

RESEARCH ARTICLE

Interpretation and Deconvolution of Nanodisc Native Mass Spectra

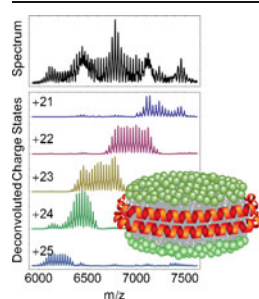
Michael T. Marty,¹ Hao Zhang,^{2,3} Weidong Cui,² Michael L. Gross,² Stephen G. Sligar^{1,4}

¹Department of Chemistry, University of Illinois Urbana-Champaign, Urbana, IL 61801, USA

²Department of Chemistry, Washington University in St. Louis, St. Louis, MO 63130, USA

³Photosynthetic Antenna Research Center (PARC), Washington University in St. Louis, St. Louis, MO 63130, USA

⁴Department of Biochemistry, University of Illinois Urbana-Champaign, Urbana, IL 61801, USA



Abstract. Nanodiscs are a promising system for studying gas-phase and solution complexes of membrane proteins and lipids. We previously demonstrated that native electrospray ionization allows mass spectral analysis of intact Nanodisc complexes at single lipid resolution. This report details an improved theoretical framework for interpreting and deconvoluting native mass spectra of Nanodisc lipoprotein complexes. In addition to the intrinsic lipid count and charge distributions, Nanodisc mass spectra are significantly shaped by constructive overlap of adjacent charge states at integer multiples of the lipid mass. We describe the mathematical basis for this effect and develop a probability-based algorithm to deconvolute the underlying mass and charge distributions. The

probability-based deconvolution algorithm is applied to a series of dimyristoylphosphatidylcholine Nanodisc native mass spectra and used to provide a quantitative picture of the lipid loss in gas-phase fragmentation.

Key words: Mass spectrometry, Nanodisc, Membrane proteins

Received: 18 September 2013/Revised: 27 October 2013/Accepted: 8 November 2013/Published online: 19 December 2013

Introduction

Membrane systems play a critical role in biology by regulating the chemical, energy, and information flow into the cell and its various compartments. Native mass spectrometry (MS), which seeks to preserve solution structures and noncovalent interactions, has emerged as a powerful technique to study membrane systems owing to its low sample requirements and unique structural information on complex structure and lipid binding [1–6]. Nanodiscs are nanoscale discoidal lipid bilayers encircled by two amphipathic membrane scaffold proteins (MSP) [7–9]. Nanodiscs offer a promising technology for native mass spectrometry of membrane systems because they are monodisperse, homogeneous, and possess a native-like lipid bilayer structure [10–12].

We previously demonstrated that intact Nanodiscs can be studied in the gas phase by native electrospray mass spectrometry [13].

The resulting mass spectra are characterized by broad distributions of narrow peaks (see Figure 1a as well as Figures S1 and S2 in the online Supplementary Information for examples). Our initial interpretation was that the broad distributions arose from two factors, the lipid count distribution and charge state distribution. Each narrow peak is due to Nanodiscs with a defined lipid count. We assumed each broader peak results from the lipid count distribution at a single charge state. Because the charge state was determined from the difference between narrow peaks, fitting the broad peaks to Gaussian distributions yielded the mean and standard deviation of the lipid count in the Nanodisc.

Subsequent measurements and theoretical exploration, however, reveal an additional factor contributing to the broad distributions observed. In addition to the lipid count and charge state distributions, the constructive overlap of adjacent charge states may play a dominant role in shaping the spectra. Similar effects have been observed in mass spectra of protein complexes, including amyloid and heat shock protein oligomers [14, 15]. For the Nanodisc system, overlap occurs specifically at m/z values near integer multiples of the lipid mass. Constructive overlap complicates peak assignments and demands a more sophisticated deconvolution of the underlying mass and charge distributions.

Electronic supplementary material The online version of this article (doi:10.1007/s13361-013-0782-y) contains supplementary material, which is available to authorized users.

Correspondence to: Stephen G. Sligar; e-mail: s-sligar@uiuc.edu

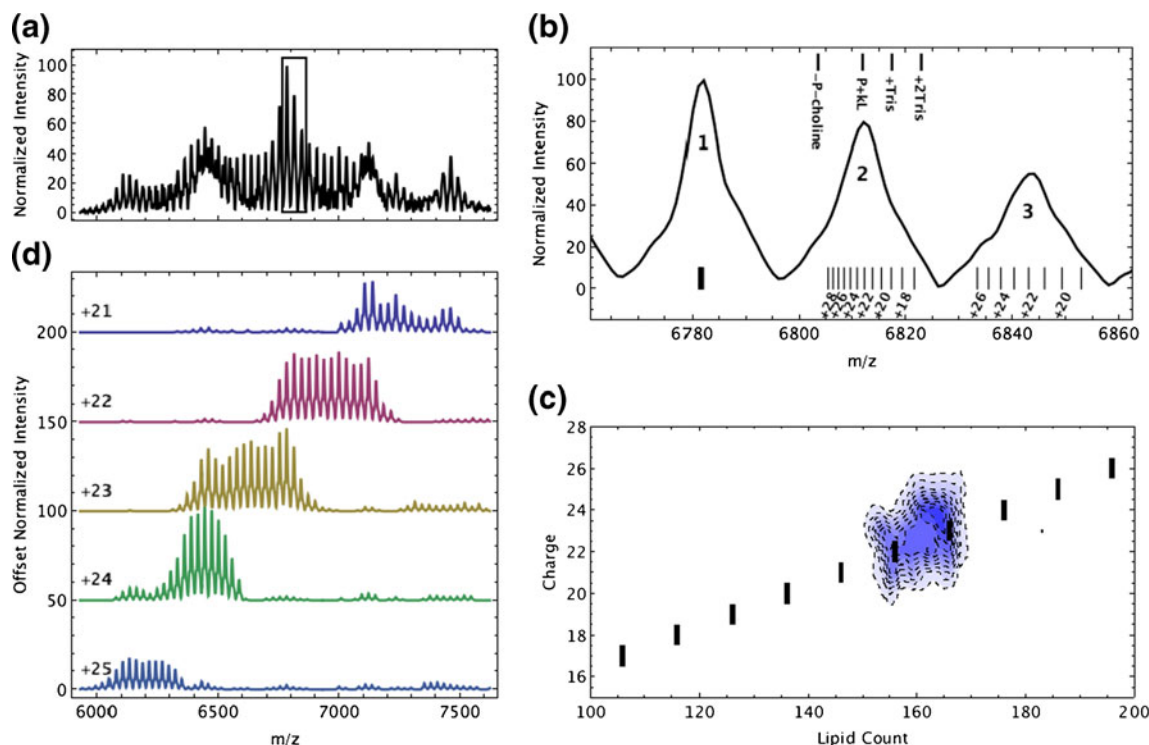


Figure 1. Native mass spectrum **(a)** of DMPC Nanodiscs at 70 V ISCAD. Broad peaks occur at integer multiples of the lipid at $n = 9, 10, \text{ or } 11$, near m/z 6102, 6780, and 7458, respectively, and halfway in between the integer values. An expansion of the boxed region in **(a)** is presented in **(b)**. The position of the nearest Nanodisc ions are marked with vertical lines below the spectrum, and ions with an even charge are annotated. Ions on peak 1 at $10L$ are very closely spaced and, hence, not labeled. Because each peak is not perfectly Gaussian, other possible species are marked above peak 2 for the center peak with the adduction of one or two Tris molecules or the loss of a phosphocholine fragment. The mass and charge values for peak 2 are marked as black boxes in the deconvolution matrix **(c)**. The blue contour plot **(c)** presents the deconvoluted lipid count and charge distributions. Projection of deconvolution into m/z space is shown in **(d)** with offset charge states for the most abundant charges

We addressed this problem with an improved model for interpreting Nanodisc native mass spectra and a probability-based algorithm for deconvolution. The deconvolution algorithm is applied to a representative series of native mass spectra from dimyristoylphosphatidylcholine (DMPC) Nanodiscs fragmented by both in-source collisionally activated dissociation (ISCAD) and infrared multiphoton dissociation (IRMPD).

We anticipate that the theory and algorithms described herein will aid in future studies of Nanodisc complexes containing more complex lipid and membrane protein systems and will facilitate the application of Nanodiscs to the compelling challenges of quantitating and studying membrane proteins. The strength of the algorithm presented herein is that it provides an unbiased deconvolution, which does not rely on a particular model of oligomeric or charge state distribution, while still factoring in the probabilities of neighboring charge and oligomeric states. This influence from neighboring states is crucial to solving the problem of overlapping peaks. As such, the probability-based deconvolution approach will likely find direct application to other systems with overlapping charge state and oligomer

distributions, such as heat shock proteins [14, 16, 17] and amyloid oligomers [15], or for complex spectra with multiple overlapping components, such as fragments of large protein complexes with multiple subunits [6, 18]. We envision that probability-based deconvolution may also be broadly applicable to heterogeneous native mass spectrometry systems such as antibody–antigen complexes [19, 20] and may prove to be useful in the analysis of proteins such as antibodies that contain complex glycosylation patterns [20, 21].

Experimental

Nanodisc Preparation

Nanodiscs were prepared as previously described [7, 9, 13]. Briefly, the lipid, 1,2-dimyristoyl-*sn*-glycero-3-phosphocholine (DMPC), purchased from Avanti Polar Lipids (Alabaster, AL, USA), was solubilized in chloroform, dried under nitrogen, and resuspended in sodium cholate (Sigma Aldrich, St. Louis, MO, USA). Cholate-solubilized DMPC was combined with a membrane scaffold protein

variant, MSP1D1(-), in a 180:90:1 cholate:DMPC:MSP ratio at 25 °C. Cholate was removed by adding Amberlite XAD-2 hydrophobic beads (Sigma Aldrich). Removal of detergent drives Nanodisc self-assembly. Nanodiscs were purified by size exclusion chromatography by using a 0.1 M ammonium acetate (Sigma Aldrich) running buffer at pH = 6.8. The Nanodiscs were collected by pooling the central fractions of the chromatographic peak. The final concentration of Nanodiscs was 10–15 μ M.

Mass Spectrometry

Mass spectrometry was performed on a Bruker Solarix 12 T Fourier transform ion cyclotron resonance (FTICR) mass spectrometer using PicoTip (New Objective, Woburn, MA, USA) capillary needles for nano-electrospray ionization. Instrumental parameters were as previously described [13]. In-source collisionally activated dissociation was performed by using skimmer cone voltages ranging from 10 to 160 V. Infrared multiphoton dissociation was performed by using a 10.6 μ m CO₂ laser (SYNRAD 48-2KAL, Mukilteo, WA, USA) at 25 % full power (32 W) varying the laser pulse duration from 0.1 to 1.5 s. All IRMPD spectra were collected by using the minimum 10 V ISCAD voltage. Several thousand scans were averaged for each spectrum and were collected at the 128 k data size. Data were exported as raw spectra in a table of x/y values and imported directly into Mathematica 8.0.4, where they were linearized and normalized. Data analysis was performed with a custom algorithm (described below) written in Mathematica 8.0.4.

Theory

Overlap of Adjacent Charge States

In addition to the intrinsic lipid count and charge state distribution, Nanodisc native mass spectra are significantly shaped by the constructive overlap of adjacent charge states. This section mathematically demonstrates that Nanodisc ions with an m/z value close to integer multiples of the lipid mass will overlap with Nanodisc ions with a different number of lipids in an adjacent charge state.

Each Nanodisc contains two molecules of MSP and a variable number of DMPC lipids. The mass of the MSP1D1(-) construct, M_{MSP} , is 22,044 Da. Because each Nanodisc contains two copies of MSP, we can define the mass of the protein belt component, B , as $B = 2M_{MSP}$. The mass of the DMPC lipid, L , is 678 Da. Adding the protein and lipid components together gives the mass of a single Nanodisc complex with k_1 lipids as $B + k_1L$. We will refer to this complex as ion 1. For convenience in this derivation, the masses of the protons added in the electrospray process are disregarded because the error introduced by the added proton mass is negligible compared with the overall mass.

Because there is a distribution of lipid count values, we can also consider a separate Nanodisc complex, ion 2, with

k_2 lipids where $k_1 > k_2$. Because k_1 and k_2 are integers, $k_1 - k_2 = n$, where n is also an integer. Assume the first ion has a charge z while the second has a charge of $z - 1$. Using simple arithmetic rearrangement, we can demonstrate that:

$$\frac{B + k_1L}{z} = \frac{B + k_2L}{z - 1} \Leftrightarrow \frac{B + k_1L}{z} = nL \quad (1)$$

In other words, if Nanodiscs of adjacent charge states have the same m/z value, the m/z value is nL , where n is equal to the difference in lipid count and L is the mass of the lipid. The converse of this statement is also true; an m/z value equal to nL implies a potential ion in an adjacent charge state with the same m/z value. An analogous argument shows that Nanodisc ions at charge z will overlap with Nanodiscs of charge $z - 2$ halfway between integer values.

In this idealized case, $\frac{B + k_1L}{z} = nL$ implies that $B = L(zn - k_1)$. Because z , n , and k_1 are all integers, B must be an integer multiple of L , and the entire system simplifies to the principle that $\frac{230L}{23} = \frac{220L}{22} = 10L$. With DMPC and MSP1D1(-), $\frac{B}{L} = 65.03$. The mass of the protein component is very close to an integer multiple of the lipid mass, so there will be nearly perfect overlap of ions 1 and 2.

However, it is not necessary for B to be an exact integer multiple of L . Consider the case where $B = L(zn - k_1) + \varepsilon = L(zn - n - k_2) + \varepsilon$, where ε is some error such that $\varepsilon < L$. Simple rearrangement shows:

$$\frac{B - \varepsilon + k_1L}{z} = \frac{B - \varepsilon + k_2L}{z - 1} \quad (2)$$

$$\frac{B + k_1L}{z} - \frac{\varepsilon}{z} = \frac{B + k_2L}{z - 1} - \frac{\varepsilon}{z - 1} \quad (3)$$

$$\frac{B + k_1L}{z} = \frac{B + k_2L}{z - 1} - \frac{\varepsilon}{z(z - 1)} \quad (4)$$

The difference between the m/z values of ion 1 and ion 2 is $\frac{\varepsilon}{z(z - 1)} < \frac{L}{z(z - 1)}$. The error introduced by the protein component is bounded and scales roughly with the inverse square of the charge state. Thus, constructive overlap can occur for any B or L masses.

This model of constructive overlap between adjacent charge states suggests a strategy to minimize the effect. Lowering the charge shifts the ions to higher m/z values and increases the value of n . At high n , the difference in the number of lipids in ions 1 and 2 (recall that $k_1 - k_2 = n$) may be larger than the intrinsic lipid distribution. In other words,

the lipid distribution is not wide enough that the adjacent charge state contributes significantly to the peak at nL . An example of this phenomenon is discussed below.

Probability-Based Deconvolution of Nanodisc Spectra

Peak assignment and interpretation of Nanodisc native MS spectra are complex and must take into account the overlap of adjacent charge states demonstrated above. To deconvolute the overlapping charge states and determine the underlying lipid count and charge distribution of the Nanodisc ions, we developed a probability-based deconvolution (PDB) algorithm.

The central goal of the algorithm is to deconvolute the one-dimensional mass/charge spectrum into a two-dimensional matrix of mass and charge values. To simplify the problem, mass is quantized by lipid count and limited to a specific range. We assume that mass is equal to $B + kL$, where $k \in \{75, 76, \dots, 225\} = K$. Charge is also quantized and limited to $z \in \{3, 4, \dots, 32\} = Z$. These limitations are empirically set at the beginning of the algorithm and are centered around previously established values for Nanodisc lipid counts and charge [8, 13].

We define the matrix of (k, z) pairs arising from sets K and Z as M . Each element of M has an m/z value defined by the function μ as $\mu(k, z) = \frac{B + kL + z}{z}$ for $(k, z) \in M$. Note that the mass of the ESI protons are now included in the overall mass for the deconvolution algorithm. Other adducts are discussed below. The probability matrix, P , is defined with the same dimensions as M . We will refer to the probability of any element, $(k, z) \in M$, as $P(k, z)$.

In preparation for analysis, the experimental spectrum is linearized and normalized. No other manipulation of the data is required, and no prior peak picking is necessary to apply the algorithm. For clarity, we define the function, $\beta(m/z)$, as the intensity of the experimental spectrum at m/z .

The simplest deconvolution strategy would be to set the probability in P of any given $(k, z) \in M$ as the intensity of the spectrum at that m/z value:

$$P(k, z) \propto \beta(\mu(k, z)) \quad (5)$$

Normalization of matrix, P , corrects for the proportionality and converts P into a true probability distribution. This simple strategy is foiled, however, by the constructive overlap of adjacent charge states because the peak at a given m/z value can be assigned to a number of potential (k, z) pairs.

To correct for the overlap effect, another factor is added to Equation 5:

$$P(k, z) \propto \beta(\mu(k, z))C(k, z) \quad (6)$$

where $C(k, z)$ is the proportion of the spectra intensity, $\mu(k, z)$, which should be assigned to the particular (k, z) pair. $C(k, z)$ is defined by three separate factors:

$$C(k, z) = \frac{\beta(\mu(k, z))N(k, z)}{\sum_{i \in K} \sum_{j \in Z} \beta(\mu(i, j)) \text{dist}((i, j), (k, z))N(i, j)} \quad (7)$$

where β was previously defined as the spectral intensity. The dist term is a distance cutoff defined by a Gaussian distribution centered at $\mu(k, z)$ with a standard deviation of σ_d :

$$\text{dist}((i, j), (k, z)) = e^{-\frac{(\mu(i, j) - \mu(k, z))^2}{2\sigma_d^2}} \quad (8)$$

The value of σ_d defined in the distance cutoff has a significant impact on the deconvolution. Some tuning is required to find the optimal cutoff distance. When σ_d is small, P is noisy with significant background signal. The algorithm overcorrects for the overlap effect and distributes the peak intensity across too many charge states. When σ_d is too large, P is too smooth, and the algorithm does not effectively account for the overlap effect. Too much of the intensity is distributed to the most probable charge state. In general, we found that a σ_d value around 1.5 times the standard deviation of the narrow peaks gave the best fit.

N is a factor designed to capture the neighborhood of each (k, z) pair. The central assumption is that the probability of any given (k, z) pair is proportional to the probability of neighboring pairs, including $(k, z - 1)$, $(k, z + 1)$, and $(k - 1, z)$, for example. We assume that there will be a low probability of any (k, z) pairs showing up in isolation. Implementation of this concept requires an iterative updating of the probability matrix, P . For each iteration, the probability matrix, P^n , is calculated based on the prior probability matrix, P^{n-1} , and then normalized. The first iteration, P^0 , is given a uniform probability, so the probability for overlapping peaks is approximately equally distributed. On each subsequent iteration of the algorithm, the prior probability, P^{n-1} , is blurred by a Gaussian filter. In other words, the probability matrix is convolved with a Gaussian function with standard deviations (σ_1, σ_2) as given by:

$$N(k, z) = \frac{\sum_{x \in K} \sum_{y \in Z} e^{-\frac{(x-k)^2}{2\sigma_1^2}} e^{-\frac{(y-z)^2}{2\sigma_2^2}} P^{n-1}(x, y)}{\sum_{x \in K} \sum_{y \in Z} e^{-\frac{(x-k)^2}{2\sigma_1^2}} e^{-\frac{(y-z)^2}{2\sigma_2^2}}} \quad (9)$$

For this study, $\sigma_1 = 2$ and $\sigma_2 = 1$ were used as values for the standard deviation. In general, larger σ_1 and σ_2 values will yield a smoother fit that is more dependent on the neighborhood, whereas smaller values will allow more local variation. We recommend some optimization of these

parameters to achieve the best fit. However, the algorithm is relatively insensitive to the precise values of σ_1 and σ_2 or the shape of the blurring filter. For example, taking the arithmetic mean of the probabilities of all complexes plus or minus one lipid and plus or minus one charge gives similar results to the Gaussian filter described in Equation 9. In any case, the algorithm typically converges to a final solution within about eight iterations.

This study utilized a uniform initial probability matrix, P^0 , to avoid biasing the deconvolution. It is possible, however, to consider using a nonuniform P^0 in the case where it is justified by prior data or knowledge. To evaluate this approach, several different biased P^0 matrices were evaluated. Distributions were chosen that were both similar and dissimilar, which we will refer to as correctly and incorrectly biased, to the final probability matrices determined from the uniform initial probability matrix. In general, the correctly biased initial matrices converged quickly to nearly identical distributions and fits. The incorrectly biased initial matrices took longer but eventually converged to similar distributions with similar but slightly poorer fits. Thus, the algorithm was fairly robust with respect to the use of nonuniform P^0 matrices but appears to behave best with a uniform starting distribution.

Following determination of P , the quality of the fit is determined by projecting the probability matrix back into a simulated mass spectrum in m/z space. This is accomplished by summing Gaussian distributions centered at each m/z value in M with the intensity of each determined from P . Some care must be taken to find the appropriate width of the distributions and to account for any adducts or fragments (see below). The relative populations of adduct species and an initial guess for peak widths are determined by fitting only the overlap peaks at integer multiples of the lipid mass. After determining the final P , the peak widths are optimized for the best fit. The sum of squared errors (SSE) may then be calculated between the experimental spectrum and the simulated spectrum determined from the deconvolution.

Results and Discussion

To illustrate the effect of overlapping charge states on Nanodisc native mass spectra and probability-based deconvolution, we first focus on a single mass spectrum of DMPC Nanodisc ions subjected to 70 V ISCAD. As shown in Figure 1a, the spectrum contains a number of sharp narrow peaks and five larger broad peaks. Three of the broad peaks occur at integer multiples of the lipid at $n = 9, 10,$ or 11 , near m/z 6102, 6780, and 7458, respectively. Two other broad peaks are found half way between these integer multiples. These are the locations predicted by the charge state overlap theory described above.

Zooming in to peak 1 at 10L (approximately 6780 Da) and its neighbors, it is clear that the overlap effect precludes assignment of the peak at m/z 6780. Potential Nanodisc ions of various charge states are marked with vertical lines below

the peaks in Figure 1b. For peak 1, the potential ions are very closely spaced (too close to label clearly). Even for peak 2, which does not overlap perfectly, the difference between many of the possible ions is small. The position of some possible m/z pairs for peak 2 are marked with black boxes in the mass/charge matrix in Figure 1c to illustrate the wide range of charge states and lipid counts that could contribute to peak 2. On the other hand, peak 3 is easier to attribute primarily to the +22 charge state with a minor contribution from the +23 charge state.

One notable feature of the spectrum is that the three peaks are not perfectly Gaussian in shape but show a similar pattern. Because all Nanodisc ions that contain integer values of lipid, protein, and charge cluster closely at peak 1, it is impossible to attribute some of these shoulders to Nanodisc ions containing only protein and lipid. One possible assignment of these peaks is to various adducts and fragments. The shoulder at lower m/z can be attributed to loss of a phosphocholine fragment of 184 Da. This loss of phosphocholine from phosphatidylcholine lipids is a well-known fragmentation reaction [22, 23].

The shoulders at higher m/z are harder to assign. They are likely due to heterogeneous adduction of water or ions from solution. The largest potential adduct is one containing Tris of 121 Da. Although we attempted to remove all Tris with size exclusion chromatography, a small amount could remain bound to the complex [24, 25]. For the deconvolution algorithm, peak 1 along with peaks near 9L and 11L were fit to four overlapping Gaussian distributions, the pure protein/lipid complex, the ion formed by loss of phosphocholine, and adducts with one or two Tris molecules, to determine the relative population of each of these species and an appropriate peak width. The peak widths are used to determine the cutoff distance, σ_d , used in Equation 8. It is possible to include adduct peaks in the probability-based deconvolution by adding the spectral intensity of the adducts to the spectral intensity of the bare ions in the β term, but we found that this addition did not improve the algorithm significantly for this system.

With an appropriate distance cutoff, we can now consider each factor contributing to P in the context of peaks 1, 2, and 3. For peak 1, the $dist$ values will all be close to unity because the possible m/z values are very close to each other. The spectral intensity factor, β , will also be very similar for all (m, z) pairs. Thus, the probability of each possible (m, z) pair close to peak 1 is primarily determined by the neighborhood factor, N . The N term is initially uniform for each possible value but converges to the final solution. In the case of peak 1, N converges to similar values for the +22 and +23 charge states and is nearly zero for all others. Figure 1c shows the final distribution and illustrates why the neighborhood factor would be small for all other charge states in light of the final distribution.

For peaks 2 and 3, N behaves similarly to peak 1. The spreading of charge states, however, causes differences in the β and $dist$ factors. The β term is highest for the +22 and

+23 charge states, especially in peak 3. The *dist* factor will limit the intensity from being assigned to charge states that are far away from the central m/z . Thus, even when the neighborhood factor is uniform in the first iteration of the algorithm, the +22 and +23 charge states have the highest probabilities. As shown in Figure 1d, the final solution assigns probabilities to both the +22 and +23 charge state for all three peaks.

Considering the whole spectrum in Figure 1a and d, we see that the deconvolution algorithm reveals the overlap of adjacent charge states contributing to the broader peaks. The probability matrix, Figure 1c, shows a centralized distribution of mass and charge. Summing the columns of the matrix yields the overall lipid distribution. For simplification, the lipid count distribution may then be fit to a Gaussian to determine the lipid count mean and standard deviation.

The same deconvolution algorithm was applied to mass spectra of Nanodiscs where increasing levels of fragmentation occur. The deconvolution matrices for DMPC Nanodiscs at various fragmentation energies with both ISCAD and IRMPD are given in Figure 2a and b, respectively. The overall lipid distributions from these matrices are shown in Figure 2c and d. Plotting the lipid count mean and standard deviation as a function of ISCAD voltage or IRMPD laser duration (Figure 3) provides a quantitative picture of Nanodisc fragmentation. The experimental data and final fits for each spectrum are shown in Supplementary Figures S1 and S2.

From these data, we observe that the lipid count is slightly higher and the distribution is slightly broader than previously reported [13]. The mean lipid count is 165 at the 10 V ISCAD, whereas the value that we previously reported is 155 [13]. The higher lipid count may be the result of improved data analysis and higher resolution spectra or could be due to sample-to-sample variation. It is also possible although unlikely that association with free lipid molecules could give rise to elevated lipid counts. The lipid count standard deviation varies from ± 4 at low fragmentation energies to ± 7 at the highest. Although there is some background outside of these central distributions, especially at higher fragmentation energies, these data suggest that Nanodiscs undergo a rather well defined fragmentation pathway, staying relatively tightly grouped as they lose lipids.

One interesting feature observed in some of the IRMPD spectra is that the constructive overlap of adjacent charge states is relatively minor (Supplementary Figure S2). This is due to charge reduction caused by the IRMPD fragmentation. At lower charge, Nanodisc ions shift to higher m/z , and the regions of potential overlap shift to higher nL values. Because $n = k_1 - k_2$, where k_1 and k_2 are the overlapping lipid counts, overlap is more likely at lower values of n . For peak 1 at 10L, lipid counts that are 10 apart have the potential to overlap. Because the lipid count distribution is ± 5 at 70 V ISCAD, overlap is likely. However, at 22L (the highest peak in the 0.9 s IRMPD

spectrum, see Supplementary Figure S2), lipid counts must be 22 lipids apart to overlap. Because the central distribution is ± 6 for this spectrum, overlap is unlikely. The weak overlap that does occur can be attributed to the presence of background peaks outside of the central distribution.

A major advantage of the probability-based deconvolution algorithm is that it does not assume a given model, such as a Gaussian distribution. Although the algorithm considers the neighborhood, the probability of each (m, z) pair is determined individually. This raises the question of whether the number of independent variables included in the probability-based deconvolution is justified. Using the K and Z sets defined above, the probability matrix contains 4530 elements. Removal of all elements of P with probabilities smaller than 1 % of the maximum probability in P , however, reduces the number of variables in the model by a half to full order of magnitude without significantly reducing the quality of the fit.

We used a Levenberg-Marquardt algorithm to fit other models for the lipid and charge distributions, including Gaussian, Cauchy, and skewed distributions. The best of these was a Cauchy distribution in the lipid count and a Gaussian distribution in charge, based on a modified square root relationship to its mass [17]. None of these distributions, however, fit the spectra well. F-tests comparing the reduced probability-based model with the Cauchy distribution model revealed that the probability-based model was significantly better than the simpler models (additional details are provided in the Supporting Information). Although improved models and fitting strategies may emerge to describe the Nanodisc charge and lipid count distributions, the probability-based deconvolution approach provides a useful flexibility for initial studies such as these where the distributions are not well characterized.

A number of different algorithms have been developed for deconvoluting electrospray mass spectra [26–32]. Although a detailed comparison is beyond the scope of this report, our probability-based deconvolution (PBD) has conceptual and algorithmic differences from the entropy-based methods such as MaxEnt [26, 27] and the algorithm developed by Reinhold and Reinhold [30]. The biggest difference is conceptual. Maximum entropy methods formally recognize that a mass spectrum is just one realization of an ensemble of spectra that could have been obtained. The realized spectrum is not an exact copy of the “true” spectrum, but is different owing to the uncertainty of measurement. A probability model of the ensemble is explicitly incorporated into maximum entropy methods, but there is no a priori assumption about which mass-to-charge values in the spectrum are important. Our PBD method, on the other hand, uses probability to distribute spectral intensity into idealized distributions of lipid count and charge. Uncertainty in the experimental spectrum is transformed into uncertainty in the distributions. As such, there is a very strong a priori assumption about which mass-to-charge values are important in the spectrum.

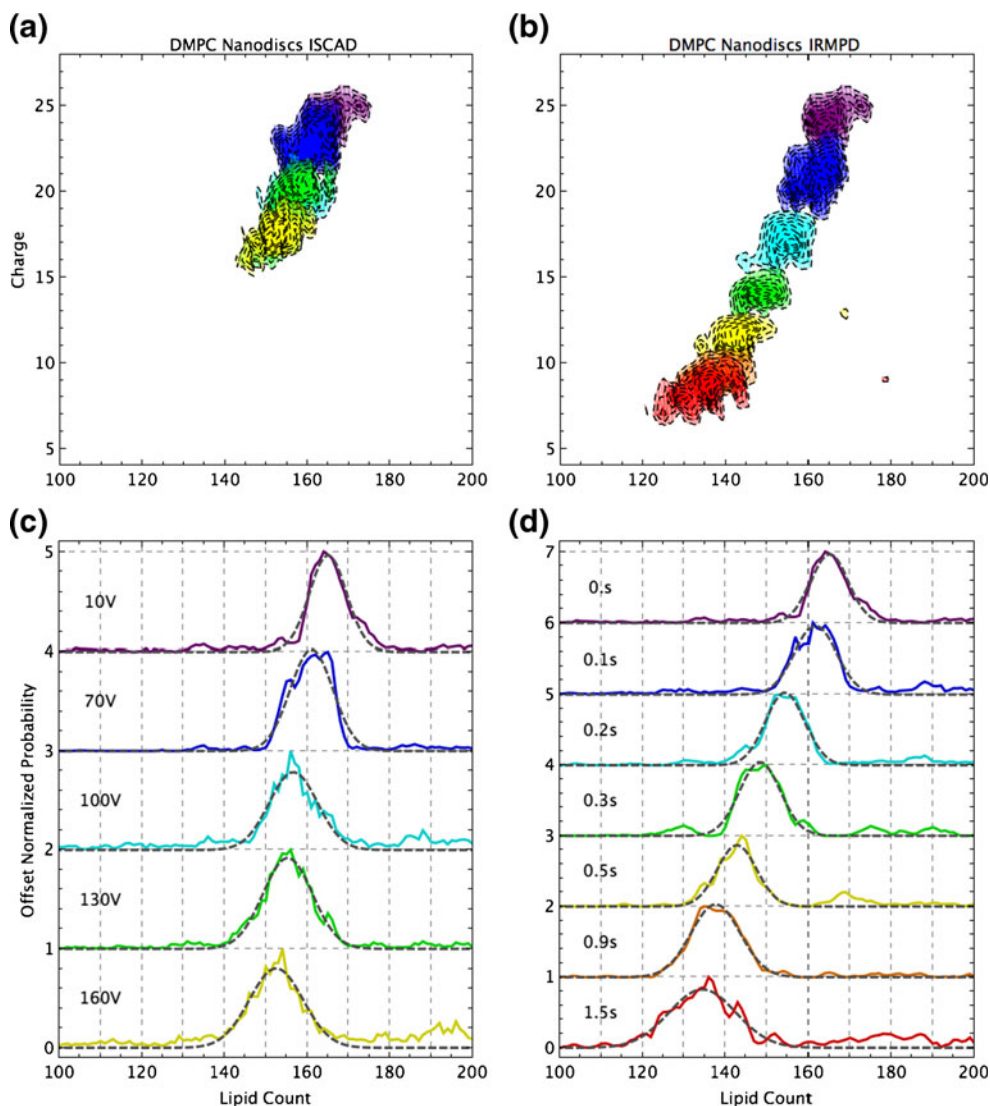


Figure 2. Deconvolution of lipid count and charge for DMPC Nanodiscs at a variety of fragmentation energies from both ISCAD **(a)** and IRMPD **(b)**. Summation of the columns from these matrices gives the lipid count distributions, which are plotted in **(c)** and **(d)**. Regions from the top contour plots correspond with the same color that is labeled in the bottom distributions. Fits of each lipid count distribution to a Gaussian distribution are shown as black dashed traces in **(c)** and **(d)**

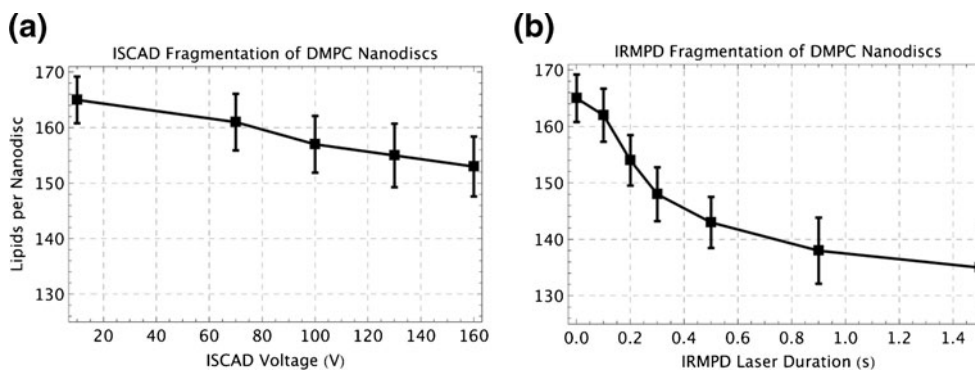


Figure 3. Lipid count distribution for DMPC Nanodiscs as a function of ISCAD voltage **(a)** and IRMPD laser duration **(b)**. Error bars are shown at 1 standard deviation. Mean and standard deviation are taken directly from the fits in Figure 2c and d

Algorithmically, MaxEnt methods ultimately utilize a search process to find the representative spectrum that produces an extreme value for some information theory based evaluation (entropy, for example). Our method utilizes a mapping (constructed in two parts) that is recursively applied to its own result. The mapping makes a correction that is designed to remove the distortion and to produce a better approximation of the idealized distributions or lipid count and charge. The mapping has a fixed point (the idealized distributions) to which the recursive sequence converges.

Conclusion

We describe here the theoretical basis for the constructive overlap of adjacent charge states in Nanodiscs and propose a probability-based algorithm to deconvolute these overlapping distributions. As demonstrated with DMPC Nanodisc spectra at a range of fragmentation energies, charge state overlap plays a significant role in shaping the spectra. The probability-based deconvolution algorithm provides an effective strategy for determining the lipid count and charge states distributions. The theoretical work and algorithms developed in this report will inform future studies of Nanodiscs with membrane proteins and varying lipid populations. Additionally, although our PBD method is not a general method, the strategy could prove useful for the study of a range of other protein complexes or synthetic and biological polymers.

Acknowledgments

The research was supported by the Photosynthetic Antenna Research Center (PARC), an Energy Frontier Research Center funded by the US Department of Energy, Office of Basic Energy Science (grant no. DE-SC 0001035 to Robert E. Blankenship), National Institute of General Medical Sciences (grant no. 8 P41 GM103422-35 to M.L.G.), National Institute of Health (grant no. R01-GM31756 and R01-GM33775 to S.G.S.), and the NSF (IBDR 0964199 to M.L.G.). M.T.M. was supported by the Robert C. and Carolyn J. Springborn Graduate Fellowship. H.Z. was funded equally by the DOE grant to Professor Robert E. Blankenship and NIGMS grant to M.L.G., and instrumentation was made available from both the DOE and NIGMS grants. The authors thank Don Rempel for helpful discussions.

References

- Barrera, N.P., Zhou, M., Robinson, C.V.: The role of lipids in defining membrane protein interactions: insights from mass spectrometry. *Trends Cell Biol.* **23**(1), 1–8 (2013)
- Barrera, N.P., Di Bartolo, N., Booth, P.J., Robinson, C.V.: Micelles protect membrane complexes from solution to vacuum. *Science* **321**(5886), 243–246 (2008)
- Barrera, N.P., Isaacson, S.C., Zhou, M., Bavro, V.N., Welch, A., Schaedler, T.A., Seeger, M.A., Miguel, R.N., Korkhov, V.M., van Veen, H.W., Venter, H., Walmsley, A.R., Tate, C.G., Robinson, C.V.: Mass spectrometry of membrane transporters reveals subunit stoichiometry and interactions. *Nat. Methods* **6**(8), 585–587 (2009)
- Barrera, N.P., Robinson, C.V.: Advances in the mass spectrometry of membrane proteins: from individual proteins to intact complexes. *Annu. Rev. Biochem.* **80**(1), 247–271 (2011)
- Hilton, G.R., Benesch, J.L.P.: Two decades of studying non-covalent biomolecular assemblies by means of electrospray ionization mass spectrometry. *J. Royal Soc. Interface* **9**(70), 801–816 (2012)
- Zhou, M., Morgner, N., Barrera, N.P., Politis, A., Isaacson, S.C., Matak-Vinkovic, D., Murata, T., Bernal, R.A., Stock, D., Robinson, C.V.: Mass spectrometry of intact V-type ATPases reveals bound lipids and the effects of nucleotide binding. *Science* **334**(6054), 380–385 (2011)
- Bayburt, T.H., Grinkova, Y.V., Sligar, S.G.: Self-assembly of discoidal phospholipid bilayer nanoparticles with membrane scaffold proteins. *Nano Lett.* **2**(8), 853–856 (2002)
- Bayburt, T.H., Sligar, S.G.: Membrane protein assembly into Nanodiscs. *FEBS Lett.* **584**(9), 1721–1727 (2010)
- Denisov, I.G., Grinkova, Y.V., Lazarides, A.A., Sligar, S.G.: Directed self-assembly of monodisperse phospholipid bilayer Nanodiscs with controlled size. *J. Am. Chem. Soc.* **126**(11), 3477–3487 (2004)
- Zhang, Y., Liu, L., Daneshfar, R., Kitova, E.N., Li, C., Jia, F., Cairo, C.W., Klassen, J.S.: Protein–glycosphingolipid interactions revealed using catch-and-release mass spectrometry. *Anal. Chem.* **84**(18), 7618–7621 (2012)
- Hopper J.T., Yu Y.T., Li D., Raymond A., Bostock M., Liko I., Mikhailov V., Laganowsky A., Benesch J.L., Caffrey M., Nietlispach D., Robinson C.V.: Detergent-free mass spectrometry of membrane protein complexes. *Nat. Methods.* **10**(12), 1206–1208 (2013)
- Zhang, Y., Deng, L., Kitova, E.N., Klassen, J.S.: Dissociation of multisubunit protein–ligand complexes in the gas phase. Evidence for ligand migration. *J. Am. Soc. Mass Spectrom.* **24**(10), 1573–1583 (2013)
- Marty, M.T., Zhang, H., Cui, W., Blankenship, R.E., Gross, M.L., Sligar, S.G.: Native mass spectrometry characterization of intact Nanodisc lipoprotein complexes. *Anal. Chem.* **84**(21), 8957–8960 (2012)
- Aquilina, J.A., Benesch, J.L., Bateman, O.A., Slingsby, C., Robinson, C.V.: Polydispersity of a mammalian chaperone: mass spectrometry reveals the population of oligomers in alphaB-crystallin. *Proc. Natl. Acad. Sci.* **100**(19), 10611–10616 (2003)
- Smith, A.M., Jahn, T.R., Ashcroft, A.E., Radford, S.E.: Direct observation of oligomeric species formed in the early stages of amyloid fibril formation using electrospray ionisation mass spectrometry. *J. Mol. Biol.* **364**(1), 9–19 (2006)
- Baldwin, A.J., Lioe, H., Robinson, C.V., Kay, L.E., Benesch, J.L.: alphaB-crystallin polydispersity is a consequence of unbiased quaternary dynamics. *J. Mol. Biol.* **413**(2), 297–309 (2011)
- Stengel, F., Baldwin, A.J., Bush, M.F., Hilton, G.R., Lioe, H., Basha, E., Jaya, N., Vierling, E., Benesch, J.L.: Dissecting heterogeneous molecular chaperone complexes using a mass spectrum deconvolution approach. *Chem. Biol.* **19**(5), 599–607 (2012)
- Morgner, N., Robinson, C.V.: Massign: An assignment strategy for maximizing information from the mass spectra of heterogeneous protein assemblies. *Anal. Chem.* **84**(6), 2939–2948 (2012)
- Atmanene C., Wagner-Roussel E., Corvaia N., Dorselaer A., Beck A., Sanglier-Cianferani S.: Noncovalent mass spectrometry for the characterization of antibody/antigen complexes. In: Beck A (Ed.) *Glycosylation Engineering of Biopharmaceuticals*, vol 988. Methods Mol. Biol. Humana Press, pp. 243–268 (2013)
- Thompson, N.J., Rosati, S., Rose, R.J., Heck, A.J.: The impact of mass spectrometry on the study of intact antibodies: from post-translational modifications to structural analysis. *Chem. Commun.* **49**(6), 538–548 (2013)
- Rosati, S., Rose, R.J., Thompson, N.J., van Duijn, E., Damoc, E., Denisov, E., Makarov, A., Heck, A.J.: Exploring an Orbitrap analyzer for the characterization of intact antibodies by native mass spectrometry. *Angew. Chem. Int. Ed.* **51**(52), 12992–12996 (2012)
- Hsu, F.-F., Turk, J.: Electrospray ionization/tandem quadrupole mass spectrometric studies on phosphatidylcholines: the fragmentation processes. *J. Am. Soc. Mass Spectrom.* **14**(4), 352–363 (2003)
- Murphy, R.C., Harrison, K.A.: Fast atom bombardment mass spectrometry of phospholipids. *Mass Spectrom. Rev.* **13**(1), 57–75 (1994)
- Freeke, J., Robinson, C.V., Ruotolo, B.T.: Residual counter ions can stabilise a large protein complex in the gas phase. *Int. J. Mass Spectrom.* **298**(1/3), 91–98 (2010)

25. Hernandez, H., Robinson, C.V.: Determining the stoichiometry and interactions of macromolecular assemblies from mass spectrometry. *Nat. Protoc.* **2**(3), 715–726 (2007)
26. Ferrige, A.G., Seddon, M.J., Green, B.N., Jarvis, S.A., Skilling, J.: Disentangling electrospray spectra with maximum-entropy. *Rapid Commun. Mass Spectrom.* **6**(11), 707–711 (1992)
27. Ferrige, A.G., Seddon, M.J., Jarvis, S.: Maximum-entropy deconvolution in electrospray mass spectrometry. *Rapid Commun. Mass Spectrom.* **5**(8), 374–377 (1991)
28. Hagen, J.J., Monnig, C.A.: Method for estimating molecular-mass from electrospray spectra. *Anal. Chem.* **66**(11), 1877–1883 (1994)
29. Mann, M., Meng, C.K., Fenn, J.B.: Interpreting mass-spectra of multiply charged ions. *Anal. Chem.* **61**(15), 1702–1708 (1989)
30. Reinhold, B.B., Reinhold, V.N.: Electrospray ionization mass spectrometry—deconvolution by an entropy-based algorithm. *J. Am. Soc. Mass Spectrom.* **3**(3), 207–215 (1992)
31. Zhang, Z., 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.* **9**(3), 225–233 (1998)
32. Zheng, H., Ojha, P.C., McClean, S., Black, N.D., Hughes, J.G., Shaw, C.: Heuristic charge assignment for deconvolution of electrospray ionization mass spectra. *Rapid Commun. Mass Spectrom.* **17**(5), 429–436 (2003)