Electronic Supplementary Material (ESI) for Analyst.

Supplementary Information for:

Distinct Classes of Multi-Subunit Heterogeneity: Analysis using Fourier Transform Methods and Native Mass Spectrometry

Sean P. Cleary[1], James S. Prell[1,2]*

1. *Department of Chemistry and Biochemistry, 1253 University of Oregon, Eugene, OR 97403-1253, USA*
2. *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 of bulk-mole-fraction-weighted Fourier frequency for two-subunit Class II ion populations satisfying the mean-proportional-variance condition.

Lemma: The product of two Gaussian functions is another Gaussian function with a mean equal to the variance-weighted average of the means of the two Gaussians.

Proof: Let $f(k) = I_A \exp(-\frac{\left(k - \frac{z}{m_A}\right)^2}{2\sigma_A^2})$ and $g(k) = I_B \exp(-\frac{\left(k - \frac{z}{m_B}\right)^2}{2\sigma_B^2})$, which represent the frequency-domain peak shapes corresponding two Gaussian mass spectral peak distributions with charge state Z, repeated subunit masses m_A and m_B, and mass spectral-domain variances $\frac{1}{\sigma_A}$ and $\frac{1}{\sigma_B}$, respectively. I_A and I_B are the peak intensities of $f(k)$ and $g(k)$ in the frequency domain, respectively. Let $h(k) = f(k) \times g(k)$.

Then $h(k) = I_A I_B \exp\left(-\frac{\left(k - \frac{Z}{m_A}\right)^2}{2\sigma_A^2} - \frac{\left(k - \frac{Z}{m_B}\right)^2}{2\sigma_B^2}\right) = I_A I_B \exp\left(-\frac{\sigma_B^2\left(k - \frac{Z}{m_A}\right)^2 + \sigma_A^2\left(k - \frac{Z}{m_B}\right)^2}{2\sigma_A^2\sigma_B^2}\right)$

Completing the square inside the exponent yields:

$$h(k) = I_A I_B \exp\left(-\left(\sigma_{tot} k - \frac{\sigma_B^2 Z/m_A + \sigma_A^2 Z/m_B}{\sigma_{tot}}\right)^2 / 2\sigma_A^2\sigma_B^2\right) \times I_C$$

where $I_C = \exp\left(-\left(\frac{\sigma_B^2 Z^2}{m_A^2} + \frac{\sigma_A^2 Z^2}{m_B^2} - \frac{\sigma_B^4 Z^2/m_A^2 + 2\sigma_A^2\sigma_B^2 Z^2/m_A m_B + \sigma_A^4 Z^2/m_B^2}{\sigma_{tot}^2}\right) / 2\sigma_A^2\sigma_B^2\right)$ and $\sigma_{tot}^2 = \sigma_A^2 + \sigma_B^2$.

Thus, $h(k)$ is again a Gaussian function of k with mean $\overline{h(k)} = \frac{\sigma_B^2 Z/m_A + \sigma_A^2 Z/m_B}{\sigma_{tot}^2}$, QED.

With the result of this Lemma in hand, we note that $\overline{h(k)}$ is the fundamental frequency that is observed in the Fourier domain upon FT of a mass spectrum that is the convolution of two underlying *m*/*z* distributions with variances $\frac{1}{\sigma_A^2}$ and $\frac{1}{\sigma_B^2}$. **Note that $\overline{h(k)}$ is independent of the *means* of the two underlying *m*/*z* distributions for A and B and therefore does *not* directly relate to the average number of either subunit type in the total ion population.**

Now let $\overline{n_A}$ and $\overline{n_B}$ be the abundance-weighted average number of subunit A and B in the respective underlying mass distributions. If we assume the variances in the number of A and B in the underlying mass distributions are each proportional to their means, i.e., $\frac{1}{\sigma_A^2} = c\overline{n_A}$ and $\frac{1}{\sigma_B^2} = c\overline{n_B}$ for some common constant c (hereafter, the "mean-proportional-variance" condition), then we have

$$\overline{h(k)} = \frac{Z/c\overline{n_B}m_A + Z/c\overline{n_A}m_B}{\sigma_{tot}^2} = \frac{\overline{n_A}Z/m_A + \overline{n_B}Z/m_B}{\overline{n_A} + \overline{n_B}} = \overline{x_A}\,Z/m_A + \overline{x_B}\,Z/m_B$$

where $\overline{x_A}$ and $\overline{x_B}$ are the mole fractions of the corresponding lipids averaged over the entire ion population. This results justifies the approach taken in this manuscript for mixed-lipid Nanodiscs and block copolymers as well as that used by Marty and coworkers,[1] crucially assuming the mean-proportional-variance condition. In other words, under the mean-proportional-variance condition, the mean of $\overline{h(k)}$ is the frequency corresponding to the mole-fraction-weighted average of the mean frequencies of $f(k)$ and $g(k)$.

Through some slightly tedious algebra, I_C, a term that tends to dampen the intensity of $h(k)$, can be rearranged to:

$$I_C = \exp\left(-\frac{Z^2(m_A - m_B)^2}{2m_A^2 m_B^2(\sigma_A^2 + \sigma_B^2)}\right) = \exp\left(-\frac{(\frac{Z}{m_B} - \frac{Z}{m_A})^2}{2\sigma_{tot}^2}\right)$$

Thus I_C, and hence the intensity of the peak in $h(k)$, tends to decrease **1)** as the difference masses of subunits A and B (hence also the difference in the means of $f(k)$ and $g(k)$) increases, and **2)** as $f(k)$ and $g(k)$ become narrower (i.e., the underlying mass distributions for A and B become wider). **These two factors are thus very important in determining whether $h(k)$ will be observable above any noise present in the frequency domain; higher resolution of $f(k)$ and $g(k)$ leads to lower signal for $h(k)$.** See, for illustration, Fig. 4 in the main text and Fig. S1B, below. High resolution of $f(k)$ and $g(k)$ occurs when the underlying mass distributions for A and B are relatively broad and/or the masses of A and B are very different.

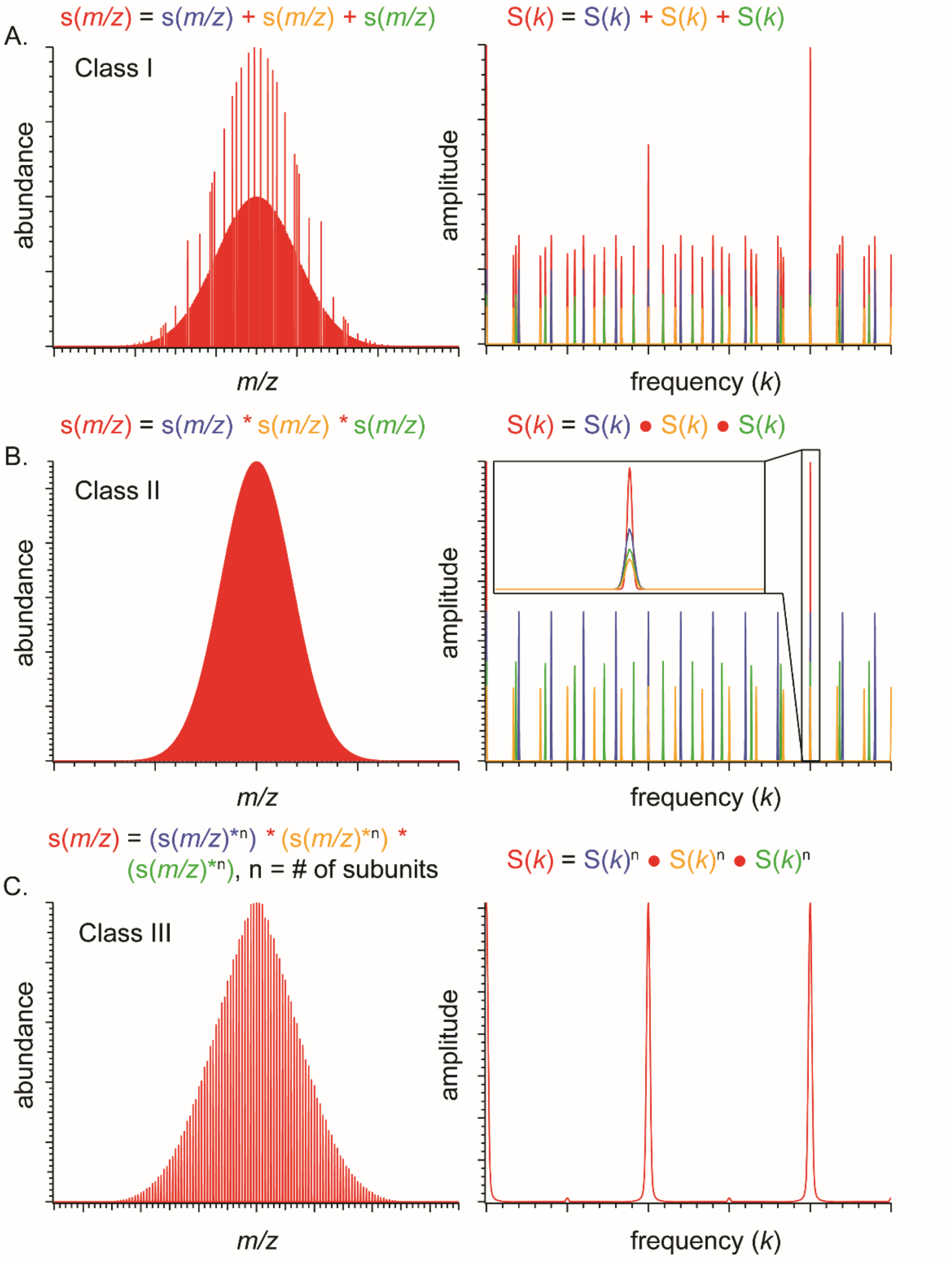

Figure S1. Modeled mass spectra for a single charge state representing different classes of multi-subunit polydispersity for ion populations composed of three different subunit types in 1:1:1 bulk mole fractions and relative masses 10:11:12 (Class I, top; Class II, middle; and Class III, bottom). The mass spectra were built with polydispersity from three different subunits, whose underlying distributions are represented by the blue, orange and green traces with underlying *m*/*z* distributions s(*m*/*z*) labeled by the same color. • denotes multiplication, * denotes convolution, and superscript **n* for Class III denotes *n*-fold autoconvolution.

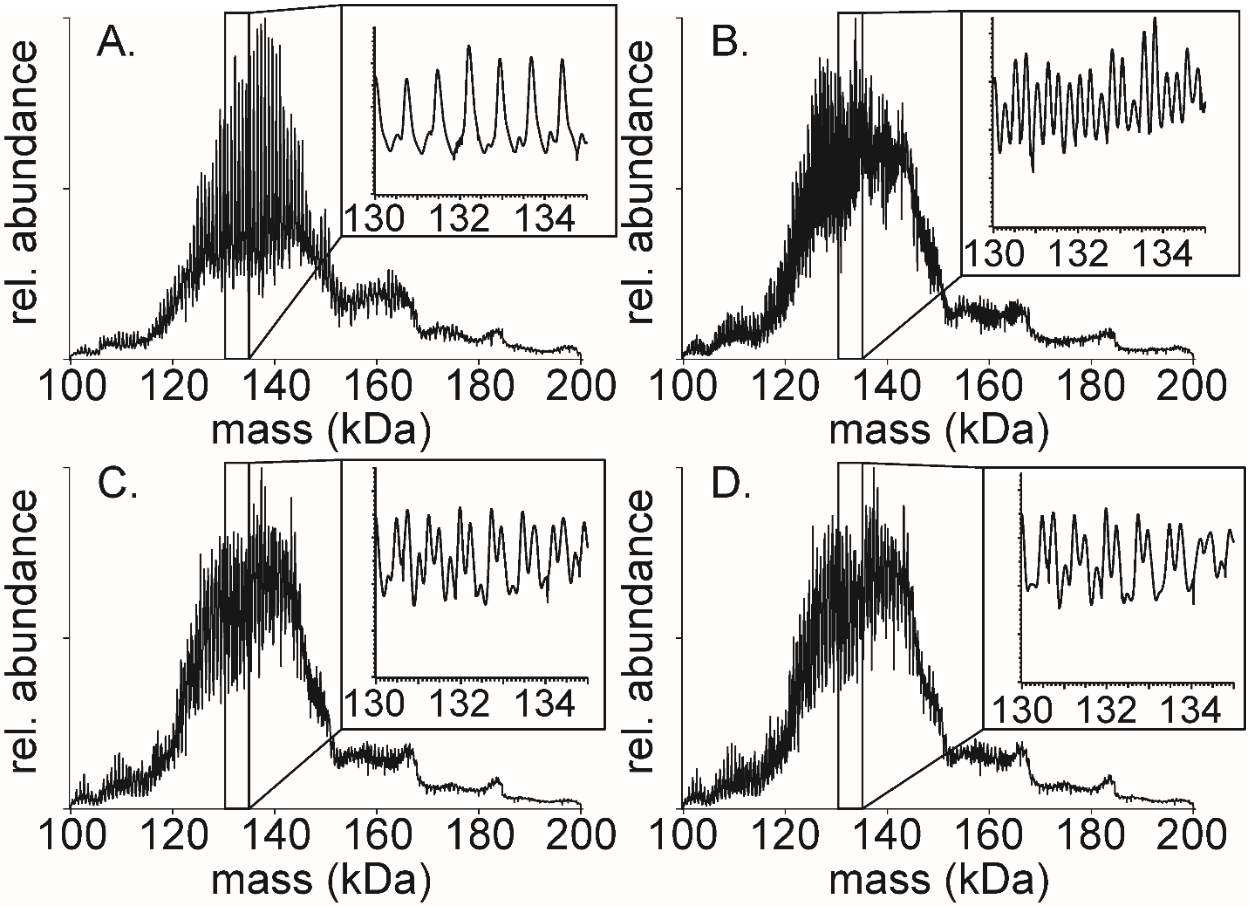

Figure S2. Zero-charge deconvolutions of Nanodisc mass spectra using UniDec v. 4.2.1 corresponding to the data in Figure 2A (A), Figure 2C (B), and Figure 2E (C and D) of the main manuscript. The UniDec Nanodisc preset was used with parameters adjusted to match the experimental data (mass range 100,000-200,000 Da, charge state range 8-14, and an appropriate comb filter for the subunit mass). Because the ion population in Figure 2E of the main manuscript contains two different lipid types, the corresponding UniDec deconvolutions in (C) and (D) use an assumed subunit mass of 734 and 760, respectively, for the comb filter.

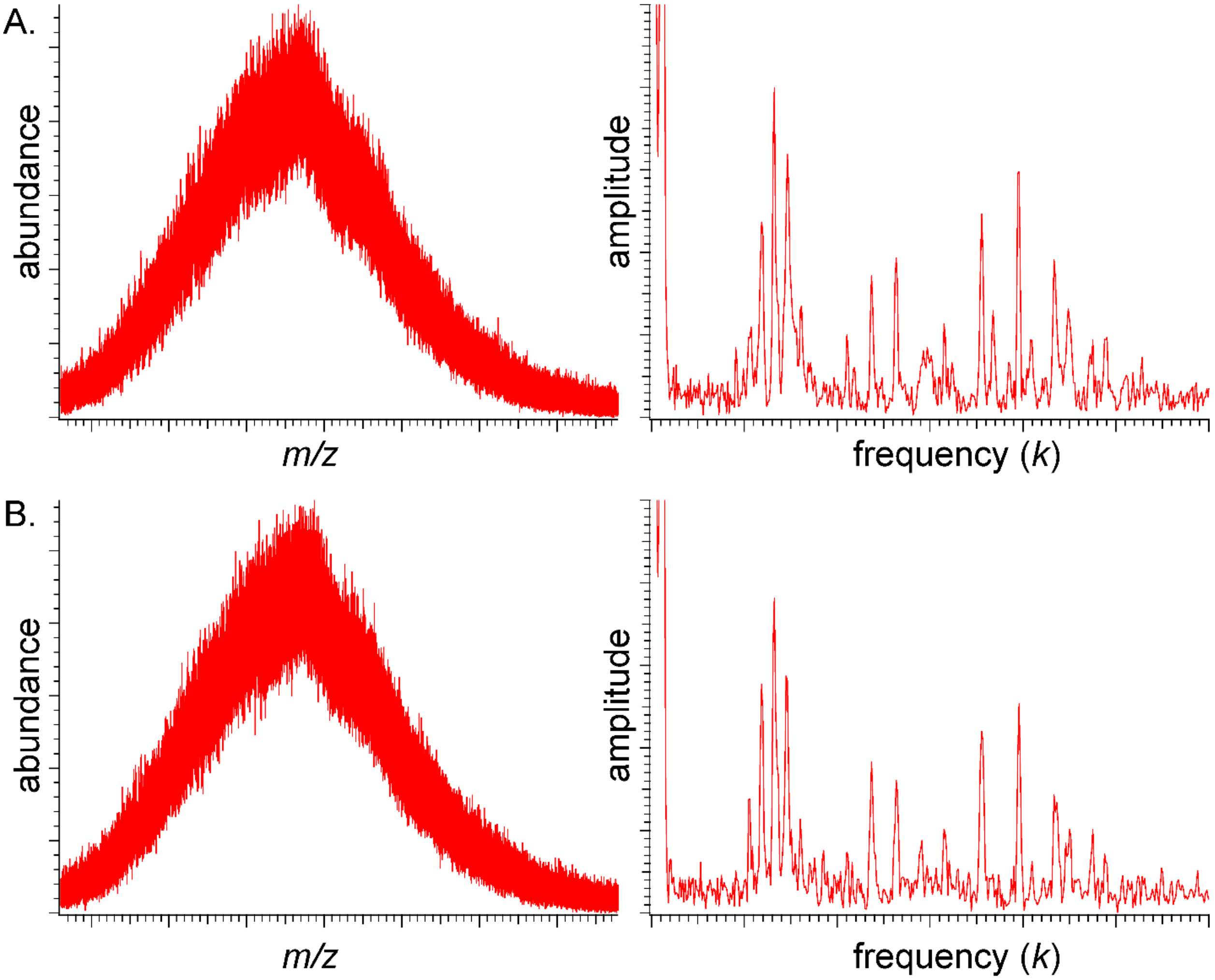

Figure S3. Experimental mass spectrometry data for mixture of two separately-prepared POPC and DPPC Nanodiscs acquired at different time intervals after mixing. Shown above are mass spectra and corresponding Fourier spectra of the same sample shown in Figure 2C in the main manuscript but at longer time intervals after mixing the Nanodisc sample: 20-40 minutes (A) and 40-60 minutes (B).

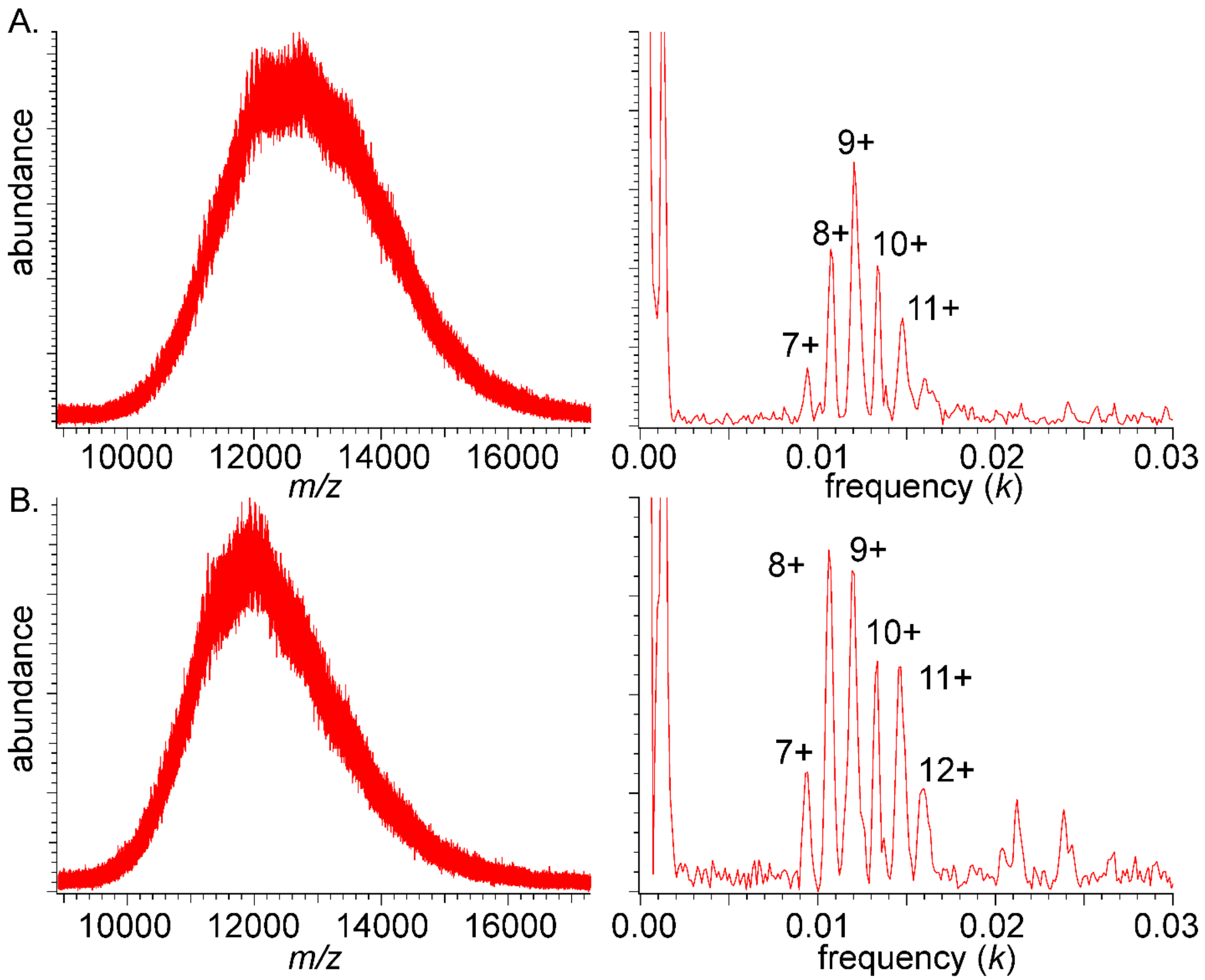

Figure S4. Experimental mass spectra and corresponding Fourier spectra for Nanodiscs assembled from a mixture of 50:50 (A) and 25:75% (B) DPPC:POPC. In contrast to the mixture of two single-lipid Nanodiscs mixed together (see Fig. S3), Nanodiscs assembled from a mixture of two different lipids produces only a single series of peaks in the Fourier spectrum, whose spacing corresponds to the bulk-mole-fraction-weighted average subunit mass.

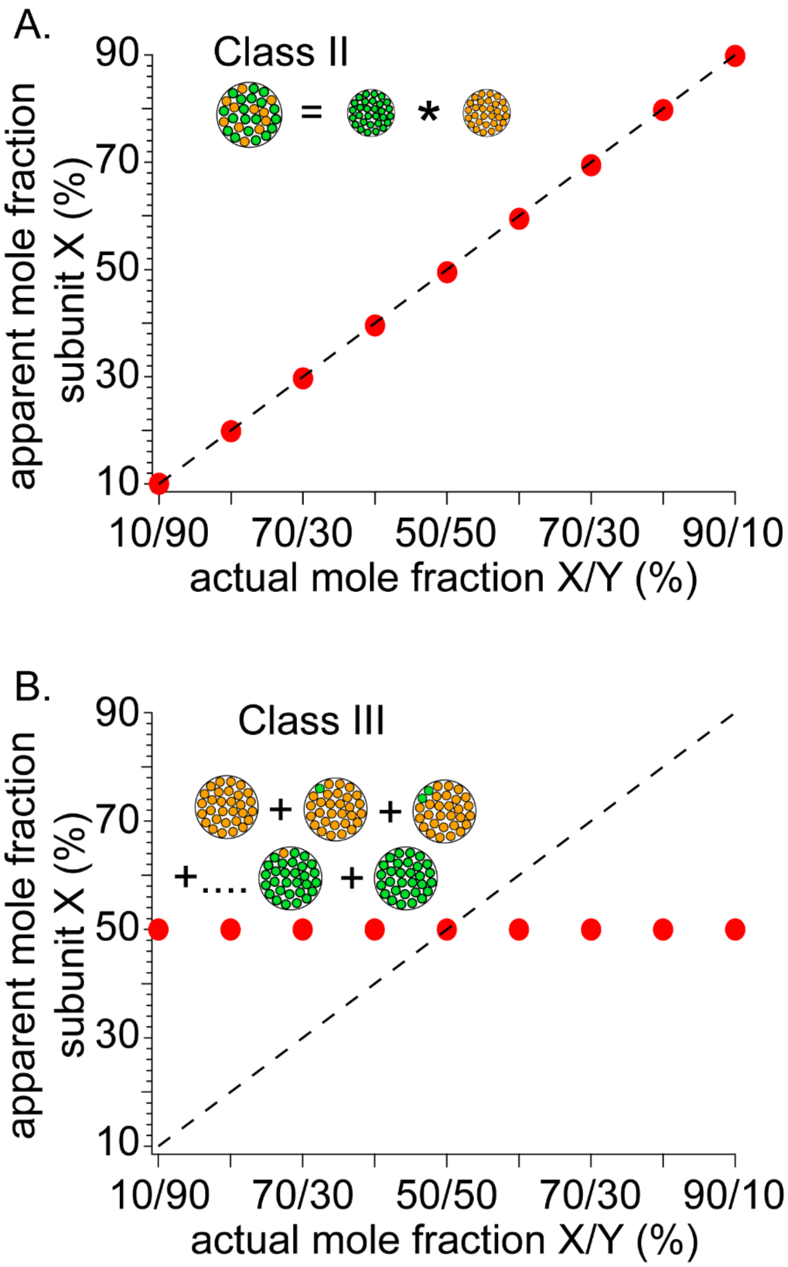

Figure S5. Simulated recovered mole fractions for Class II (A) and Class III (B) ion populations consisting of two types of repeated subunits, X and Y, with masses 44 and 58 Da, respectively, and population average ion mass 1,000 Da. Red dots represent recovered mole fractions determined using Fourier Transform method described in main text under the mean-proportional-variance condition. Dotted line represents perfect agreement with actual mole fractions. Excellent agreement is obtained for Class II, but recovered mole fractions are essentially constant (at ~50/50) for Class III and generally do not reflect actual mole fractions.

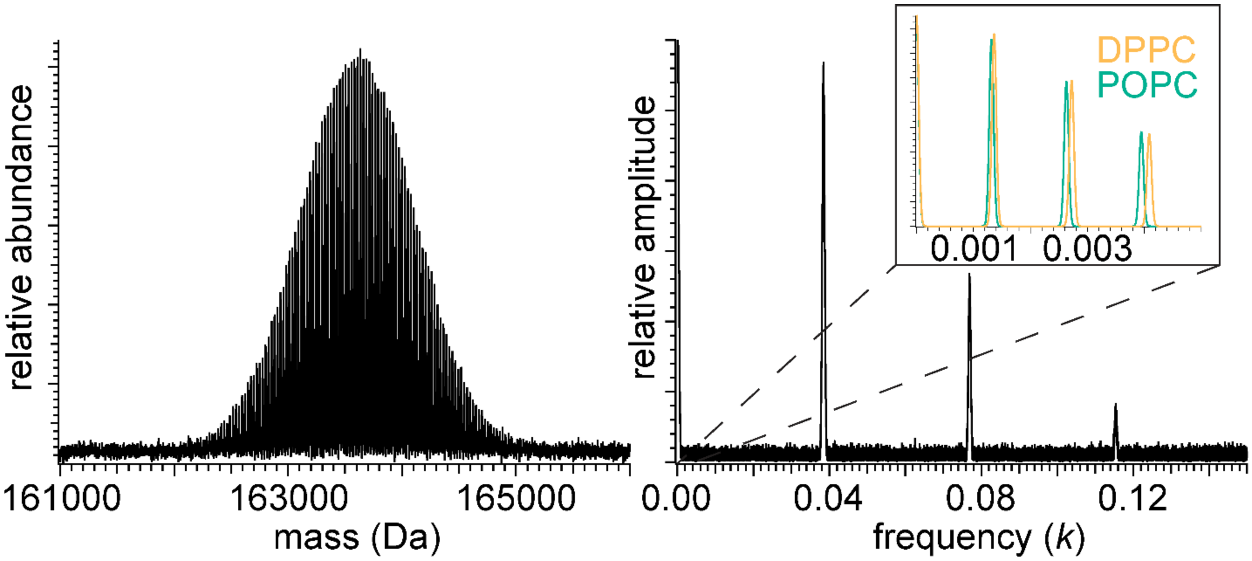

Figure S6. Simulated mass spectrum and corresponding Fourier spectrum for a Class III Nanodisc population formed from a 1:1 bulk mole fraction of DPPC (nominal mass 734 Da) and POPC (nominal mass 760 Da) lipids. A base mass of 44,087.8 Da was used to simulate the mass of two MSP1D1 membrane scaffold proteins. The total number of lipids was held constant at 160 lipids, and all combinations of the different lipids (e.g., 80 DPPC and 80 POPC, 81 DPPC and 79 POPC, etc.) were used to determine the different expected masses. Signal-to-noise of 20:1 was used to illustrate noise tolerance. Peaks in the Fourier spectrum (shown in black) correspond to the reciprocal of the difference of the two lipids (0.038) and the resulting higher harmonics. Also shown in the inset of the Fourier spectrum are the (much lower) frequencies where peaks would be expected for a Class II population formed from DPPC (orange) and POPC (blue) lipids; peaks at these frequencies are clearly absent in the spectrum for the Class III Nanodisc population.

References

1. 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*, 6199.