

UniDecCD: Deconvolution of Charge Detection-Mass Spectrometry Data

Marius M. Kostelic,¹ Ciara K. Zak,¹ Yang Liu,² Victor Shugui Chen,² Zhuchun Wu,² Jared Sivinski,³ Eli Chapman,³ and Michael T. Marty^{1, 4, *}

¹Department of Chemistry and Biochemistry, University of Arizona, Tucson, AZ 85721, USA

²REGENXBIO Inc. 9804 Medical Center Dr, Rockville, MD 20850, USA

³Department of Pharmacology and Toxicology, University of Arizona, Tucson, AZ 85721

⁴Bio5 Institute, University of Arizona, Tucson, AZ 85721, USA

*mtmarty@arizona.edu

KEYWORDS : *Native mass spectrometry, charge detection mass spectrometry, deconvolution, nanodiscs, adeno-associated virus capsids*

ABSTRACT: Native mass spectrometry (MS) has become a versatile tool for characterizing high-mass complexes and measuring biomolecular interactions. Native MS usually requires resolution of different charge states produced by electrospray ionization to measure the mass, which is difficult for highly heterogeneous samples that have overlapping and unresolvable charge states. Charge detection-mass spectrometry (CD-MS) seeks to address this challenge by simultaneously measuring the charge and m/z for isolated ions. However, CD-MS often shows uncertainty in the charge measurement that limits the resolution. To overcome this charge state uncertainty, we developed UniDecCD (UCD) software for computational deconvolution of CD-MS data, which significantly improves the resolution of CD-MS data. Here, we describe the UCD algorithm and demonstrate its ability to improve CD-MS resolution of proteins, megadalton viral capsids, and heterogeneous nanodiscs made from natural lipid extracts. UCD provides a user-friendly interface that will increase the accessibility of CD-MS technology and provide a valuable new computational tool for CD-MS data analysis.

INTRODUCTION:

Since the introduction of electrospray ionization-mass spectrometry (ESI-MS), deconvolution of the charge and mass dimensions of mass spectra has been a critical step in their analysis.¹ Because ESI-MS of peptides and proteins produces ions with multiple charges, measuring the mass requires first assigning the charge states. The most direct way to assign charge states in ESI-MS is from the isotope distribution, which leverages the known mass spacing between isotopologues. For larger proteins or spectra without isotopic resolution, the charge state distribution pattern can be used to assign the charge states. However, for protein complexes that are very large or very polydisperse, overlap between the charge states can create unresolvable data where individual charge states cannot be assigned.

To address this challenge, charge detection-mass spectrometry (CD-MS) seeks to simultaneously measure the m/z ratio and the charge for isolated ions using a charge sensitive detector. CD-MS was first performed on custom-built instruments,²⁻⁷ but conventional Orbitrap detectors have been recently applied for CD-MS by measuring the image current per single ion, which expands the potential reach of the technology.⁸⁻¹⁴ CD-MS is uniquely suited for mass analysis of large and/or polydisperse proteins and protein

complexes, including vaccine and gene therapy delivery vehicles.^{10,15-16}

The primary limitation of CD-MS analysis, as noted in several recent reports, is uncertainty in the charge measurement.¹⁰⁻¹¹ Although great advancements have been made in CD-MS measurements with Orbitraps, the standard deviations of the charge measurements are still 1–3 charges, which cause significant uncertainties in the mass.¹⁰⁻¹¹ Custom instruments have achieved the highest charge resolution with uncertainties less than a single charge, but charge resolution can still limit the measurements in some cases.^{2,5,17}

Here, we address this challenge in charge uncertainty with UniDecCD (UCD), which provides the first deconvolution algorithm for CD-MS data. UCD builds on our widely used UniDec software for deconvolution of ESI-MS data and translates the UniDec algorithm to CD-MS data. As a graphical software package, it provides an open-source, cross-platform, and vendor-neutral interface for CD-MS data analysis and visualization that taps into the wide range of data processing and analysis tools available in UniDec, such as smoothing, background subtraction, binning, and tools for peak assignment and integration. As a deconvolution algorithm, UCD enables fast deconvolution of two-dimensional CD-MS data to computationally reduce the uncertainties in

charge assignment, which dramatically improves the accuracy of CD-MS data from Orbitrap detectors. We demonstrate the utility of CD-MS deconvolution with protein complexes, lipid nanodiscs, and adeno-associated viral (AAV) capsids.

METHODS:

Materials and methods for protein, viral capsid, and nanodisc sample preparation as well as conventional native MS analysis parameters are provided in the Supplemental Methods.

CD-MS Analysis. CD-MS was performed on a Q-Exactive HF quadrupole-Orbitrap mass spectrometer equipped with ultra-high mass range (UHMR) modifications (Thermo Fisher Scientific, Bremen). For all CD-MS data sets, the resolution was set to 240,000, which is a transient time of 512 ms, and the noise threshold was set to 0. CD-MS data sets were collected for 2–15 minutes. Typically, each CD-MS replicate contained at least 5,000 ions after filtering. Ion injection times varied from 0.06–512 ms between samples to decrease or increase the number of ions in the Orbitrap to achieve sufficiently low ion levels for CD-MS, as previously described.¹⁰ Additional instrument settings are provided in Table S1.

For CD-MS analysis of MSP1D1(–), alcohol dehydrogenase (ADH), and bovine serum albumin (BSA) nitrogen was used as the collision gas and the trapping gas pressure was set to 0.1–1. Low m/z detector optimization and low m/z transfer optics were used for BSA and MSP1D1(–). For ADH, low m/z detector optimization was used with high m/z transfer optics. Interestingly, it was crucial to use the correct m/z detector optimization for proteins with different m/z . If the wrong detector optimization was used, we found that the single ion data was biased towards either higher or lower S/N values. ADH, which had a range of 5,300–6,000 m/z , required low m/z detector optimization and samples with higher m/z required high m/z detector optimization. However, this may be different between instruments. Proteins were analyzed with a spray voltage of 1100 V and a m/z range of 1,000–8,000, 2,000–10,000, or 3,000–10,000 m/z . 30 V of HCD voltage was used for BSA to remove adducts.

DPPC, *E. coli*, and brain nanodiscs required a trapping gas pressure setting of 1, high m/z detector optimization, high m/z transfer optics, and a m/z range of 2,000–25,000 m/z was used for natural lipid extract nanodiscs. For DPPC nanodiscs, a m/z range of 6,000–25,000 m/z was used, and the source fragmentation was set to 50 V. For *E. coli* nanodiscs and brain nanodiscs, the source fragmentation was lowered to 30 V to retain more single ions.

GroEL and AAV particles were analyzed with argon as the collision gas. For GroEL, the trapping gas pressure varied from 5–7 to achieve sufficient single-ion levels. For AAV particles, the m/z range was set to 5,000–45,000, the HCD voltage was set to 200 V, the spray voltage was set to 2200–2600 V, and the trapping gas setting varied from 8–10. Detailed deconvolution settings are provided in the Supporting Methods and Table S2.

UCD Software. UCD is available as a module of the open-source UniDec software package in version 5.0.0 (<https://github.com/michaelmarty/UniDec>). UCD was programmed in Python 3 and uses a Model-Presenter-View architecture. The UCD engine (CDEng.py) can be used via command line or scripts for automated processing. It relies primarily on the NumPy and SciPy libraries,^{18–19} but the core deconvolution algorithm has been written to also use CuPy, a CUDA-based GPU-accelerated library, when it is available.²⁰ The GPU acceleration increased the speed of the deconvolution by a factor of 5–20 in our preliminary tests. The UCD presenter (CDPres.py) coordinates between the engine and specific GUI components such as the control panel, menu, and main window. The UCD engine and presenter inherit classes from the UniDec presenter and engine and reuse much of the same code, including key classes for peaks, data, and configuration parameters. Thus, users familiar with UniDec will find the same design and functions with UCD. The primary difference is that conventional UniDec has the core deconvolution algorithm programmed in parallelized C, whereas UCD has the core deconvolution algorithm in Python. Using linearized data and GPU-acceleration has enabled UCD to be developed in Python to make it more accessible at a reasonable speed. The UCD algorithm is broken into two main parts: data processing and deconvolution. A schematic overview of the algorithm is provided in Figure S1.

UCD Data Processing. We will begin by describing the data processing algorithm. First, data is imported from the provided file path. UCD currently supports Thermo Raw (Windows only), mzML, zipped mzML, text, CSV, binary, and NumPy compressed data formats. For text, CSV, binary, and NumPy compressed data, it assumes a format where the first column is the m/z values, the second column are the intensities, and the third column is the optional scan number, which can be left out. For Thermo Raw and mzML data, the m/z , intensities, and scan numbers are imported directly from the data. They also import the injection time. The intensities in the Thermo data need to be corrected by multiplying by the injection time in units of seconds, as previously described.¹⁰ Importantly, for all data types, the data needs to be supplied in centroid mode. If needed, MSConvert can be used to convert Thermo data from profile to centroid prior to opening with UCD.²¹ After importing the data, UCD saves the three columns (m/z , intensity, and scans) to a NumPy array. This array is then saved as a NumPy compressed binary file so that it can be quickly read if the data is reopened.

Finally, the noise level is measured. In the Orbitrap, the main component of the noise is amplifier electrical noise,^{22–23} and it is far more abundant than the chemical noise from ions. We tested several different ways to calculate the noise, including the median noise and the pre-calculated noise included in the Raw file.¹¹ However, we found that the most consistent definition of the noise was the highest peak in the second derivative of the noise intensity distribution above the median, which captures the upper edge of the noise distribution (Figure S2). The lower end of the noise values tended to shift with m/z , which made the median noise shift

as the m/z range was changed. The upper edge of the noise values was more constant and reliable for calibration.

The next step in data processing is filtering the data array to remove unwanted ions. These steps largely follow the procedure outlined by Heck and coworkers.¹⁰ First, ions are filtered by their m/z within a specific range. Next, ions are removed if their centroids are too close together within the same scan. Finally, ions are filtered by their intensity range. We have defined the intensity filtering by charge states because this is more intuitive for users. The intensity range is calculated from the user-supplied charge range based on the intensity/charge or signal-to-noise/charge slope parameter (see below).

After ion filtering, the intensity values for each ion are converted to charge states. There are two options for charge state conversion. The first is to divide the intensities by the slope of the charge vs. intensity plot. Heck and coworkers found the slope to be 12.51 intensity/charge, and we have observed similar slope values in some of our data.¹⁰ However, we found that changing some instrumental parameters such as m/z range can affect the observed slope. To correct for these changes, the slope can be calculated as intensity/(charge * noise) or signal-to-noise/charge.²² Here, the intensity values are divided by the slope times the noise to get the charge. For non-Orbitrap instruments, different slopes can be used, or the charge can be calculated prior to importing into UCD. In this case, a slope of 1 will return the same charge values unaltered.

Next, the list of ions with m/z and z values are pooled with a two-dimensional histogram to get a 2D array of ion counts as a function of m/z and z . The linear bin size for the histogram can be set for both dimensions, but the default bin size for the charge dimension is 1 to give each bin an integer charge value. The charge bin can be set to a non-integer such as 0.1, but the smooth charge states feature currently requires an integer charge bin of 1. The histogram array can then be processed with several techniques, including 2D Gaussian smoothing, intensity thresholding, and background subtraction. Finally, the processed histogram array is then converted from an m/z vs. z array into a resampled mass vs. charge array. The resampling of m/z into mass can be done via interpolation or integration, and the linear bin size for the mass sampling is set by the user. Finally, the mass vs. charge array is summed along the charge axis to produce the final mass distribution. After the final zero-charge mass distribution is assembled, the plots show the m/z vs. charge array, the charge distribution, the m/z distribution, and the mass distribution. The mass vs. charge array can also be plotted.

UCD Deconvolution Algorithm. After data processing, CD-MS data can be deconvolved with a Bayesian deconvolution algorithm analogous to the conventional UniDec algorithm.²⁴ At the core is a Richardson-Lucy deconvolution algorithm.²⁴⁻²⁶ Here, the user defines a 2D "point spread function", which specifies how much an ideal delta function of ions would spread in the instrument in both the m/z and z dimensions. The charge spread function can be measured by analyzing a well-resolved protein where we can confidently assign a given m/z range to a single charge state, and

tools in UCD allow the user to select specific charge states and fit them to Gaussian, Lorentzian, or split Gaussian/Lorentzian functions. The width of the point spread function can be set to 0 in either dimension to assume no spreading in either m/z or z . Setting both to 0 will turn off Richardson-Lucy deconvolution altogether. The Richardson-Lucy deconvolution uses an iterative series of convolutions to remove the point spread function out of the data and has the effect of narrowing 2D peak widths within the data.

Because the charge state is already known from the CD-MS measurement, the Richardson-Lucy deconvolution alone can be sufficient for achieving a good deconvolution. However, other components of the UniDec algorithm can also be useful in improving the deconvolution. One of the core components of the original UniDec algorithm is the smoothing of the charge state distribution, which is essential for assigning charge states in conventional native MS data without isotopic resolution.²⁴ In UCD, a similar smoothing of charge state distribution was implemented. Assuming spectra should have a smooth charge state distribution, this feature increases the intensity of charge states with support from neighboring charge states and decreases the intensity of those with neighboring charge states that have lower intensity. Although not essential, it significantly improves the quality of the deconvolution when distinct charge states are visible. A similar smoothing can be applied in the mass dimension where known mass spacings are expected, which is beneficial for spectra from lipid nanodiscs where peaks are separated by the mass of the lipid. However, charge and mass smoothing may be less useful for highly unresolved and heterogeneous species, where they could produce artifacts if not used carefully.

Finally, two additional deconvolution tactics are available. A SoftMax function can be applied to favor a single charge state per m/z value.²⁷ Unlike conventional UniDec deconvolution, lower SoftMax beta parameters of 1 or 5 were more useful. Also, point smoothing can be applied to smooth the intensities in each charge state among neighboring points.²⁸ In total, there are five potential approaches used during deconvolution: Richardson-Lucy deconvolution, charge state smoothing, mass smoothing, point smoothing, and artifact suppression via SoftMax. Each of these can be used or not used in combination with the others, and the performance of each is largely analogous to conventional UniDec. Following deconvolution, the core histogram array is modified to include the deconvolved histogram array, and the deconvolved m/z vs. z histogram is converted to mass vs. z as described above. Additional mathematical details on the algorithm are provided in the supporting information.

RESULTS:

Calibration of Orbitrap CD-MS Data. Our overall goal was to use computational deconvolution to reduce the charge uncertainty in CD-MS data. Previous publications have shown that the charge uncertainty is most likely due to collisions or dephasing of ions in the Orbitrap from ion-ion repulsion in the Orbitrap.^{11,23} First, we set out to calibrate our Orbitrap mass analyzer for CD-MS data collection. In the Orbitrap, a single ion's image current is directly

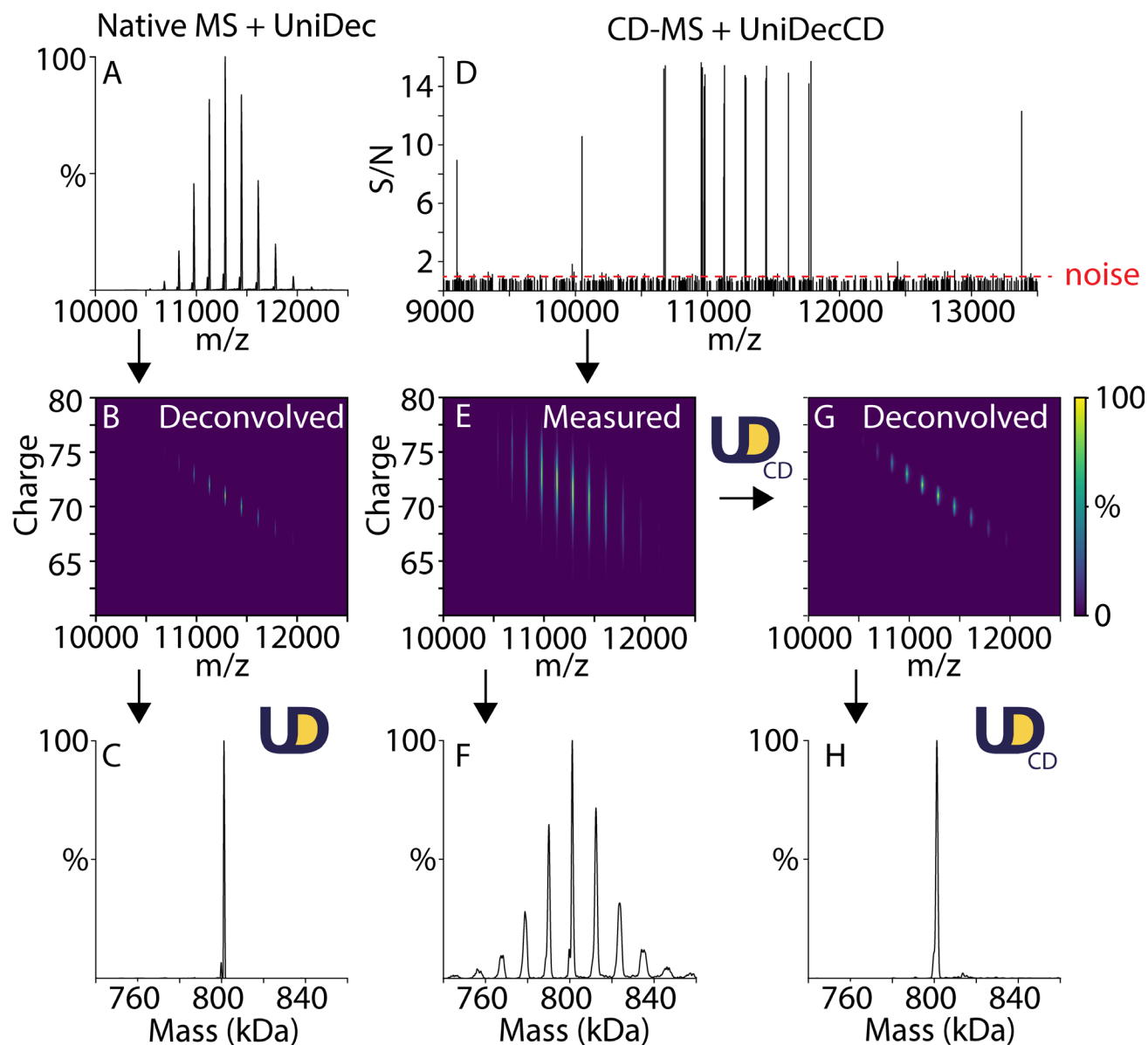


Figure 1. UCD resolves CD-MS of GroEL similarly to conventional native MS. Conventional native MS of GroEL with A) the raw mass spectrum, B) deconvolved charge vs. m/z , and C) deconvolved mass spectrum with UniDec. CD-MS of GroEL with D) one representative scan of single ions with S/N on the y-axis and the noise level indicated as a red dashed line, E) histogram of charge vs. m/z based on the raw data, and F) mass distribution for the raw CD-MS data, which shows significant uncertainty. In contrast, G) the UCD deconvolved charge vs. m/z and H) the deconvolved CD-MS mass distribution show single charge state resolution.

proportional to its charge, and prior studies have shown they scale linearly.^{10,13} We therefore set out to create a calibration plot, similar to Wörner et al.,¹⁰ that would convert a single ion's intensity to charge. We analyzed a variety of well-resolved protein complexes ranging from 22–801 kDa with CD-MS and plotted their signal intensity vs. their known charge states from their native mass spectra. However, we found that the signal intensity/charge slope in our Orbitrap was variable due to different instrument parameters that were required between different samples (Figure S3A). This variability could be reduced to an acceptable fit by limiting the m/z range to a narrow window around the analyte of interest (Figure S4), but this approach would be

limited for samples with ions spread across a wider range of m/z values.

Therefore, similar to Makarov and coworkers,¹¹ we used the signal-to-noise (S/N) ratio rather than the raw intensity to measure the charge,²² which internally controls for any unexpected deviations in the injection time. Here, the noise level likely serves as an internal correction for the injection time or other instrumental settings that may affect the absolute intensity per second that is reported. For our S/N ratio, as described above, we defined the noise as the peak of the second derivative of the intensity distribution above the median noise level, which quantifies the upper edge of the noise distribution (Figure S2). We found that this definition

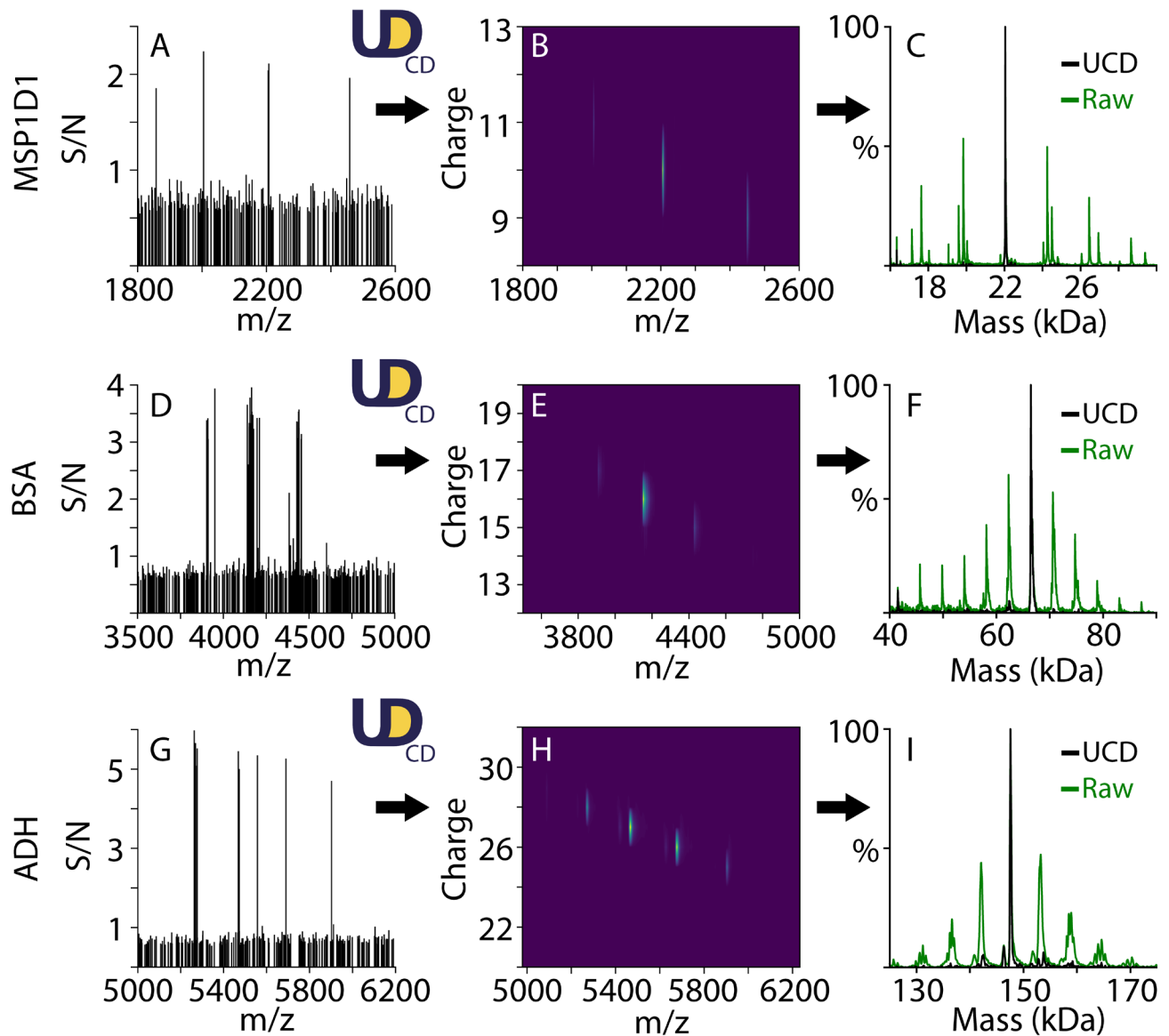


Figure 2. UCD deconvolution of CD-MS data from MSP1D1(–) (A–C), BSA (D–F), and ADH (G–I). Each row shows the (A, D, G) raw mass spectrum of one scan of single ions with S/N on the y-axis, (B, E, H) charge vs. m/z histogram after deconvolution, and (C, F, I) deconvolved mass distribution in *black* overlaid with the raw mass distribution without UCD in *green*.

of noise was consistent across the m/z spectrum, unlike the median and lower end of the noise distribution, which varied across the m/z range and were thus susceptible to changing when the m/z range was adjusted. Our calibration yielded a slope of 0.2083 S/N per charge (Figure S3B). However, it would be best for users to perform their own calibration because there may be variability instrument to instrument, drift over time, and variability due to different transient times and different Orbitrap sizes.

To compare the raw intensity and S/N calibration methods, we measured the root mean square deviation (RMSD) between the median measured charge values from our standards and the fit charge values from the calibrations (Figure S4). We found that the signal intensity to charge conversion with a narrow m/z range had an average RMSD

of 1.38 e, and the S/N to charge conversion had a slightly lower RMSD of 0.71 e, indicating a moderately more accurate fit for this data set.

Deconvolution Improves Charge Precision. After we had a reliable calibration to convert S/N to charge, we incorporated the UCD deconvolution algorithm. Although the charge can be directly measured with CD-MS, we found that the raw CD-MS data gave typical charge state standard deviations of around 2 charges. For example, although GroEL gave nicely resolved charge states when analyzed with conventional native MS (Figure 1A–C), the raw CD-MS data had a standard deviation of ± 2.2 charges (as seen by the vertical spread in Figure 1E). These uncertainties caused multiple artifact mass peaks due to incorrect charge assignments (Figure 1F).

By applying Richardson-Lucy deconvolution and charge state smoothing, deconvolution with UCD removed the charge state uncertainty from the CD-MS data, and we were able to resolve single charge states of GroEL (Figure 1G). In essence, UCD removed the point spread function to reduce the charge state uncertainty from around ± 2 charges to almost 0. The removal of the charge state uncertainty removed the artifact mass peaks and resulted in a single mass peak for GroEL that matched the conventional native mass spectrum (Figure 1H and 1C). Even with no charge state uncertainty, the peaks are slightly broader than their conventional native MS counterparts, and we expect that this loss

of m/z resolution is due to frequency drifting of ions over the longer transient times used for CD-MS.²³

We then tested UCD on other proteins and protein complexes with known charge states: MSP1D1(-), BSA, and ADH. With UCD deconvolution, we were able to remove the charge uncertainty from the CD-MS spectra and confidently assign single charge states for MSP1D1(-), BSA, and ADH (Figure 2). The raw CD-MS data (shown in green in Figure 2) showed multiple artifact mass peaks because each charge state had a standard deviation of around 2 charges. UCD removed the charge state uncertainty and gave single mass

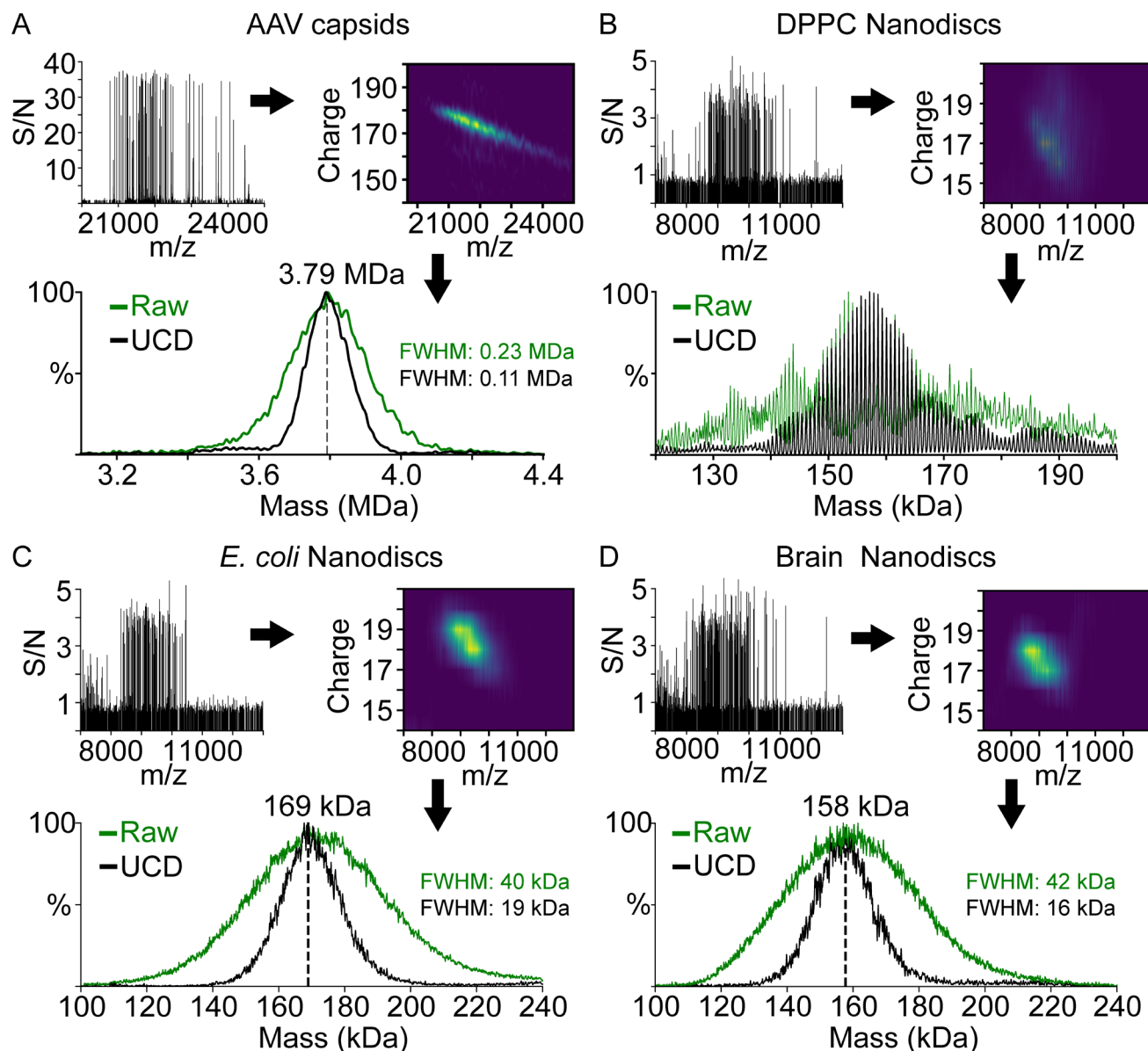


Figure 3. Systems with overlapping charge states measured by CD-MS and deconvolved using UCD. A) AAV viral capsids, B) DPPC nanodiscs, C) *E. coli* nanodiscs, and D) brain nanodiscs. Each panel (A–D) has a representative scan of single ions (top left), their total deconvolved charge vs. m/z histogram (top right), their deconvolved mass spectrum (bottom) with the raw CD-MS data in green and the UCD deconvolved mass distribution in black. The FWHM values are indicated for raw CD-MS in green and deconvolved in black. The center of the mass and FWHM are listed for a single replicate, and the text indicates the average across three replicate samples.

peaks that corresponded to the correct mass for each protein that matched their conventional native MS data (Figure S5).

It is also possible to remove the charge uncertainty with the charge bin set to 0.1, and the Richardson-Lucy algorithm gives similar results independent of the charge binning (Figure S6). However, the charge state smoothing, which currently requires a charge bin of 1, is necessary to fully remove the uncertainty. These data demonstrated that the addition of the UCD deconvolution provided precise assignment of charge states that gave more confident mass assignments for CD-MS data.

Deconvolution of AAV particles. After we validated our UCD deconvolution method on proteins with well-resolved charge states, we began measuring the charge of samples with increasing complexity to test the algorithm on complexes with unresolved charge states. First, we measured the mass of empty AAV capsids, which are used as gene therapy delivery systems.²⁹⁻³⁰ AAV capsids are complex, and it is currently difficult to measure small differences between capsids. Recently, important progress has been made with characterizing differences between empty capsids, partially filled, and filled capsids with CD-MS,^{10,16} and we wanted to test UCD with a basic AAV capsid system. Without UCD deconvolution, from the raw CD-MS data we found that the AAV capsids had an average mass of 3.81 ± 0.01 MDa (standard deviation from the average mass), which matched the expected mass of 3.78-3.84 MDa (Figure 3A). The expected mass is calculated based on the 60-mer stoichiometry of the three AAV8 viral proteins as 1:1:10, and similar masses have been previously reported.^{16,31-32} The width of the mass distribution is likely caused by slight deviations in the 1:1:10 stoichiometry. With UCD, we found an average mass distribution of 3.82 ± 0.02 MDa (also $n=3$). An F-test showed no statistical difference in the standard deviation of the mean, so the deconvolution algorithm did not significantly impact the accuracy or precision of the measured mass.

However, UCD significantly improved the resolution of the CD-MS measurement. The mass distribution for the AAV capsids from the raw CD-MS data had an average FWHM of 274 kDa. UCD deconvolution lowered the average FWHM to 144 kDa (Figure 3A). In this case, we did not use the charge state smoothing with this unresolved data to avoid introducing artifacts. Instead, we simply used Richardson-Lucy deconvolution with point spread function extrapolated based on the GroEL data. This data shows that UCD can deconvolve and improve the resolution of CD-MS data obtained from high-mass, heterogeneous complexes that are not resolvable by conventional native MS.

CD-MS of Nanodiscs. We next analyzed DPPC nanodiscs with CD-MS to test UCD with a complex sample with well-resolved but overlapping charge states. By adding the mass smoothing function from UniDec into UCD, we were able to deconvolve DPPC nanodiscs with resolved lipid mass peaks, similar to the conventional native MS data (Figure 3B and S7-9), demonstrating excellent mass resolution. However, the average FWHM of the mass peaks were somewhat larger in the CD-MS measurements than the conventional native MS measurements (18.6 kDa vs. 10.5 kDa), which likely

reflects some residual uncertainty in the charge measurement not removed during deconvolution.

We next set out to study nanodiscs made from natural lipid extracts. Mixed lipid nanodiscs can be resolved by native MS if the lipids are close in mass or resonant,^{28,33} but natural lipid membranes have many lipids with different masses, making them essentially impossible to resolve with conventional native MS. As expected, nanodiscs made with lipids from *E. coli* polar extract and porcine brain polar extract showed broad and unresolvable humps with conventional native MS. With CD-MS, we found their charge and mass values were similar to well-resolved DPPC nanodiscs (Figure 3C-D), but the mass distributions were very broad. Using UCD with only the Richardson-Lucy deconvolution (no charge or mass smoothing) significantly reduced the width of the mass distribution, resulting in a polydispersity similar to the well-resolved DPPC nanodiscs (Figure 3C-D).

Although we could not resolve the individual lipid mass peaks in natural lipid nanodiscs, the intact mass distribution gave useful information on their size and polydispersity that helped to optimize their assembly. Using the optimized assembly conditions, *E. coli* nanodiscs had a center of mass of 166.0 ± 4.1 kDa (standard deviation of the average mean from three replicates), and brain nanodiscs had a center of mass of 152.8 ± 3.8 kDa (Figure 3C, 3D, and S10). Porcine brain lipid extract nanodiscs had a lower center of mass compared to DPPC nanodiscs (152.8 ± 3.8 kDa vs. 161.9 ± 3.6 kDa), which indicated looser lipid packing compared to DPPC nanodiscs. *E. coli* nanodiscs were slightly larger but not statistically different than DPPC nanodiscs. Interestingly, the average FWHM of the mass distributions of *E. coli* (20.2 kDa) and brain (15.4 kDa) nanodiscs matched control DPPC nanodiscs (18.6 kDa). The close similarities in the measured FWHM of the mass distributions between the DPPC nanodiscs and natural lipid extract nanodiscs demonstrates the ability of UCD to improve the precision of complex CD-MS data and shows a similar polydispersity between natural and synthetic lipid nanodiscs.

Overall, UCD deconvolution greatly improved the resolution of *E. coli* and brain nanodiscs by reducing their charge uncertainty. Natural lipid extract nanodiscs provide a more native environment to study membrane proteins and have been used for obtaining structures of membrane proteins with cryo-electron microscopy.³⁴ CD-MS aided by UCD provides a useful technique for the precise characterization of the mass and polydispersity of natural lipid nanodiscs that greatly helps in optimizing nanodisc assembly.

CONCLUSIONS:

Here, we have demonstrated that computational deconvolution reduces the charge uncertainty and allows single charge state resolution of CD-MS data. We first showed that UCD is a reliable method for measuring the charge states of known proteins. We then used UCD to measure the charge and mass of AAV viral particles that were not resolvable with conventional native MS. Finally, UCD allowed the resolution of nanodiscs made with *E. coli* and brain lipid extract, providing the first native MS-based measurement of natural membrane nanodiscs.

Rising advancements in CD-MS technology such as selective temporal overview of resonant ions (STORI) plots and frequency tracking promise to further increase the resolution gained with Orbitrap-based CD-MS and will also contribute to reducing the charge state uncertainty.^{14,23} However, these approaches do not currently provide single charge state resolution, so we anticipate that UCD will be highly valuable for this type of data as well. Even for custom-built instruments where single charge state resolution is achievable and deconvolution is unnecessary,³⁵ UCD provides an easy-to-use interface for CD-MS data analysis and visualization. It also enables CD-MS users to take advantage of the suite of auxiliary analysis tools in UniDec for peak assignment, intensity extraction, and other quantitative analysis methods.

Overall, UCD provides an open-source, graphical CD-MS data analysis and deconvolution software that will advance the use of CD-MS for studying complex, high-mass, and heterogeneous samples and will push the boundaries of what problems can be solved by MS.

AUTHOR INFORMATION

Corresponding Author

* mtmarty@arizona.edu

SUPPORTING INFORMATION

The Supporting Information contains the materials and methods for protein, viral capsid, and nanodisc sample preparation. This also contains supplemental figures and supplemental tables that thoroughly describe the instrumental parameters and deconvolution parameters used between samples as well as a mathematical description of the algorithm. The Supporting Information is available free of charge via the internet at <http://pubs.acs.org>

ACKNOWLEDGMENTS

We thank the REGENXBIO upstream and downstream development department for their support on this project. pMSP1D1 plasmid was a gift from Stephen Sligar (Addgene plasmid #20061). The authors thank Maria Reinhardt-Szyba, Kyle Fort, and Alexander Makarov at Thermo Fisher Scientific for their support on the UHMR Q-Exactive HF instrument and for helpful suggestions on tuning for CD-MS. This work was funded by the National Science Foundation (CHE-1845230) and National Institutes of Health (T32 GM08804).

ABBREVIATIONS

UCD, UniDec charge detection; CD-MS, charge-detection mass spectrometry; DPPC, dipalmitoyl-phosphocholine; S/N, signal to noise.

REFERENCES

1. Rolland, A. D.; Prell, J. S., Approaches to Heterogeneity in Native Mass Spectrometry. *Chem. Rev.* **2021**, 10.1021/acs.chemrev.1c00696.
2. Keifer, D. Z.; Shinholt, D. L.; Jarrold, M. F., Charge Detection Mass Spectrometry with Almost Perfect Charge Accuracy. *Anal. Chem.* **2015**, 87 (20), 10330-10337.
3. Fuerstenau, S. D.; Benner, W. H., Molecular weight determination of megadalton DNA electrospray ions using charge

detection time-of-flight mass spectrometry. *Rapid. Commun. Mass Spectrom.* **1995**, 9 (15), 1528-1538.

4. Doussineau, T.; Kerleroux, M.; Dagany, X.; Clavier, C.; Barbaire, M.; Maurelli, J.; Antoine, R.; Dugourd, P., Charging megadalton poly(ethylene oxide)s by electrospray ionization. A charge detection mass spectrometry study. *Rapid. Commun. Mass Spectrom.* **2011**, 25 (5), 617-623.
5. Todd, A. R.; Barnes, L. F.; Young, K.; Zlotnick, A.; Jarrold, M. F., Higher Resolution Charge Detection Mass Spectrometry. *Anal. Chem.* **2020**, 92 (16), 11357-11364.
6. Contino, N. C.; Jarrold, M. F., Charge detection mass spectrometry for single ions with a limit of detection of 30 charges. *Int. J. Mass Spectrom.* **2013**, 345-347, 153-159.
7. Todd, A. R.; Alexander, A. W.; Jarrold, M. F., Implementation of a Charge-Sensitive Amplifier without a Feedback Resistor for Charge Detection Mass Spectrometry Reduces Noise and Enables Detection of Individual Ions Carrying a Single Charge. *J. Am. Soc. Mass. Spectrom.* **2020**, 31 (1), 146-154.
8. Rose, R. J.; Damoc, E.; Denisov, E.; Makarov, A.; Heck, A. J. R., High-sensitivity Orbitrap mass analysis of intact macromolecular assemblies. *Nat. Methods.* **2012**, 9 (11), 1084-1086.
9. Kafader, J. O.; Melani, R. D.; Senko, M. W.; Makarov, A. A.; Kelleher, N. L.; Compton, P. D., Measurement of Individual Ions Sharply Increases the Resolution of Orbitrap Mass Spectra of Proteins. *Anal. Chem.* **2019**, 91 (4), 2776-2783.
10. 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.
11. Denisov, E.; Damoc, E.; Makarov, A., Exploring frontiers of orbitrap performance for long transients. *Int. J. Mass Spectrom.* **2021**, 466, 116607.
12. McGee, J. P.; Melani, R. D.; Yip, P. F.; Senko, M. W.; Compton, P. D.; Kafader, J. O.; Kelleher, N. L., Isotopic Resolution of Protein Complexes up to 466 kDa Using Individual Ion Mass Spectrometry. *Anal. Chem.* **2021**, 93 (5), 2723-2727.
13. Makarov, A., Electrostatic Axially Harmonic Orbital Trapping: A High-Performance Technique of Mass Analysis. *Anal. Chem.* **2000**, 72 (6), 1156-1162.
14. Kafader, J. O.; Beu, S. C.; Early, B. P.; Melani, R. D.; Durbin, K. R.; Zabrouskov, V.; Makarov, A. A.; Maze, J. T.; Shinholt, D. L.; Yip, P. F.; Kelleher, N. L.; Compton, P. D.; Senko, M. W., STORI Plots Enable Accurate Tracking of Individual Ion Signals. *J. Am. Soc. Mass. Spectrom.* **2019**, 30 (11), 2200-2203.
15. Wörner, T. P.; Shamorkina, T. M.; Snijder, J.; Heck, A. J. R., Mass Spectrometry-Based Structural Virology. *Anal. Chem.* **2021**, 93 (1), 620-640.
16. Pierson, E. E.; Keifer, D. Z.; Asokan, A.; Jarrold, M. F., Resolving Adeno-Associated Viral Particle Diversity With Charge Detection Mass Spectrometry. *Anal. Chem.* **2016**, 88 (13), 6718-6725.
17. Todd, A. R.; Jarrold, M. F., Dynamic Calibration Enables High-Accuracy Charge Measurements on Individual Ions for Charge Detection Mass Spectrometry. *J. Am. Soc. Mass. Spectrom.* **2020**, 31 (6), 1241-1248.
18. Harris, C. R.; Millman, K. J.; van der Walt, S. J.; Gommers, R.; Virtanen, P.; Cournapeau, D.; Wieser, E.; Taylor, J.; Berg, S.; Smith, N. J.; Kern, R.; Picus, M.; Hoyer, S.; van Kerkwijk, M. H.; Brett, M.; Haldane, A.; del Río, J. F.; Wiebe, M.; Peterson, P.; Gérard-Marchant, P.; Sheppard, K.; Reddy, T.; Weckesser, W.; Abbasi, H.; Gohlke, C.; Oliphant, T. E., Array programming with NumPy. *Nature* **2020**, 585 (7825), 357-362.
19. Virtanen, P.; Gommers, R.; Oliphant, T. E.; Haberland, M.; Reddy, T.; Cournapeau, D.; Burovski, E.; Peterson, P.; Weckesser, W.; Bright, J.; van der Walt, S. J.; Brett, M.; Wilson, J.; Millman, K. J.; Mayorov, N.; Nelson, A. R. J.; Jones, E.; Kern, R.; Larson, E.; Carey, C. J.; Polat, I.; Feng, Y.; Moore, E. W.; VanderPlas, J.; Laxalde, D.; Perktold, J.; Cimrman, R.; Henriksen, I.; Quintero, E. A.; Harris, C. R.; Archibald, A. M.; Ribeiro, A. H.; Pedregosa, F.; van Mulbregt, P.; Vijaykumar, A.; Bardelli, A. P.; Rothberg, A.; Hilboll, A.; Kloeckner, A.; Scopatz, A.; Lee, A.; Rokem, A.; Woods, C. N.; Fulton, C.; Masson, C.; Häggström, C.; Fitzgerald, C.; Nicholson, D. A.; Hagen, D. R.; Pasechnik, D. V.; Olivetti, E.; Martin, E.; Wieser, E.; Silva, F.; Lenders, F.; Wilhelm, F.; Young, G.; Price, G. A.

- Ingold, G.-L.; Allen, G. E.; Lee, G. R.; Audren, H.; Probst, I.; Dietrich, J. P.; Silterra, J.; Webber, J. T.; Slavič, J.; Nothman, J.; Buchner, J.; Kulick, J.; Schönberger, J. L.; de Miranda Cardoso, J. V.; Reimer, J.; Harrington, J.; Rodríguez, J. L. C.; Nunez-Iglesias, J.; Kuczyński, J.; Tritz, K.; Thoma, M.; Newville, M.; Kümmerer, M.; Bolingbroke, M.; Tartre, M.; Pak, M.; Smith, N. J.; Nowaczyk, N.; Shebanov, N.; Pavlyk, O.; Brodtkorb, P. A.; Lee, P.; McGibbon, R. T.; Feldbauer, R.; Lewis, S.; Tygier, S.; Sievert, S.; Vigna, S.; Peterson, S.; More, S.; Pudlik, T.; Oshima, T.; Pingel, T. J.; Robitaille, T. P.; Spura, T.; Jones, T. R.; Cera, T.; Leslie, T.; Zito, T.; Krauss, T.; Upadhyay, U.; Halchenko, Y. O.; Vázquez-Baeza, Y.; SciPy, C., SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods*. **2020**, *17* (3), 261-272.
20. Okuta, R.; Unno, Y.; Nishino, D.; Hido, S.; Crissman, CuPy: A NumPy-Compatible Library for NVIDIA GPU Calculations. In *Proceedings of Workshop on Machine Learning Systems (LearningSys) in The Thirty-first Annual Conference on Neural Information Processing Systems (NIPS)* [Online] 2017. http://learningsys.org/nips17/assets/papers/paper_16.pdf.
21. Chambers, M. C.; Maclean, B.; Burke, R.; Amodei, D.; Ruderman, D. L.; Neumann, S.; Gatto, L.; Fischer, B.; Pratt, B.; Egertson, J.; Hoff, K.; Kessner, D.; Tasman, N.; Shulman, N.; Frewen, B.; Baker, T. A.; Brusniak, M.-Y.; Paulse, C.; Creasy, D.; Flashner, L.; Kani, K.; Moulding, C.; Seymour, S. L.; Nuwaysir, L. M.; Lefebvre, B.; Kuhlmann, F.; Roark, J.; Rainer, P.; Detlev, S.; Hemenway, T.; Huhmer, A.; Langridge, J.; Connolly, B.; Chadick, T.; Holly, K.; Eckels, J.; Deutsch, E. W.; Moritz, R. L.; Katz, J. E.; Agus, D. B.; MacCoss, M.; Tabb, D. L.; Mallick, P., A cross-platform toolkit for mass spectrometry and proteomics. *Nat. Biotechnol.* **2012**, *30* (10), 918-920.
22. Makarov, A.; Denisov, E., Dynamics of Ions of Intact Proteins in the Orbitrap Mass Analyzer. *J. Am. Soc. Mass. Spectrom.* **2009**, *20* (8), 1486-1495.
23. Wörner, T. P.; Aizikov, K.; Snijder, J.; Fort, K. L.; Makarov, A. A.; Heck, A. J. R., Frequency chasing of individual megadalton ions in an Orbitrap analyzer improves precision of analysis in single molecule mass spectrometry. *bioRxiv* **2021**, 10.1101/2021.06.15.448530, 2021.06.15.448530.
24. Marty, M. T.; Baldwin, A. J.; Marklund, E. G.; Hochberg, G. K.; Benesch, J. L.; Robinson, C. V., Bayesian deconvolution of mass and ion mobility spectra: from binary interactions to polydisperse ensembles. *Anal. Chem.* **2015**, *87* (8), 4370-4376.
25. Lucy, L. B., An iterative technique for the rectification of observed distributions. *Astron.* **1974**, *79*, 745.
26. Richardson, W. H., Bayesian-Based Iterative Method of Image Restoration. *J. Opt. Soc. Am.* **1972**, *62* (1), 55-59.
27. Marty, M. T., Eliminating Artifacts in Electrospray Deconvolution with a SoftMax Function. *J. Am. Soc. Mass. Spectrom.* **2019**, *30* (10), 2174-2177.
28. Kostelic, M. M.; Ryan, A. M.; Reid, D. J.; Noun, J. M.; Marty, M. T., Expanding the Types of Lipids Amenable to Native Mass Spectrometry of Lipoprotein Complexes. *J. Am. Soc. Mass. Spectrom.* **2019**, *30* (8), 1416-1425.
29. Naso, M. F.; Tomkowicz, B.; Perry, W. L.; Strohl, W. R., Adeno-Associated Virus (AAV) as a Vector for Gene Therapy. *BioDrugs* **2017**, *31* (4), 317-334.
30. Vandenberghe, L. H.; Wilson, J. M.; Gao, G., Tailoring the AAV vector capsid for gene therapy. *Gene Ther.* **2009**, *16* (3), 311-319.
31. Nam, H.-J.; Gurda, B. L.; McKenna, R.; Potter, M.; Byrne, B.; Salganik, M.; Muzyczka, N.; Agbandje-McKenna, M., Structural studies of adeno-associated virus serotype 8 capsid transitions associated with endosomal trafficking. *Journal of virology* **2011**, *85* (22), 11791-11799.
32. Wörner, T. P.; Bennett, A.; Habka, S.; Snijder, J.; Frieze, O.; Powers, T.; Agbandje-McKenna, M.; Heck, A. J. R., Adeno-associated virus capsid assembly is divergent and stochastic. *Nat. Commun.* **2021**, *12* (1), 1642-1642.
33. Kostelic, M. M.; Zak, C. K.; Jayasekera, H. S.; Marty, M. T., Assembly of Model Membrane Nanodiscs for Native Mass Spectrometry. *Anal. Chem.* **2021**, *93* (14), 5972-5979.
34. Hofmann, S.; Janulienė, D.; Mehdipour, A. R.; Thomas, C.; Stefan, E.; Bruchert, S.; Kuhn, B. T.; Geertsma, E. R.; Hummer, G.; Tampe, R.; Moeller, A., Conformation space of a heterodimeric ABC exporter under turnover conditions. *Nature* **2019**, *571* (7766), 580-583.
35. Miller, L. M.; Barnes, L. F.; Raab, S. A.; Draper, B. E.; El-Baba, T. J.; Lutomski, C. A.; Robinson, C. V.; Clemmer, D. E.; Jarrold, M. F., Heterogeneity of Glycan Processing on Trimeric SARS-CoV-2 Spike Protein Revealed by Charge Detection Mass Spectrometry. *J. Am. Chem. Soc.* **2021**, *143* (10), 3959-3966.

TOC Figure:

