2019,** *91* (22), 14765-14772.
39. Kendrick, E., A Mass Scale Based on CH_2 = 14.0000 for High Resolution Mass Spectrometry of Organic Compounds. *Anal. Chem.* **1963,** *35* (13), 2146-2154.
40. Hughey, C. A.; Hendrickson, C. L.; Rodgers, R. P.; Marshall, A. G.; Qian, K., Kendrick Mass Defect Spectrum: A Compact Visual Analysis for Ultrahigh-Resolution Broadband Mass Spectra. *Anal. Chem.* **2001,** *73* (19), 4676-4681.
41. Fouquet, T.; Sato, H., Extension of the Kendrick Mass Defect Analysis of Homopolymers to Low Resolution and High Mass Range Mass Spectra Using Fractional Base Units. *Anal. Chem.* **2017,** *89* (5), 2682-2686.
42. Civjan, N. R.; Bayburt, T. H.; Schuler, M. A.; Sligar, S. G., Direct solubilization of heterologously expressed membrane proteins by incorporation into nanoscale lipid bilayers. *BioTechniques* **2003,** *35* (3), 556-563.
43. Denisov, I. G.; Sligar, S. G., Nanodiscs for structural and functional studies of membrane proteins. *Nat. Struct. Mol. Biol.* **2016,** *23* (6), 481-486.
44. Denisov, I. G.; Sligar, S. G., Nanodiscs in Membrane Biochemistry and Biophysics. *Chem. Rev.* **2017,** *117* (6), 4669-4713.
45. Denisov, I. G.; Grinkova, Y. V.; Lazarides, A. A.; Sligar, S. G., Directed Self-Assembly of Monodisperse Phospholipid Bilayer Nanodiscs with Controlled Size. *J. Am. Chem. Soc.* **2004,** *126* (11), 3477-3487.
46. Ritchie, T. K.; Grinkova, Y. V.; Bayburt, T. H.; Denisov, I. G.; Zolnerciks, J. K.; Atkins, W. M.; Sligar, S. G., Reconstitution of Membrane Proteins in Phospholipid Bilayer Nanodiscs. *Methods Enzymol.* **2009,** *464*, 211-231.

47. Reid, D. J.; Diesing, J. M.; Miller, M. A.; Perry, S. M.; Wales, J. A.; Montfort, W. R.; Marty, M. T., MetaUniDec: High-Throughput Deconvolution of Native Mass Spectra. *J. Am. Soc. Mass Spectrom.* **2018,** *30*, 118-127.
48. Marty, M. T., Eliminating Artifacts in Electrospray Deconvolution with a SoftMax Function. *J. Am. Soc. Mass Spectrom.* **2019,** *30* (10), 2174-2177.
49. Bleistein, N.; Handelsman, R. A., Fourier integrals and the method of stationary phase. In *Asymptotic expansions of integrals*, 1st ed.; Holt, Rinehart and Winston: New York, NY, 1975; pp 219-223.
50. Harris, D. C., Quality assurance and calibration methods. In *Quantitative chemical analysis*, 8th ed.; Byrd, M. L., Ed. Marshall, Clancy: New York, NY, 2010; pp 100-105.
51. Zeng, G. L., Gibbs artifact reduction by nonnegativity constraint. *J. Nucl. Med. Technol.* **2011,** *39* (3), 213-219.
52. Kostelic, M. M.; Zak, C. K.; Jayasekera, H.; Marty, M. T., Assembly of Model Membrane Nanodiscs for Native Mass Spectrometry. *Anal. Chem.* **2021,** *93* (14), 5972-5979.
53. Murphy, A. J.; Hoang, A.; Aprico, A.; Sviridov, D.; Chin-Dusting, J., Anti-Inflammatory Functions of Apolipoprotein A-I and High-Density Lipoprotein Are Preserved in Trimeric Apolipoprotein A-I. *J. Pharmacol. Exp. Ther.* **2013,** *344* (1), 41-49.
54. Cerofolini, L.; Giuntini, S.; Carlon, A.; Ravera, E.; Calderone, V.; Fragai, M.; Parigi, G.; Luchinat, C., Characterization of PEGylated Asparaginase: New Opportunities from NMR Analysis of Large PEGylated Therapeutics. *Chem. Eur. J.* **2019,** *25* (8), 1984-1991.
55. Ferrige, A. G.; Seddon, M. J.; Jarvis, S.; Skilling, J.; Aplin, R., Maximum entropy deconvolution in electrospray mass spectrometry. *Rapid Commun. Mass Spectrom.* **1991,** *5* (8), 374-377.

For Table of Contents Only



Supporting Information for:

Fourier-Transform Approach for Reconstructing Macromolecular Mass Defect Profiles

Andrew K. Swansiger,[1] Michael T. Marty,[2] and James S. Prell*[1,3]

Submitted to *Journal of the American Society for Mass Spectrometry*

21 October 2021

Revised 30 November 2021

1 Department of Chemistry and Biochemistry, 1253 University of Oregon, Eugene, OR 97403-1253, USA

2 Department of Chemistry and Biochemistry, 1306 University of Arizona, Tucson, AZ 85721, USA

3 Materials Science Institute, 1252 University of Oregon, Eugene, OR 97403-1252, USA

*Correspondence should be addressed to jprell@uoregon.edu

Tel: (541) 346-2597

Fax: (541) 346-4643

Table of Contents:

Mathematical Proof

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

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

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

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

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

By the Fourier Convolution Theorem,

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

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

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

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

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

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

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

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

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

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

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

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

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

Equivalently,

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

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

To combine $p(x;Z)$ from all observed values of Z into a single charge-state-averaged MMD profile $p(x)$, one can simply translate each $p(x;Z)$ cyclically to the left by (mass of the charge carrier/1), normalize for charge by replacing x by Zx, and interpolate and sum the resulting "zero-charge" MMD profiles over the set of all observed Z. Note that the axis transformation $x \mapsto Zx$ naturally accounts for the $1/Z$ scaling of $e(x; x_{max}(Z), Z)$ such that each charge state's contribution to the summed MMD profile ($p(x)$) is simply proportional to its total ion population.

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

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

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

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

Table S2 List of all UniDec parameters of used for spectral deconvolution of melittin-embedded Nanodisc data, in order as they appear in UniDec Version 4.2.1.

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

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

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

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

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

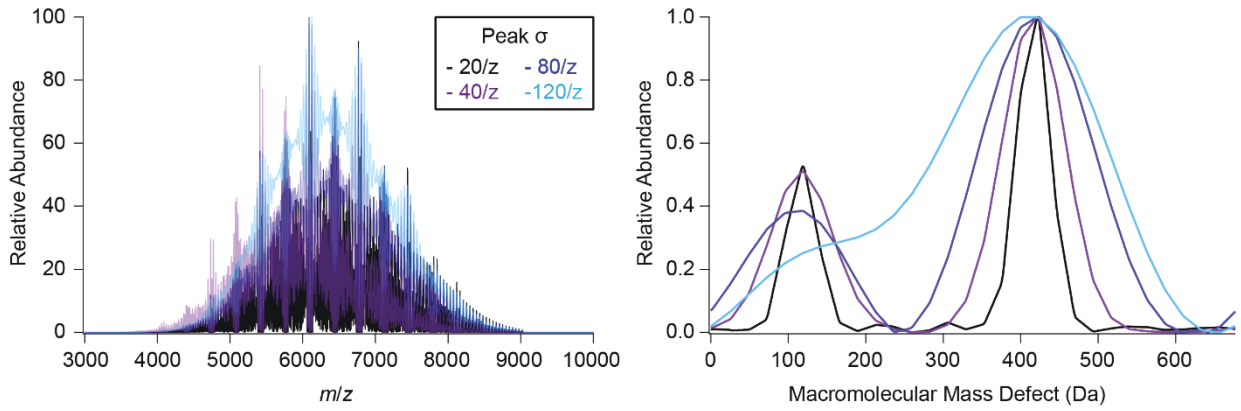


Figure S1 Comparison of simulated spectra and MMD profiles as peak width is varied. Wider peaks in the mass spectra lead to greater superposition, visually appearing as a higher curved baseline. The reduced resolution of the simulated mass spectra is mirrored in their MMD profiles, and inconsistent peak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

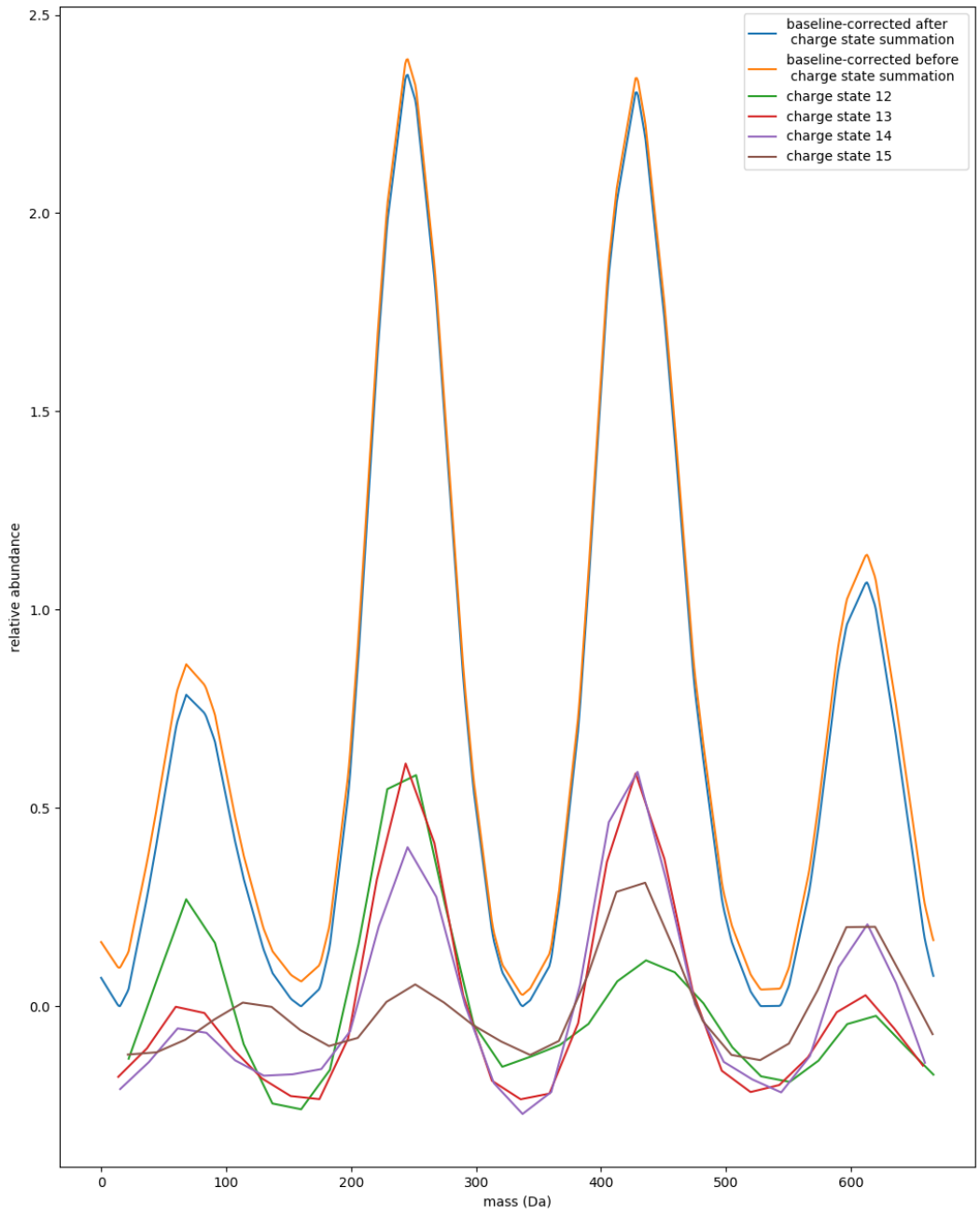


Figure S2 Charge-state-specific (green, red, purple and brown traces) and total MMD profiles (blue and orange traces) for melittin-embedded Nanodiscs prepared at a bulk melittin:Nanodisc ratio of 6:1. Baseline correction was applied to either the individual charge state profiles (orange) or their interpolated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

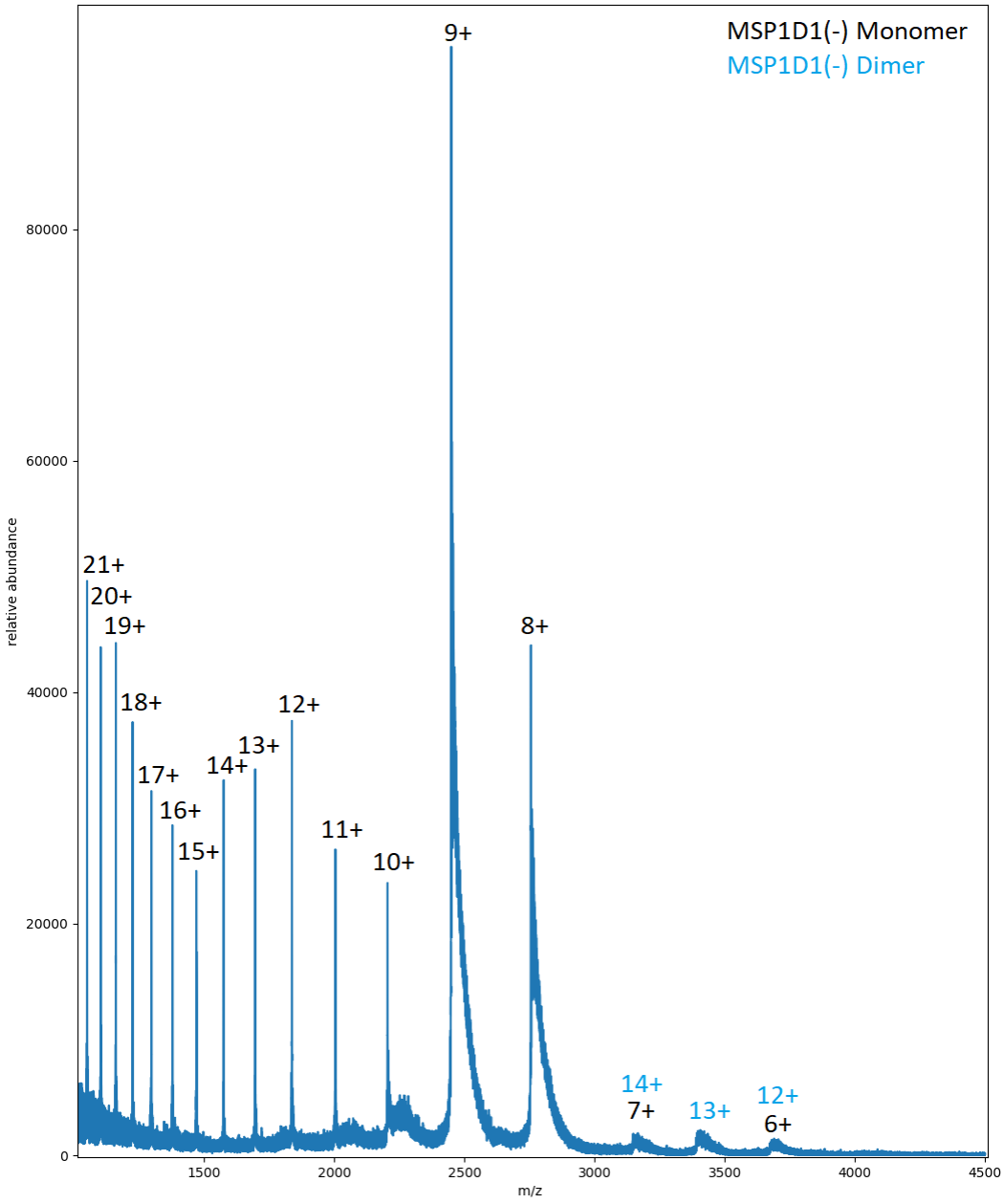


Figure S3 Example native mass spectrum of membrane scaffold protein MSP1D1(−), taken with a trap voltage of 5 V and trap gas flow rate of 3 mL/min, showing both native monomer (black charge states) and dimer (blue charge states) signals. The presence of higher charge states (~11-21+) indicates some partially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

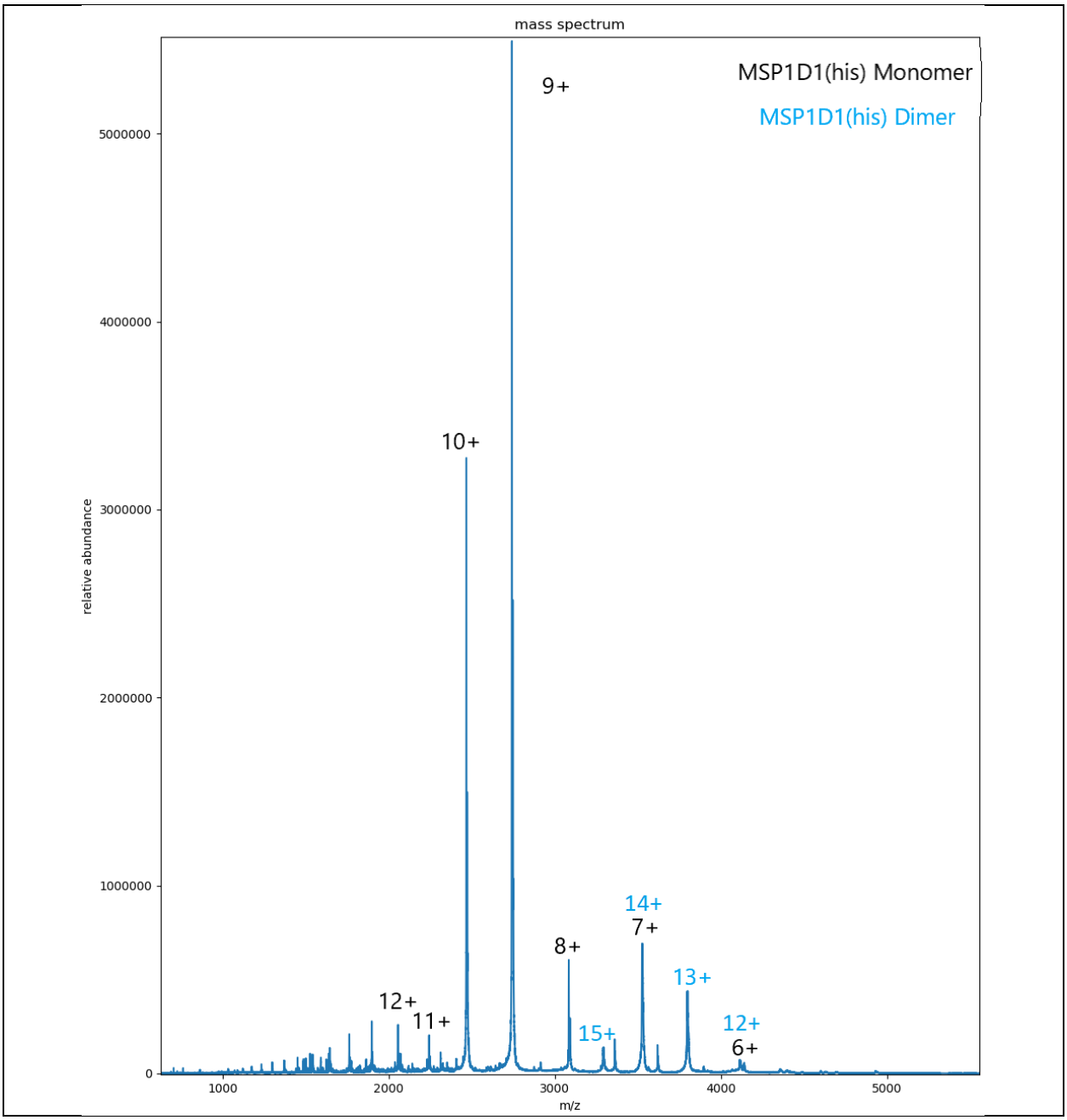


Figure S4 Example native mass spectrum of membrane scaffold protein MSP1D1His, taken with a trap voltage of 25 V and trap gas flow rate of 5 mL/min, showing both native monomer (black charge states) and dimer (blue charge states) signals. The presence of higher charge states (~11+ and above) indicates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