



Published in final edited form as:

Nat Methods. 2020 April ; 17(4): 391–394. doi:10.1038/s41592-020-0764-5.

Multiplexed Mass Spectrometry of Individual Ions Improves Measurement of Proteoforms and Their Complexes

Jared O. Kafader[†], Rafael D. Melani[†], Kenneth R. Durbin[†], Bon Ikwuagwu[†], Bryan P. Early[†], Ryan T. Fellers[†], Steven C. Beu[‡], Vlad Zabrouskov[‡], Alexander A. Makarov[§], Joshua T. Maze[#], Deven L. Shinholt[#], Ping F. Yip[‡], Danielle Tullman-Ercek[†], Michael W. Senko[‡], Philip D. Compton^{*,†}, Neil L. Kelleher^{*,†}

[†]Departments of Chemistry, Molecular Biosciences, and Chemical and Biological Engineering, the Chemistry of Life Processes Institute, the Proteomics Center of Excellence at Northwestern University, Evanston, Illinois 60208, United States

[‡]S.C. Beu Consulting, Austin, Texas 78729, United States

[‡]Thermo Fisher Scientific, San Jose, California 95134, United States

[#]Thermo Fisher Scientific, Austin, Texas 78728, United States

[§]Thermo Fisher Scientific, Bremen 28199, Germany

Abstract

A new Orbitrap-based ion analysis procedure is shown to be possible by determining the direct charge for numerous individual protein ions to generate true mass spectra. The deployment of an Orbitrap system for charge detection enables the characterization of highly complicated mixtures of proteoforms and their complexes in both denatured and native modes of operation, revealing information not obtainable by traditional measurement of an ensemble of ions.

For decades, mass spectrometry has used ions to measure the mass-to-charge ratio of molecules once lifted into the gas phase¹. Denatured and native electrospray ionization of intact proteins and their complexes pose many complications due to sample heterogeneity and large charge state envelopes in the m/z domain. To simplify analysis, Charge Detection Mass Spectrometry (CDMS)^{2–5} has enabled the generation of true *mass* spectra with the direct readout of an ion's integer charge value (e.g. 35+, 34+, etc.). Here, we bring the commercially available Orbitrap mass analyzer into the CDMS universe to multiplex and

Users may view, print, copy, and download text and data-mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use:http://www.nature.com/authors/editorial_policies/license.html#terms

*Corresponding authors: philip-compton@northwestern.edu and n-kelleher@northwestern.edu.

Author Contributions

Data acquisition and analysis was performed by J.O.K with assistance from R.D.M. and K.R.D with additional input from all authors. R.D.M. prepared samples with help from J.O.K, while B.I. and D.T.E. assembled and purified the MS2 virus-like particles. J.O.K., B.P.E., K.R.D., R.T.F., S.C.B., P.F.Y., M.W.S., and P.D.C. developed models and created programs to analyze the data. K.R.D., V.Z., A.A.M., J.T.M., D.L.S., M.W.S., P.D.C., and N.L.K. provided additional technical support and guidance. N.L.K., P.D.C., and M.W.S. collected funding support and oversaw the project. J.O.K. and N.L.K. wrote the manuscript. All authors critically reviewed the manuscript.

COI Statement

V.Z., A.A.M., J.T.M., D.L.S., P.F.Y., and M.W.S. are employees of Thermo Fisher Scientific.

regularize this approach for creation of mass spectra on extremely complex samples. The approach is shown to be general for measuring complex proteoform mixtures and their complexes, covering a range of masses from 8 kDa to 3.2 MDa, with Orbitrap-based Individual Ion Mass Spectrometry (I²MS) capable of precise mass determination where ensemble measurements fail or require prior separation.

Previously, single ion sensitivity has been demonstrated using an Orbitrap mass analyzer^{6, 7}, analogous to linear geometry trapping instruments more frequently used for CDMS. In 2018, we used an Orbitrap analyzer to acquire one-at-a-time measurements on single ions, and showed that >20-fold increases of resolving power over ensemble measurements were possible by centroiding m/z values of individual ions⁸. Here, we extend this methodology substantially with a complete workflow for CDMS that we call I²MS (Fig. 1) to directly measure the charge of individual ions. Further, we show this method to be fully established and functional for the simultaneous measurement of 200 individual ions.

In Step 1 (Fig. 1) of the I²MS approach, ~120 ions are observed per acquisition in a random-style trapping event. Parallelizing ion observation to >100 ions per acquisition decreases the data acquisition burden by ~100-fold over current CDMS techniques⁹. Due to the low ion count needed for I²MS, protein solutions can be diluted down by ~three orders of magnitude (into the high pM to low nM range). In Step 2 (Fig. 1), the frequency of each ion signal is determined and analyzed independently. At this processing stage, precise information for each ion including frequency, intensity, and m/z values is established. Step 3 (Fig. 1) determines the ion's signal strength using a data plotting and analysis process that assesses the current induced by an ion on the Orbitrap detection electrodes as a function of acquisition time. For simplicity we call this signal strength determination the "STORI" process¹⁰ with the slope of a STORI plot being proportional to the charge of the ion (Step 3, Fig. 1). In Step 4, the charge of the ion is determined by a slope-to-charge calibration function. The STORI slope of an ion with unknown charge is assigned the closest integer charge state on the calibration function, determined just once for each instrument (two different Q Exactive-style instruments were used in this study). Using the integer charge (z) and mass-to-charge ratio (m/z), it is possible to determine the mass of each ion and produce a spectrum in the true *mass* domain with quite different spectral properties and increased resolution via centroiding and binning single ion signals⁸. As a result of our charge assignment algorithm, further explained in the Online Methods, charge misassignments do occur at the 1–4% level of the total ion count in I²MS spectra; these artifacts are recognized readily as satellite peaks due to the ± 1 error in charge state assignment. Further experimental optimization and software development is currently underway to completely eliminate charge misassignment. The I²MS process further uses validated raw data reduction and a reduced DC offset for the Orbitrap center electrode (–5 to –1 kilovolt) enabling greater than 70% of the ions to survive the ~2–4 second detection period (Supplemental Fig. 1). A more in-depth description of these steps is available in the Online Methods. We now focus on the application of the method to complex proteoform mixtures and detection of large complexes via native I²MS.

Initial experiments on a mixture of intact protein standards from 8–47 kDa highlighted an immediately apparent set of differences in I²MS readout from conventional ensemble mass

spectrometry analysis. Basic figures of merit for the standards are shown in Supplemental Fig. 2c, illustrating high resolving power for intact proteins and the accurate determination of their monoisotopic mass. Note that I²MS produces simple spectra with an x-axis of mass, removing the normal requirement for the inference step to convert m/z to mass. The remarkably clean baselines of I²MS spectra proved highly insensitive to partially desolvated ions (Supplemental Fig. 2a vs. 2b), which extended the dynamic range for intact mass determination even when complexity was high (*vide infra*). The strict filtering criteria during STORI slope determination removed non-stable ion signals stemming from desolvation, ion fragmentation in the Orbitrap analyzer (such as neutral loss), and intermittent signals produced from electronics (electronic noise). In essence, I²MS transforms the instrument into an actual MASS spectrometer, removing the normal requirement for the inference step in conventional mass spectrometry.

We then used I²MS to analyze a mixture of heavy and light chains resulting from disulfide reduction of an IgG₁ (Supplemental Fig. 3). Here, the removal of non-stable ion signals yielded highly sensitive I²MS spectra of species and adduct identification that was not possible in ensemble measurements. Ultimately, these mass domain readouts greatly simplify data analysis for non-experts.

To stress-test the I²MS process, it was compared to standard direct infusion with automatic gain control (AGC) set to 1×10^6 charges/spectrum for high complexity samples like the entire <30 kDa range of the human proteome from HEK293 T-Cells fractionated by GELFrEE technology¹¹. A silver-stained SDS-PAGE gel (inset of Fig. 2a) shows the range of proteins that were analyzed by both standard MS and I²MS approaches (Fig. 2a vs. 2b). Although no protein masses could be determined using standard MS (Fig. 2a), I²MS (Fig. 2b) detected over 500 proteoform masses that were assigned using the intact mass tag approach^{12, 13} and referencing prior top-down proteomics data where tandem MS (LC-MS/MS) was used to create a list of high quality proteoforms present in the Human Proteoform Atlas¹⁴. A comparable number of ions (~10 million) was utilized to generate the ensemble and I²MS spectra. While all identified proteoforms via I²MS analysis are listed in Supplemental Dataset 1, small insets in Fig. 2b illustrate 17 identified proteoforms that are listed in Supplemental Table 1. In addition, isotopic distributions for higher mass proteoforms between 20 – 25 kDa illustrated in the red box on Fig. 2b have not previously been identified and characterized by prior LC-MS/MS analysis. For conventional mass spectrometry, larger proteins (>20 kDa) are burdened with charge state distributions consisting of dozens of visible charge states. Appropriating a fixed number of ions between this many charge states lowers the S:N of each one. Consolidating the many low S:N charge states into one higher S:N mass distribution greatly increases the probability for their detection, annotation and characterization. I²MS, with the high intra-spectral dynamic range of ~2,000 clearly identifies these high mass features usually not detectable in lower dynamic range traditional m/z analysis. I²MS reduces charge state ambiguity and consolidates each charge state distribution into one mass channel, increasing detection capabilities for larger mass species in mixtures with extreme complexity.

Another major advantage of I²MS is the ability to work with samples containing large protein complexes that are lifted into the gas phase using native electrospray ionization (so

called native MS)¹⁵. Pressing I²MS further, we examined standard protein complexes^{16, 17} and also applied it to engineered virus-like particles (VLP). As demonstrated by Supplemental Fig. 4, protein complexes including pyruvate kinase (Supplemental Fig. 4a, b, and c) and GroEL (Supplemental Fig. 4d, e, and f) with well resolved charge states validate I²MS implementation for native complexes. The charge state distributions from I²MS match those determined using standard MS, and the relative intensity of the charge assignments mirror their corresponding spectra in the *m/z* domain. Accurately assigned individual ions from resolved charge states produce well-defined peaks in the mass domain for pyruvate kinase (231.9 kDa) and GroEL (801.2 kDa). These higher mass assemblies are accompanied with spectral resolution values of 560 – 800. Further, minor species of the pyruvate kinase complex formed from loss of a C-terminal lysine in one or two of the subunits within the tetramer (Supplemental Fig. 4b) are easily recognized.

The single-ion analysis advanced here was conducted on two virus-like particles (VLPs) engineered from viral capsids carrying varying amounts of DNA and mRNA cargo loads after heterologous production in *E. coli*. Wild-type (WT) or mutant (MINI) versions of these VLPs were constructed from the MS2 wild-type coat protein or a Ser37Pro mutant, respectively. Recently, these particles have been investigated through electron microscopy and X-ray crystallography to determine capsid structure^{18–20}. Briefly, WT and MINI VLPs are composed of exactly 180 and 60 copies of the 13.7 kDa coat protein (CP), generating theoretical masses of 2.47 and 0.82 MDa, and particles with 27 and 17 nm diameters. In turn, cavity volumes can be calculated around 10,300 and 2,600 nm³, assuming an approximate spherical nature of the VLPs. The WT and MINI monomer amino acid composition was verified through MS analysis in denatured mode (Supplemental Fig. 5). Although the structure of the VLPs is established, characterization of cargo such as therapeutic proteins and RNA would be highly valuable to quickly determine the stoichiometry and loading of drug-delivery within the particles.

To probe composition and cargo of the VLPs, I²MS data for the WT and MINI particles were compared to each other and contrasted with data from standard MS data produced in the *m/z* domain (Fig. 3c,d and Fig. 3e,f). No visible charge states were present in the *m/z* domain spectra of either the WT (Fig. 3c) or MINI (Fig. 3d) species. The lack of distinct charge states in the native MS data was attributed to the molecular heterogeneity of varying lengths of DNA and mRNA¹⁹. Given this molecular heterogeneity and no distinct charge state peaks present in the *m/z*-domain spectra, no information regarding mass of the VLP plus cargo could be discerned. This is an often-encountered situation in obtaining size distributions of highly complex samples in mass spectrometry and the reason other methods like gel electrophoresis and light scattering are used to provide broad molecular weight estimations.

In contrast to standard native MS, the I²MS spectra produced mass distributions of the WT (Fig. 3e) and MINI (Fig. 3f) species with masses of 3190 ±38 kDa and 990 ±16 kDa (mean ± s.d.). Assigned charge states for the set of single ions were centered around +137 and +67 for WT (Fig. 3g) and MINI species (Fig. 3h) with the expected Gaussian-like distributions. In addition, consolidating charge states in the creation of the VLP I²MS mass spectra increased resolution greater than 2-fold when compared to their *m/z* spectra counterparts.

Determined I²MS masses minus proposed capsid masses reveal WT and MINI cargo masses of approximately 0.72 ± 0.038 MDa and 0.17 ± 0.016 MDa, respectively. Although the length of the DNA base pairs and mRNA nucleotides that form the VLP cargo loads vary across a wide distribution of lengths, MINI VLPs having a much smaller volume incorporate shorter DNA and mRNA strands than their WT counterparts. DNA and mRNA lengths averaged 900 (± 400) vs. 375 (± 150) base pairs (bp) and 500 (± 250) vs. 100 (± 50) nucleotides (nt) for WT vs. MINI species¹⁹. The mass distribution for the WT particles determined via I²MS (Fig. 3e) is consistent with the encapsulation of 1 average length DNA (900 bp) and mRNA (500 nt) strand together in the VLP (~ 0.74 MDa). The WT mass distribution with a FWHM of only 75 kDa means that as a larger DNA strand is added as cargo, it is partnered with a smaller length mRNA component during production and self-assembly within the cytosol of *E. coli*. The mass distribution for the MINI particles determined via I²MS (Fig. 3f) is consistent with VLP encapsulation of either 1 DNA strand or multiple smaller mRNA strands as cargo. The incorporation of the average length DNA strand (375 bp) matches the higher mass tailing feature around 1.06 MDa. In turn, the smaller VLP volume can contain multiple (~ 5) of the smaller ~ 100 nt mRNA strands, accounting for the majority of the MINI mass distribution. The ratio, 4.2, of the cargo masses ($\frac{MS_{WT}}{MS_{MINI}}$) corresponds to the ratio of cavity volumes, (4.0) further supporting assignment of cargo content based on I²MS data for these spherical WT and MINI VLPs. The measurement of complex masses through I²MS, unobtainable via traditional mass spectrometry, reveals pertinent information to their composition and cargo. As a result, I²MS analysis is highly applicable to VLP engineering, can be extended to other promising systems including Q β virions, adeno-associated virus (AAV), and hepatitis B, and can yield important information for more complex systems including bacterial microcompartments which contain an assortment of coat protein sizes.

With increased dynamic range, resolution, clean baselines and intuitive data interpretation, I²MS shows increased value for determination of mass values particularly as mixture complexity increases. Being an Orbitrap-based method, I²MS provides a readily deployable technique for single molecular MS of denatured proteins and native complexes with masses ranging from 5 kDa to 3.5 MDa. The ability to accurately determine mass values directly from dilute and complex samples regardless of charge state overlap will greatly inform and expand the utility of mass spectrometric analysis of proteins and their assemblies in improved characterization of molecular complexity for both endogenous and synthetic biology.

Online Methods

Cell growth, lysis, and protein fractionation

Human embryonic kidney cells HEK-293 (ATCC CRL-1573) as reported within the Life Sciences Reporting Summary were grown on 15 cm plates in DMEM media (Life Technologies) supplemented with 10% fetal bovine serum (Life Technologies), containing 1% Penicillin/Streptavidin (Life Technologies), and they were incubated at 37°C temperature under 5% CO₂ atmosphere. Triplicate of confluent plates containing 1×10^7 cells were treated with 100 nM rapamycin or a DMSO control for 8 hours and harvested

with 0.05% trypsin (Life Technologies) for 5 min. For all I²MS experiments reported here, cells were grown at control conditions. Cells were pelleted at 300 x g for 10 min. and washed twice with PBS (Life Technologies). Cells were incubated with lysis buffer (1% w/v N-lauroylsarcosine, 20 mM Tris Base pH 7.5, 100 mM NaCl, and HALT Inhibitor Cocktail (Thermo Fisher Scientific) for 10 min. on ice, 750 units of Benzonase nuclease (Sigma) was added, the solution was incubated for 15 min. at 37°C, and the protein content was estimated using Pierce BCA assay (Thermo Fisher Scientific). Aliquots of 400 µg of protein were precipitated for 30 min. at -80°C with 4 volumes of cold acetone, resuspended in GELFrEE Tris-Acetate Sample Buffer (Expedeon), and proteins were fractionated on a GELFrEE 8100 (Expedeon) system using an 8% cartridge according to manufacture procedures. Only the first fraction containing proteins up to 30 kDa was collected.

Denatured mode top-down proteomics using LC-MS/MS

Fractions were methanol/chloroform/water precipitated¹¹, resuspended in buffer A (5% acetonitrile, 94.8% water, and 0.2% formic acid), and submitted to LC-MS/MS. Samples were fractionated on a Dionex UltiMate 3000 LC system (Thermo Fisher Scientific) using an in-house packed PLRP-S (5 µm particle, 1000 Å pore size) (Agilent) column (75 µm ID × 25 cm length) after loading on a PLRP-S trap column (150 µm ID × 2 cm). The loading pump was operated at 3 µL/min., and the nano pump gradient flow rate was set at 300 nL/min. Buffer A and buffer B (5% water, 94.8% acetonitrile, and 0.2% formic acid) were used for LC separation under the following gradient: 5% B at 0 min., 15% B at 5 min., 55% B at 55 min., 95% B from 58 to 61 min., 5% B from 64 to 80 min. Column was kept at constant temperature (55°C), the outlet was online with a 15 µm i.d. electrospray emitter (New Objective) packed with 3 mm of PLRP-S resin, and attached to a nanoelectrospray ionization source built in-house. LC was online with an Orbitrap Fusion Lumos (Thermo Fisher Scientific) mass spectrometer operating in “protein mode” with 2 mTorr of N₂ pressure. Transfer capillary temperature was set at 320 °C, ion funnel RF was set at 30%, and a 15V of source CID was applied. MS¹ spectra were acquired at 120,000 resolving power (at 200 m/z), AGC target value of 500,000 charges/acquisition, 200 msec. of maximum injection time, and 4 µscans. Data-dependent top two MS² method using 23 NCE for higher-energy collisional dissociation (HCD) was used to generate MS² spectra that were acquired at 60,000 resolving power @ 200 m/z, with target AGC values of 500,000 charges/acquisition, 800 msec. of maximum injection time, and 4 µscans. Precursors were quadrupole isolated using a 3 m/z isolation window, dynamic exclusion of 60 sec. duration, and threshold of 2×10⁴ intensity.

Protein Identification

Raw files from LC-MS/MS experiments were searched against a UniProt *Homo sapiens* (Taxon ID: 9606, Proteome ID: UP000005640) protein database using TDPportal (<http://galaxy.kelleher.northwestern.edu/>), as previously described^{21, 22}. Proteoform search space was generated *in silico* allowing up to 11 PTMs or sequence variations per candidate proteoform. Two different search types were performed in parallel: Absolute Mass search with precursor tolerance of 2.2 Da and fragments tolerance of 10 ppm, and a Biomarker search with 10 ppm precursor and fragment tolerance. Estimation of false discovery rate

(FDR) at the protein entry and proteoform level was performed²³. Proteoforms above a 10% FDR threshold were reported and used for GELFrEE fraction I²MS spectrum annotation.

Proteoforms passing the FDR cutoff were transformed into theoretical +1 (M+H) isotopic distributions²³. The I²MS mass spectrum was also converted into singly-charged data for comparison. With the identified proteoforms serving as theoretical distributions, the I²MS distributions were matched against the theoretical using an isotope fitting routine with a 25 ppm tolerance^{24, 25}. Matches were loaded into TDValidator for visualization and manual validation (Proteinaceous, Inc.).

Virus-like Particle Expression and Purification

Glycerol stocks of *E. coli* DH10B cells containing plasmids encoding for either the MS2 wild-type coat protein or the mutant (mini) variant were grown in 10 mL 2xYT overnight at 37°C. These cells were then subcultured into 1 L 2xYT to a starting OD600 of 0.05 and grown with shaking at 37°C. When the OD600 reached 0.5, protein expression was induced with 0.1% arabinose. Expression continued overnight for ~ 16 h. After harvesting the cells by centrifugation, the cellular protein was extracted and VLPs were purified according to standard procedures using ammonium sulfate precipitation and fast protein liquid chromatography (FPLC)²⁰.

Additional Sample Preparation and MS Analysis

To produce the 4 protein mixture (Supplemental Fig. 2) ubiquitin (bovine; ID: P0CH28), myoglobin (equine heart; ID: P68082), carbonic anhydrase (bovine; ID: P00921), and enolase (baker's yeast; ID: P00924) purchased from Sigma Aldrich were added together to produce a 2 mL solution with final protein concentrations of 20, 40, 50, and 90 nM, respectively. To produce the reduced version of the NIST antibody standard (Supplemental Fig. 3) 30 µL (300 µg) of stock solution (NIST) was mixed with 300 µL of 6 M GdHCl. 30 mM TCEP-HCl was added, the solution was incubated at 37°C for 1.5 h, and finally diluted to a 500 nM concentration. HEK293 GELFrEE fraction 1 (Fig. 1) was methanol/chloroform/water precipitated¹¹, resuspended in HPLC grade water and run through a HiPPR (Thermo Fisher Scientific) detergent removal column, following the manufacture's protocol. The fraction protein concentration was determined with a BSA protocol (Thermo Fisher Scientific) and diluted to a 5 nM concentration. Before dilution all samples were desalted six times at 10,000 x g for 5 min. in 3 kDa Amicon Ultra centrifugal filters (Merck Milipore) with 100 mM ammonium acetate. Once desalted, all samples were diluted and electrosprayed under denaturing conditions in a 40% acetonitrile and 0.2% acetic acid solution.

Pyruvate kinase and GroEL protein complexes (Supplemental Fig. 4) were purchased from Sigma Aldrich and diluted to final concentrations of 1 µM. GroEL was precipitated as previously described²⁶. Both protein complexes were desalted six times at 10,000 x g for 4 min. in 100 kDa Amicon Ultra centrifugal filters (Merck Milipore) with 100 mM ammonium acetate. Once desalted, both samples were diluted and electrosprayed under native conditions in 100 mM ammonium acetate.

Denatured samples with molecular weights <100 kDa were introduced into a Q Exactive Plus (Thermo Fisher Scientific) mass spectrometer with a custom nano-electrospray source as described previously²⁷ using +0.8 to 1.6 kV spray voltage. Due to significant ion decay rates observed for denatured proteins with molecular weights between 8–150 kDa, the Q Exactive Plus mass spectrometer was modified to reduce the central electrode voltage of the Orbitrap to –1 kV during ion detection. Samples >200 kDa ionized with native-mode electrospray were introduced into a Q Exactive Ultra High Mass Range (UHMR) (Thermo Fisher Scientific) mass spectrometer with a Nanospray Flex Ion Source (Thermo Fisher Scientific) with spray voltages between +1.4 to 2 kV. As large native complexes/VLPs are large relative to the size of the background gas molecules, collisions are ineffective at either reducing the momentum of the species or transferring enough internal energy to cause dissociation. As a result, less decay for these types of ions were observed within the Orbitrap analyzer and thus the central electrode voltage for the Q Exactive UHMR was kept at –5 kV for ion detection. For both instruments, acquisition rates were reduced down to 0.5 spectra/sec, corresponding to a 2 second detection period per acquisition event and the HCD-pressure was set between 0 to 0.5 (UHV Pressure < 5×10^{-11} torr) to reduce ion and background-gas collisions from a more typical setting of 1 (UHV Pressure $\sim 9 \times 10^{-11}$ torr). Enhanced Fourier-Transform (eFT) was turned off, as it was not necessary to process single ion signals in this manner. A source temperature of 320°C, an in-source collision-induced dissociation (SID) value of 0 to 15 V, and an in-source trapping (UHMR) value of –50 to –150 V was optimized on a per species basis.

Ion Collection and Data Acquisition of the I²MS Process

The overall scheme describing I²MS is shown in Fig. 1. Below is an extended technical description of this new process.

Ion Collection (Step 1)

Similar to traditional CDMS techniques, I²MS evaluates single ion signals to produce a true mass spectrum. However, Orbitrap CDMS analysis differs from traditional linear ion trap CDMS in the efficiency of single ion collection^{2–5}. With the implementation of image current detectors our single ion definition is not confined to one ion per acquisition event, but one ion at a defined frequency per acquisition, meaning that multiple ions are analyzed per acquisition as long as they do not correspond to multiple ions at the same m/z value. To lower the number of ions entering the Orbitrap analyzer, denatured samples were diluted to concentrations between 5–50 nM, while samples of large complexes were analyzed by native electrospray in the 1–3 μ M regime. In addition, Automatic Gain Control (AGC) was disabled and ion injection times were set between 0.03 – 1 ms to manually attenuate signal to minimize multi-ion events and maximize the collection individual ions. As this is a random style trapping event, the exact number of individual ion signals collected varies slightly for each acquisition event. In some cases, the ion optics including the C-trap entrance lens were detuned to further reduce the number of ions entering the Orbitrap analyzer.

Individual Ion STORI Slope and Centroid m/z Determination (Fig. 1, Steps 2 and 3)

Within each acquisition event, every detected ion was analyzed individually. Both the mass-to-charge ratio (m/z) and charge (z) were necessary to determine the mass of the ion in question. In FT-MS Orbitrap analysis, m/z was determined from the frequency of ion rotation around the central electrode. To determine the centroid m/z value, the apex of the profile peak was determined by a quadratic fit to the three most intense m/z points. The z was measured from the rate of the induced charge on the Orbitrap outer electrode, otherwise known as Selective Temporal Overview of Resonant Ions (STORI) as previously described¹⁰. Briefly, STORI analysis was the integration of the ion induced charge over the course of the time domain acquisition period at the specified single ion frequency value. The slope of the STORI plot remained constant until either the end of the acquisition event or premature ion loss within the Orbitrap analyzer. The STORI plot slope of a particular ion was proportional to the charge of the observed ion. Supplementary Fig. 6 demonstrates the dependence depicting increasing STORI slopes for the +13 charge state of ubiquitin, the +25 charge state of myoglobin, and the +39 charge state of carbonic anhydrase. STORI slope is only dependent on the charge of the ion and not the m/z or mass of the species in question. The charge detection limit for I^2MS analysis is such that ions only with a charge of +6 or greater can be measured. Ion signals with a charge state below that threshold cannot be distinguished from electronic noise. Electronic noise in FT decreases as a function of $\frac{1}{\sqrt{\text{time}}}$ and longer acquisition time or hardware modifications will be necessary to further decrease this charge detection limit.

The variability in STORI slope decreases proportionately to the $\sqrt{\text{Ion Survival Time}}$. To accurately determine the charge of an ion from its STORI slope, long acquisition times are necessary for which the ion signal must persist (ideally for the full 2 s detection period). Many factors contribute to how long an ion signal persists. A few include the amount of background gas that enters the Ultra-high vacuum (UHV) region of the mass spectrometer, velocity of the ions, length of the acquisition, and finally the collisional cross-section of the ion. Although we have successfully increased ion signal persistence through the reduction of ion kinetic energy (decreasing ion velocity) and limiting the amount of background gas entering the UHV region of the mass spectrometer, reference 8 gives a brief overview of how collisional cross-section can impact individual ion decay for denatured protein ions.

Charge Determination and Slope Averaging Technique (Fig. 1, Step 4)

In order to accurately assign the unknown z of an ion with a known STORI slope value, each ion slope was calibrated to a charge calibration function. Supplemental Fig. 7 illustrates 2 linear calibration functions created from isolated ions with known charge states analyzed with a Q Exactive Plus (red points/trace) and Q Exactive UHMR (blue points/trace). To produce the trace for the Q Exactive Plus various known charge states from denatured ubiquitin, myoglobin, carbonic anhydrase, and enolase were analyzed. To produce the trace for the Q Exactive UHMR various known charge states from natively electrosprayed carbonic anhydrase, monoclonal NIST antibody standard, pyruvate kinase, and groEL were analyzed. Thousands of single ions from each known charge state were analyzed and their median STORI slope value was calculated (represented by each black point). Each

calibration file was manually curated to ensure the STORI slopes utilized to calculate the average value did not include slopes from two ion signals at the same m/z value or slopes from contaminating species that overlap in m/z space. The points on the calibration were fit with a linear regression yielding the equation below with a strong statistical relationship ($R^2 = 0.9997$):

$$z = (\text{slope} + 2317) / 52622$$

It is good to note that the charge calibration curve was specific to each instrument. Small differences in ion trajectories into the Orbitrap and differing imperfections of the Orbitrap analyzer for each instrument affect the slope of the calibration function. However, once the calibration was produced, the instrument required no recalibration as long as the injection energy of the ions from the C-Trap into the Orbitrap analyzer was not changed. To support this conclusion, testing on the Q Exactive UHMR over a 3-month period revealed, on average, only a 0.1% drift in the median charge state assigned using initial STORI plots. This negligible shift in the calibration points over time demonstrates the longitudinally robust nature of the detection process.

To determine the quantized z with a >96% rate of correct assignment, the ion STORI slope was scored based on proximity to the nearest median charge calibration value and assigned. Without any statistical processing methods, current instrumentation limits the correct charge assignment of ions to approximately 50%. The other half of ion signals are mis-assigned to a charge state that either high or low by one integer value. Supplemental Fig. 8 shows the charge assignment of approximately 800 single ion signals for the isolated myoglobin +20 charge state collected over a 2 second transient. Although all the ion signals collected were known to correspond to a +20 charge, slight variations in their STORI slopes resulted in the incorrect charge assignment of 47.5% of the ion signals collected, with the incorrect charge assignments having a Gaussian-like distribution to higher and lower charge states. A slope averaging mechanism, more commonly known as the central limit theorem, was implemented to correct the misassignments.

When STORI slope values from a known charge state were averaged, there was a reduction in charge misassignment proportional to $1/\sqrt{n}$, where n is equal to the number of ion STORI slope values being averaged. Supplemental Fig. 8 illustrates the implementation of this averaging mechanism, where the +20 charge assignment for the myoglobin species improved to 81.1% and 96% when averaging 4 and 16 ion STORI slope values, respectively, and subsequently assigning charge. This averaging of 16 STORI slopes was cemented into the algorithm for data acquisition utilized in the creation of all denatured I²MS spectra in this report to accurately assign charge to individual analyte ions from 6 – 100 kDa carrying +6 to +80 charge states in the 500 – 5,000 m/z range. Averaged STORI slope counts increased up to 100 to assign >96% of individual ions to the correct charge state for native species in the 10,000 – 23,000 m/z range. Before averaging STORI slope values, first all ions collected were ordered as a function of their m/z values. A slope tolerance for STORI slope values 30% above or below the median slope value of the points being averaged in addition to a maximum 5 m/z window for ordered STORI slopes to be averaged provided

strict guidelines to remove outlier values. STORI slope values are averaged pre-charge calibration and the resulting calculated z was parsed back to all input single ions. Once an accurate quantized charge was calculated for each single ion the mass was determined:

$$Mass = \left(\frac{m}{z} * z\right) - (z * M_{proton})$$

Extensive testing revealed that for the Orbitrap detector, with a pre-amplifier lacking a high-pass filter, there was no deviation from the linearity of the linear function that converts STORI slope to charge for either the Q Exactive Plus or Q Exactive UHMR for ions with differing m/z or mass values. This coupled with the high correlation of the calibration functions $R^2=0.9997$ and $R^2=0.9999$, respectively, showed that extrapolation to higher non-surveyed charge states for VLP samples was reliably accomplished. To further validate this assertion, the collection of 2 GroEL +70 ion signals at the same frequency value within one acquisition, accurately yields a STORI slope that corresponds to a single +140 (+70 \times 2) ion signal.

Plotting of I²MS Spectra (Fig. 1, Step 5)

Once the mass of each ion was calculated, the thousands of acquired individual ions were binned to create the I²MS spectrum. Histogram bin sizes varied based on the species analyzed. Masses under 200 kDa ions were binned in 0.2 Da increments. As masses increased beyond 200 and 1,000 kDa and resolution decreased, bin sizes were increased to 30 and 500 Da, respectively. In tandem, the calculated charge of each ion utilized to create the I²MS spectrum was binned in quantized domains to validate charge assignment. A simple protein example demonstrating these outputs is shown in Supplemental Fig. 9. Supplemental Fig. 9a,b,c represent the traditional m/z mass spectrum of myoglobin, corresponding to the I²MS spectrum in the mass domain, and single ion charge assignment histogram for ions utilized to create the I²MS spectrum, respectively. Supplemental Fig. 9a mirrors the charge state distribution in Supplemental Fig. 9c.

Statistical Analyses

Graph visualizations and calculations were performed using Thermo Xcalibur Qual Browser software, Microsoft Excel 2016 software, or custom compiled C# code. Although experiments were generally performed twice, thousands of individual scan events were utilized to produce I²MS spectra with statistically meaningful mass distributions. These experimental mass distributions were then verified with previous datasets or matched to theoretical mass distributions. Furthermore, the method as a whole was anchored to a set of standards to validate I²MS as an accurate and reproducible platform.

Data Availability

Datasets utilized for the I²MS analyses can be found on the MassIVE repository, MSV000083840 doi:[10.25345/C5XD01](https://doi.org/10.25345/C5XD01). Other data that support the findings of this study are available from the corresponding authors upon request.

Code Availability

Custom code associated with the I²MS creation process is available from the corresponding authors upon reasonable request.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

This work was funded by the Intensifying Innovation program from Thermo Fisher Scientific and was carried out in collaboration with the National Resource for Translational and Developmental Proteomics under Grant P41 GM108569 from the National Institute of General Medical Sciences, National Institutes of Health with additional support from the Sherman Fairchild Foundation, and the instrumentation award (S10OD025194) from NIH Office of Director. In addition, we would like to thank Luca Fornelli and Timothy K. Toby for collecting and analyzing the HEK-293 LC/MS runs utilized for our intact mass tag I²MS analysis.

References

1. Aebersold R & Mann M Mass spectrometry-based proteomics. *Nature* 422, 198–207 (2003). [PubMed: 12634793]
2. Elliott AG, Harper CC, Lin H-W & Williams ER Mass, mobility and MSn measurements of single ions using charge detection mass spectrometry. *Analyst* 142, 2760–2769 (2017). [PubMed: 28636005]
3. Keifer DZ, Pierson EE & Jarrold MF Charge detection mass spectrometry: weighing heavier things. *Analyst* 142, 1654–1671 (2017). [PubMed: 28443838]
4. Benner WH A Gated Electrostatic Ion Trap To Repetitiously Measure the Charge and m/z of Large Electrospray Ions. *Analytical Chemistry* 69, 4162–4168 (1997).
5. Schmidt HT, Cederquist H, Jensen J & Fardi A Conetrap: A compact electrostatic ion trap. *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms* 173, 523–527 (2001).
6. Makarov A & Denisov E Dynamics of Ions of Intact Proteins in the Orbitrap Mass Analyzer. *Journal of the American Society for Mass Spectrometry* 20, 1486–1495 (2009). [PubMed: 19427230]
7. Rose RJ, Damoc E, Denisov E, Makarov A & Heck AJR High-sensitivity Orbitrap mass analysis of intact macromolecular assemblies. *Nature Methods* 9, 1084 (2012). [PubMed: 23064518]
8. Kafader JO et al. Measurement of Individual Ions Sharply Increases the Resolution of Orbitrap Mass Spectra of Proteins. *Analytical Chemistry* 91, 2776–2783 (2019). [PubMed: 30609364]
9. Contino NC & Jarrold MF Charge detection mass spectrometry for single ions with a limit of detection of 30 charges. *International Journal of Mass Spectrometry* 345–347, 153–159 (2013).
10. Kafader JO et al. STORI Plots Enable Accurate Tracking of Individual Ion Signals. *Journal of The American Society for Mass Spectrometry* (2019).
11. Tran JC et al. Mapping intact protein isoforms in discovery mode using top-down proteomics. *Nature* 480, 254 (2011). [PubMed: 22037311]
12. Gómez SM, Nishio JN, Faull KF & Whitelegge JP The Chloroplast Grana Proteome Defined by Intact Mass Measurements from Liquid Chromatography Mass Spectrometry. *Mol. Cell Proteomics* 1, 46–59 (2002). [PubMed: 12096140]
13. Smith RD et al. An accurate mass tag strategy for quantitative and high-throughput proteome measurements. *Proteomics* 2, 513–523 (2002). [PubMed: 11987125]
14. The Human Protein Atlas <https://www.proteinatlas.org> (2019).
15. Lermyte F, Tsybin YO, O'Connor PB & Loo JA Top or Middle? Up or Down? Toward a Standard Lexicon for Protein Top-Down and Allied Mass Spectrometry Approaches. *Journal of The American Society for Mass Spectrometry* 30, 1149–1157 (2019). [PubMed: 31073892]

16. Schachner LF et al. Standard Proteoforms and Their Complexes for Native Mass Spectrometry. *Journal of The American Society for Mass Spectrometry* 30, 1190–1198 (2019). [PubMed: 30963455]
17. Sobott F & Robinson CV Characterising electrosprayed biomolecules using tandem-MS—the noncovalent GroEL chaperonin assembly. *International Journal of Mass Spectrometry* 236, 25–32 (2004).
18. Valegård K, Liljas L, Fridborg K & Unge T The three-dimensional structure of the bacterial virus MS2. *Nature* 345, 36–41 (1990). [PubMed: 2330049]
19. Asensio MA et al. A Selection for Assembly Reveals That a Single Amino Acid Mutant of the Bacteriophage MS2 Coat Protein Forms a Smaller Virus-like Particle. *Nano Letters* 16, 5944–5950 (2016). [PubMed: 27549001]
20. Hartman EC et al. Quantitative characterization of all single amino acid variants of a viral capsid-based drug delivery vehicle. *Nature communications* 9, 1385–1385 (2018).
21. Fornelli L et al. Advancing Top-down Analysis of the Human Proteome Using a Benchtop Quadrupole-Orbitrap Mass Spectrometer. *J Proteome Res* 16, 609–618 (2017). [PubMed: 28152595]
22. Anderson LC et al. Identification and Characterization of Human Proteoforms by Top-Down LC-21 Tesla FT-ICR Mass Spectrometry. *J Proteome Res* 16, 1087–1096 (2017). [PubMed: 27936753]
23. LeDuc RD et al. Accurate Estimation of Context-Dependent False Discovery Rates in Top-Down Proteomics. *Mol. Cell Proteomics* 18, 796–805 (2019). [PubMed: 30647073]
24. Ntai I et al. Applying Label-Free Quantitation to Top Down Proteomics. *Analytical Chemistry* 86, 4961–4968 (2014). [PubMed: 24807621]
25. Durbin KR, Skinner OS, Fellers RT & Kelleher NL Analyzing Internal Fragmentation of Electrosprayed Ubiquitin Ions During Beam-Type Collisional Dissociation. *Journal of The American Society for Mass Spectrometry* 26, 782–787 (2015). [PubMed: 25716753]
26. Freeke J, Robinson CV & Ruotolo BT Residual counter ions can stabilise a large protein complex in the gas phase. *International Journal of Mass Spectrometry* 298, 91–98 (2010).
27. Skinner OS et al. An informatic framework for decoding protein complexes by top-down mass spectrometry. *Nature Methods* 13, 237 (2016). [PubMed: 26780093]

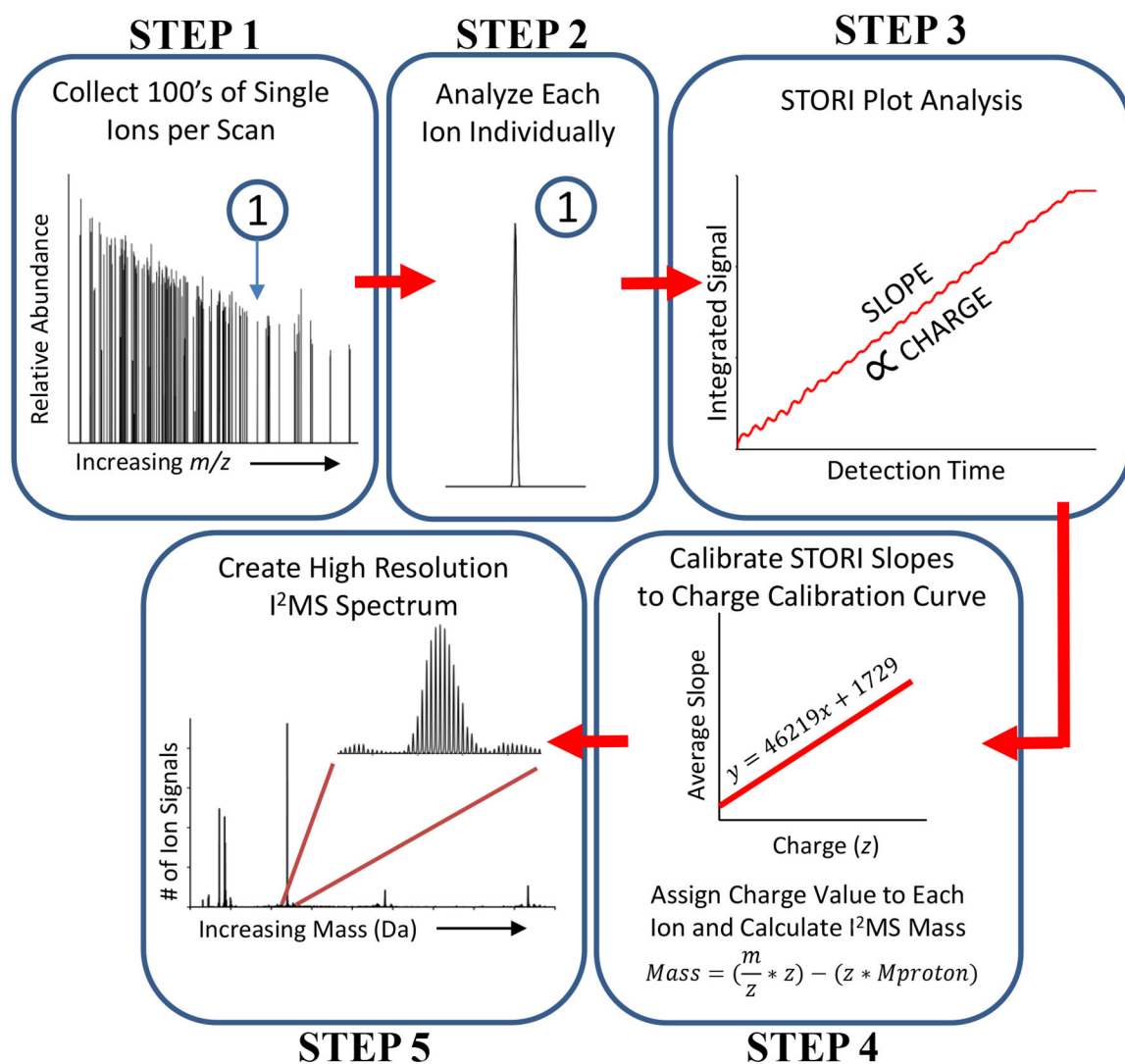


Figure 1. The five step workflow for the I²MS process.

Includes data acquisition and post-acquisition processing steps that culminate in the plotting of an I²MS spectrum (Step 5). A detailed explanation of each of these five steps is found in the accompanying Online Methods section.

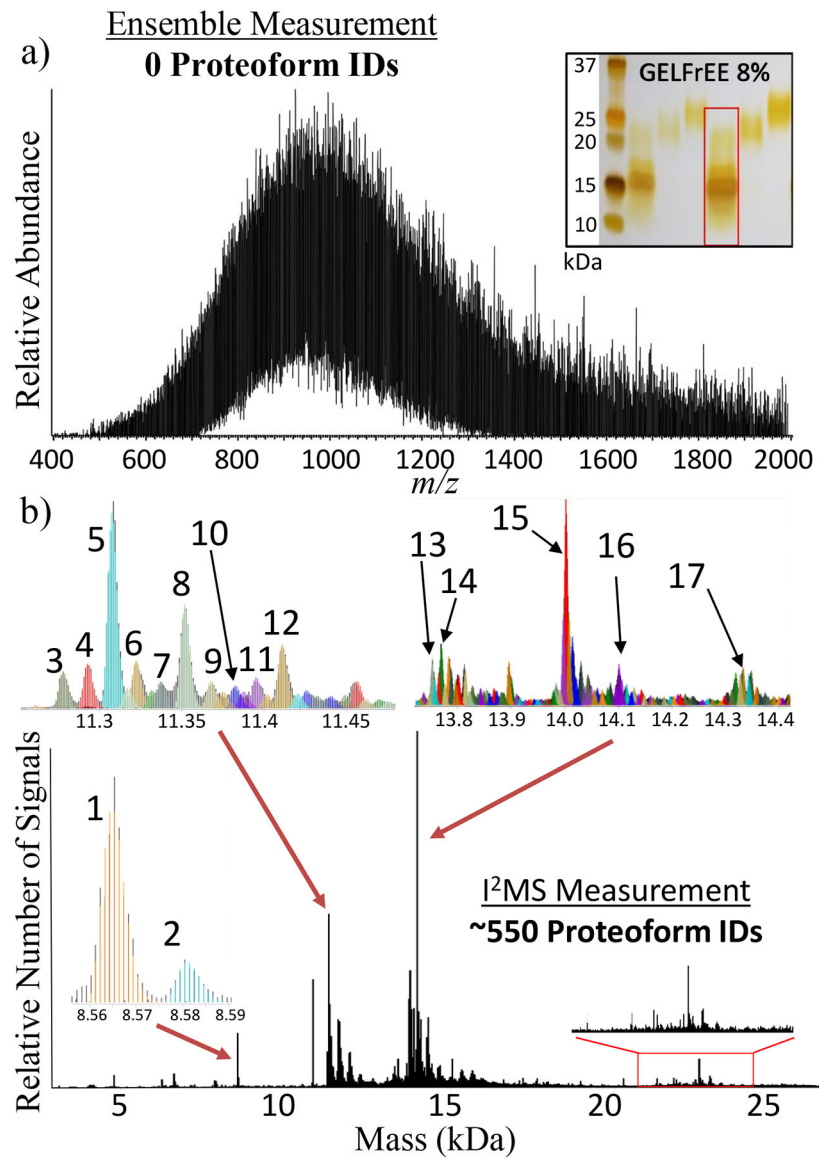


Figure 2. Deconvolution of Complex Proteoform Mixtures.

Comparison of the number of proteoform assignments possible using either conventional spectral acquisition with an ensemble of ions in a population (a) or the I²MS process (b). This comparative analysis involved directly infusing a mixture of 0 – 30 kDa proteins created by fractionation of whole extracts from HEK-293 human cells. The inset in panel (a) shows an analytical, silver-stained gel, with the fraction analyzed highlighted with a red rectangle. The 3 insets in panel (b) highlight 17 of the 550 total proteoforms assigned, which are listed in Supplemental Table 1 and include their proteoform record numbers (PFRs). The 20–25 kDa region highlighted with a red rectangle in panel (b) corresponds to proteoforms of higher mass proteins in the complex mixture previously unidentified using LC-MS/MS for top-down proteomics.

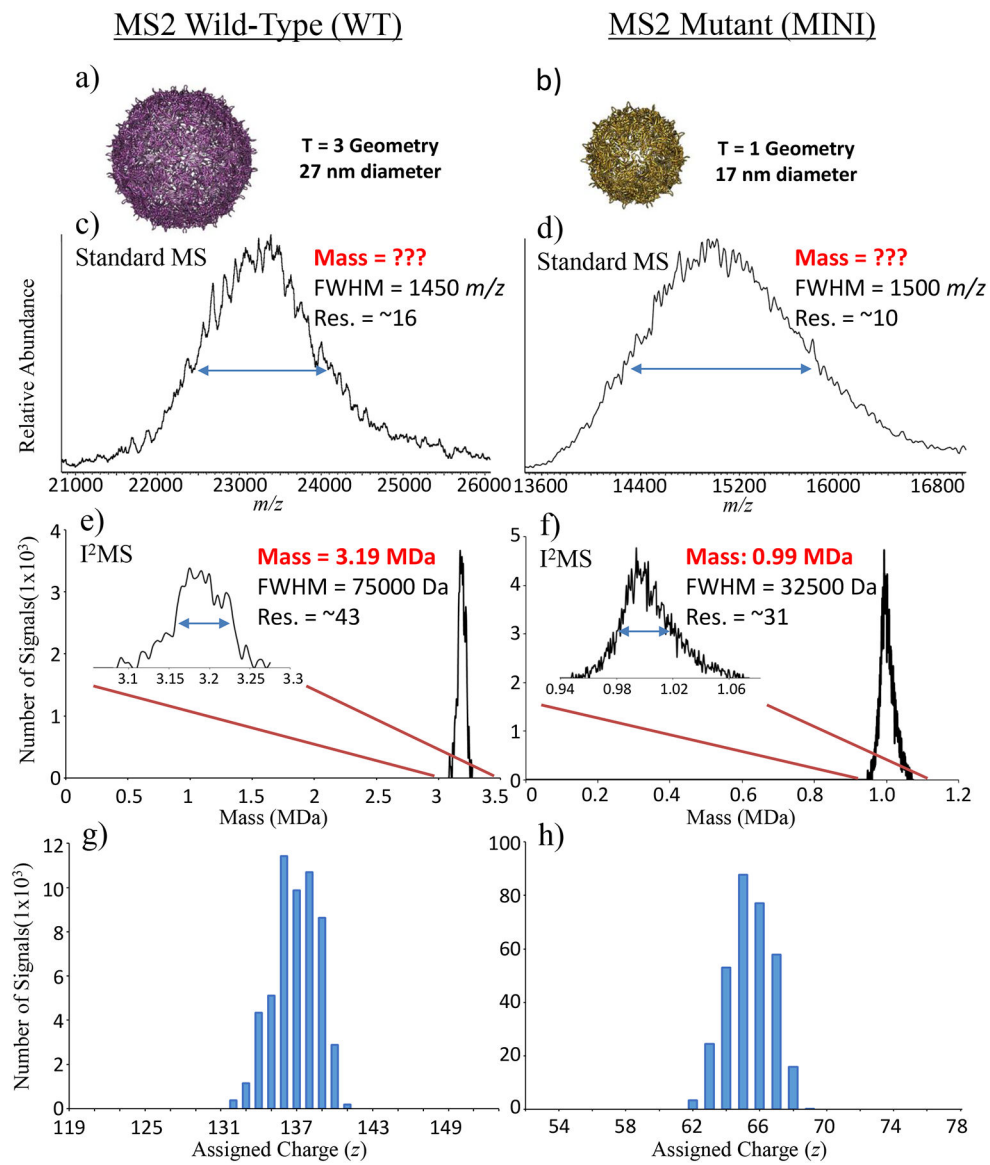


Figure 3. Mass Determination of Large Native Complexes.

Crystal structure renderings (a,b) and mass spectra of assembled WT (c,e,g) vs. MINI (d,f,h) MS2 virus-like particles, with approximate masses and capsid diameters of 3.2 vs. 1 MDa and 27 vs. 17 nm, respectively. In panels (c) and (d), the standard MS spectra plotted in m/z space results in a loss of charge state resolution due to heterogeneity of cargo inside the capsid of the virus-like particles. However, the corresponding I^2MS spectra in panels (e) and (f) allow mass assignment to the distribution of particles even without resolution of the individual charge states (a normal requirement of conventional data produced by electrospray MS). The individual ions assigned to their underlying charge distributions are shown in panels (g) and (h). Mass and resolution (Res.) at full-width of the half-maximum (FWHM) of the peak heights are labeled beside each peak.