Title

Native and Denaturing MS Protein Deconvolution for Biopharma: Monoclonal Antibodies and Antibody-Drug Conjugates to Polydisperse Membrane Proteins and Beyond.

Journal

Analytical chemistry, 91(15)

ISSN

0003-2700

Authors

Campuzano, Iain DG
Robinson, John H
Hui, John O
et al.

Publication Date

2019-08-01

DOI

10.1021/acs.analchem.9b00062

Peer reviewed

analytical chemistry

Article

Cite This: Anal. Chem. 2019, 91, 9472−9480

pubs.acs.org/ac

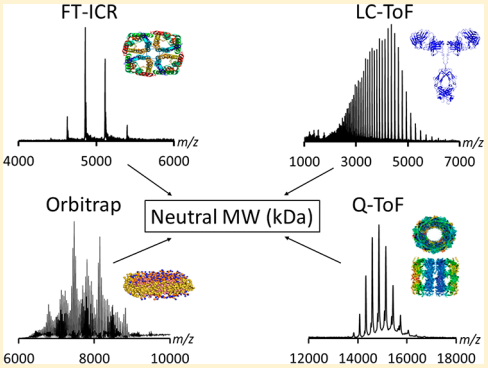
Native and Denaturing MS Protein Deconvolution for Biopharma: Monoclonal Antibodies and Antibody−Drug Conjugates to Polydisperse Membrane Proteins and Beyond

Iain D. G. Campuzano,*[,†] John H. Robinson,[†] John O. Hui,[†] Stone D.-H. Shi,[†] Chawita Netirojjanakul,[‡] Michael Nshanian,[§] Pascal F. Egea,[⊥] Jennifer L. Lippens,[†,¶] Dhanashri Bagal,[∥] Joseph A. Loo,[⊥] and Marshall Bern*[,#]

[†]Discovery Attribute Sciences, Amgen Research, One Amgen Center Drive, Thousand Oaks, California 91320, United States
[‡]Hybrid Modality Engineering, Amgen Research, One Amgen Center Drive, Thousand Oaks, California 91320, United States
[§]Department of Chemistry and Biochemistry, University of California-Los Angeles, Los Angeles, California 90095, United States
[∥]Discovery Attribute Sciences, Amgen Research, Veterans Boulevard, South San Francisco, California 94080, United States
[⊥]Department of Biological Chemistry, University of California-Los Angeles, Los Angeles, California 90095, United States
[#]Protein Metrics, Cupertino, California 95010, United States

Supporting Information

ABSTRACT: Electrospray ionization mass spectrometry (ESI-MS) is a ubiquitously used analytical method applied across multiple departments in biopharma, ranging from early research discovery to process development. Accurate, efficient, and consistent protein MS spectral deconvolution across multiple instrument and detector platforms (time-of-flight, Orbitrap, Fourier-transform ion cyclotron resonance) is essential. When proteins are ionized during the ESI process, a distribution of consecutive multiply charged ions are observed on the m/z scale, either positive $[M + nH]^{n+}$ or negative $[M - nH]^{n-}$ depending on the ionization polarity. The manual calculation of the neutral molecular weight (MW) of single proteins measured by ESI-MS is simple; however, algorithmic deconvolution is required for more complex protein mixtures to derive accurate MWs. Multiple deconvolution algorithms have evolved over the past two decades, all of which have their advantages and disadvantages, in terms of speed, user-input parameters (or ideally lack thereof), and whether they perform optimally on proteins analyzed under denatured or native-MS and solution conditions. Herein, we describe the utility of a parsimonious deconvolution algorithm (explaining the observed spectra with a minimum number of masses) to process a wide range of highly diverse biopharma relevant and research grade proteins and complexes (PEG-GCSF; an IgG1k; IgG1- and IgG2-biotin covalent conjugates; the membrane protein complex AqpZ; a highly polydisperse empty MSP1D1 nanodisc and the tetradecameric chaperone protein complex GroEL) analyzed under native-MS, denaturing LC-MS, and positive and negative modes of ionization, using multiple instruments and therefore multiple data formats. The implementation of a comb filter and peak sharpening option is also demonstrated to be highly effective for deconvolution of highly polydisperse and enhanced separation of a low level lysine glycation post-translational modification (+162.1 Da), partially processed heavy chain lysine residues (+128.1 Da), and loss of *N*-acetylglucosamine (GlcNAc; −203.1 Da).



Mass spectrometry plays a critical role in multiple stages of pharmaceutical research. From small molecule medicinal chemistry research efforts[1] to high throughput screening efforts of drug targets[2] to monoclonal antibody separation and accurate molecular weight determination,[3] mass spectrometry (MS) is a ubiquitous analytical method used throughout biopharma. All the aforementioned examples rely on either liquid chromatographic (LC) separation or solid phase extraction (SPE) prior to MS analysis. A far less routine MS analytical method, in pharma at least, is native-MS, where noncovalent protein−drug or protein−protein[4] interactions remain intact within the gas phase of the MS. In both cases, algorithmic spectral deconvolution is routinely performed within pharma, for routine accurate and rapid molecular weight (MW) determination, on data derived from multiple instrument platforms (time-of-flight, Orbitrap, and Fourier transform ion cyclotron resonance MS systems).

Since the initial demonstration of native MS experiments on proteins and complexes by Katta and Chait,[5] Loo,[6] and others,[7,8] this unique area of MS has steadily grown from what

Received: January 4, 2019
Accepted: June 13, 2019
Published: June 13, 2019

DOI: 10.1021/acs.analchem.9b00062
Anal. Chem. 2019, 91, 9472−9480

was initially a niche area to a fully established research field, described as gas-phase structural biology.[9,10] The proteins investigated using native-MS and solution conditions (typically 200 mM ammonium acetate, pH 6−7[11,12]) have ranged from the original holo-myoglobin[5] and the alcohol dehydrogenase homotetrameric complex[13] to multisubunit complexes such as GroEL[14] and valinyl-oxidase,[15] described in the late 1990s and early 2000s, to the present day, highly polydisperse nanodisc molecules,[16,17] membrane proteins,[18] and mega-Dalton virus capsids.[19] Denaturing LC-MS has also proven itself to be a highly enabling platform for the rapid determination of accurate MWs of denatured proteins[20] and is routinely used within pharmaceutical research for monoclonal antibody characterization, bispecific antibodies, and proteins of therapeutic interest.[2,3,21,22]

MW determination of a protein or complex can be performed either manually or through software by performing data smoothing and centroid processing, followed by adjacent peak assignment (Figure S1) based on the following formulas reproduced from Fenn:[23]

$$z(m1) = \frac{x(m2 - mp)}{m2 - m1} \quad (1)$$

where z is the calculated charge for $m1$; $m2$ is the ion with x less charge (therefore, in this case $x = 1$, appearing higher in the m/z scale); x can also be 2, 3, 4, 5, etc., as long as the correct m/z value for $m2$ is chosen; mp is the mass of the proton (1.00728 Da). The numerical values for $m1$ and $m2$ are based on MS derived m/z values. Once the value for z is determined, the intact MW can easily be calculated:

$$\text{neutral MW} = z(m1 - mp) \quad (2)$$

For the more complex MS spectra, such as those derived from tandem MS experiments of polydisperse molecules, such as alpha-Crystallin B,[24] choatropic partial disruption of transcription factor iEF3,[25] and the highly polydisperse empty MSP1D1 nanodisc,[16,17] manual peak picking is challenging, if not impossible. In these cases, algorithmic deconvolution is a prerequisite for accurate MW determination. However, recently, effective manual MW determination of an empty MSP1D1 nanodisc has been described.[17]

One of the first and arguably the most heavily used protein deconvolution algorithms is the Bayesian probability-based Maximum Entropy.[26] Maximum Entropy was originally designed to deconvolve MS data of low MW, denatured, multiply charged protein spectra, acquired on low resolution quadrupole-based instruments.[20] On MS instruments where the proteins are analyzed in near neutral pH aqueous solutions,[11,12] the measured peaks for the charge states are typically wider than the expected isotopic peak width distribution and also asymmetric due to solution, buffer, and salt adducting.[11]

To date, there are multiple algorithms available for protein electrospray ionization mass spectrometry (ESI-MS) spectral deconvolution that have evolved over the last 20 years, some of which have focused on denaturing protein spectral data.[23,26,27] However, recently with the advent of native-MS,[9,10] there has been a renewal of interest in charge deconvolution algorithms.[28−33] UniDec[29] and FFT-based deconvolution[32] represent a significant step forward in protein spectral deconvolution. UniDec has the ability to efficiently deconvolve ion mobility-based and highly polydisperse native MS data, such as those generated on empty MSP1D1 nanodiscs. UniDec also incorporates a comb filter, which allows the user to explicitly define the MW repeat of the incorporated phospholipid. The FFT-based deconvolution method developed in the Prell lab[32] does not require any prior known charge or repeat unit values but solely relies on the fundamental frequencies and higher harmonics for MW determination of highly polydisperse and polymeric ions such as nanodiscs and polyethelene glycol. The latest algorithm development called PMI Intact (Protein Metrics) enables rapid and efficient deconvolution of native-MS and denaturing LC-MS spectral data.

Herein, we present the deconvolution of native and denatured MS spectral data for a monoclonal antibody (NIST IgG1k), an antibody−drug conjugate-like molecule (IgG1 and IgG2 conjugated to biotin), the PEG-GCSF protein, a membrane protein (AqpZ), an empty nanodisc (MSP1D1), and a chaperone protein complex (GroEL), acquired on a quadrupole time-of-flight (Q-ToF), an LC-ToF, a Fourier transform ion cyclotron resonance (FTICR), and an Orbitrap MS instrument using both positive and negative modes of ionization, all of which are discussed in the context of a biopharmaceutical relevant universal deconvolution algorithm. The deconvolution algorithm described is the PMI Intact Mass algorithm,[33,34] which uses both forward (m to m/z) and backward (m/z to m) mappings. The Intact Mass algorithm also includes a step to bias the deconvoluted neutral mass spectrum to a "parsimonious" solution with minimal mass peaks as necessary to explain the observed m/z spectrum.

MATERIALS AND METHODS

Mass Spectrometry. Nanoelectrospray ionization (nESI) native-MS was performed using the following source voltages and pressures: Q-ToF Synapt G1 (Waters Corporation): source temperature, 25 °C, source backing pressure, 6.0 mbar, sample cone, 25−200 V, trap collision voltage, 75−125 V, in cC_4F_8; Orbitrap-EMR (ThermoScientific): source transfer capillary temperature, 250 °C, source collision induced dissociation, 80 V, higher-energy collision induced dissociation (HCD), 20 V, in N_2; 15 T solariX FTICR (Bruker Daltonics): source transfer capillary temperature, 100 °C, Skimmer 1, 50 V, collision cell voltage, 30 V, in SF_6. Protein samples were all in the concentration range of 10−20 μM in 200 mM ammonium acetate and introduced in to the MS systems using a gold coated glass nESI needle (long thin wall, M956232AD1-S; Waters Corporation) in positive and negative ionization mode. High m/z calibration was performed under both positive and negative nESI modes of acquisition using a 100 μg/μL solution of CsI (Figure S2). Denaturing LC-MS was performed on an open access enabled 6230 ToF MS (Agilent) connected to an Infinity 1290 LC (Agilent) system operated with a Zorbax SB300, C8 50 × 2.1 mm, 3.5 μm analytical column. More detailed native-MS and LC-MS conditions are noted in the Supporting Information.

Materials. The following proteins and complexes were used in this study: a homotetrameric membrane protein AqpZ;[35] an IgG1 mAb biotin covalent conjugate[34] (deglycosylated using PNGaseF, QA Bio, E-PNG01); an IgG2 conjugated with NHS-PEG_{12}-Biotin (ThermoFisher Scientific, 21312; 2.5 mol equivalents prepared in an identical manner to those described in ref 34); an empty MSP1D1 nanodisc;[17] the PEG-GCSF protein;[36] a tetradecameric chaperone complex GroEL[12] and the NIST IgG1k mAb.[37] All proteins and complexes were buffer exchanged into aqueous 200 mM ammonium acetate

DOI: 10.1021/acs.analchem.9b00062
Anal. Chem. 2019, 91, 9472−9480

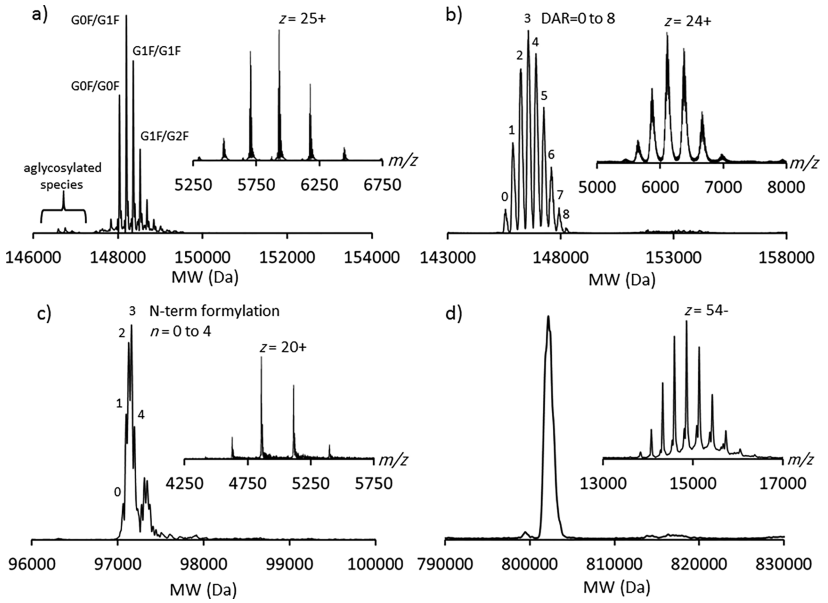
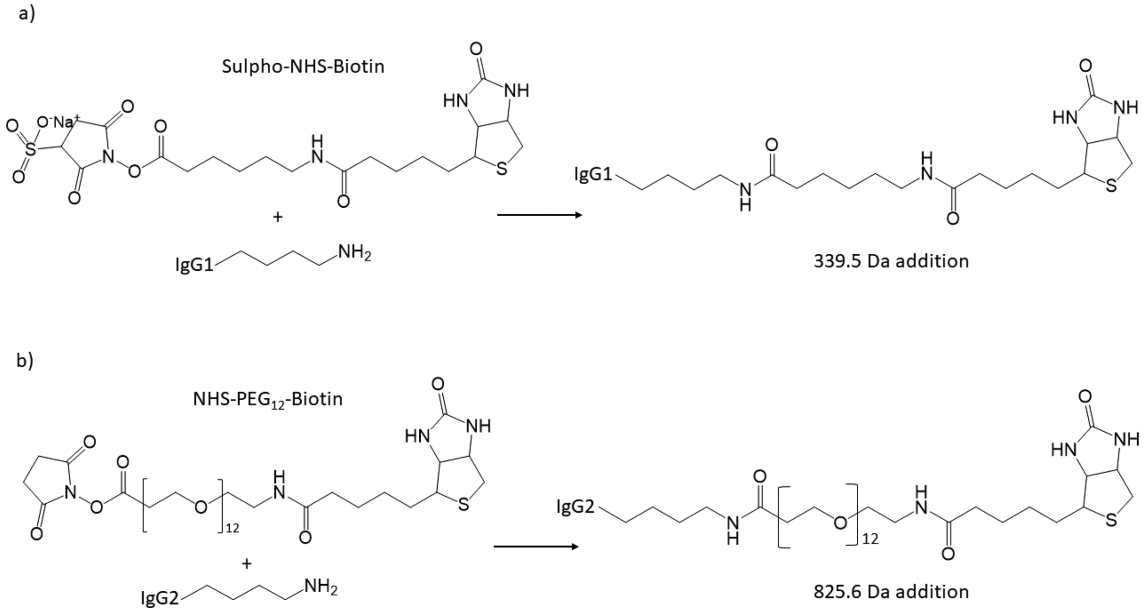


Figure 1. Native MS analyses using multiple MS instrumentation and subsequent algorithmic deconvolution of a diverse range of pharmaceutically relevant and research grade protein constructs: (a) the NIST IgG1k mAb standard analyzed by nESI native-MS mode by Orbitrap-EMR MS; (b) an IgG1 lysine-biotin (10 molar biotin equivalent;[34] PNGaseF treated) conjugate, analyzed by nESI native-MS by Q-ToF MS; (c) aquaporin-Z analyzed by nESI native-MS by FTICR MS;[35] (d) GroEL analyzed by negative native-MS nESI by Q-ToF MS. Insets display the unprocessed data. The most intense charge state, the formylation stoichiometry, and the drug-to-antibody ratio (DAR) values are annotated.

Scheme 1. Synthetic Scheme of Lysine Conjugation Used to Covalently Modify the IgG1 and IgG2 mAbs Described Herein[a]



[a](a) Sulfo-NHS-LC-biotin lysine conjugation;[34] (b) NHS-PEG_{12}-biotin lysine conjugation. A single native lysine residue is represented as only the primary amine side chain. Observed MW additions for a single conjugation are annotated (Da, average MW).

(diluted from Sigma-Aldrich 7.5 M stock, A2705-500 ML) using a P6 micro biospin filter (BioRad, 7326221). The AqpZ 200 mM ammonium acetate solution also contained 1.1% w/v octylglucoside.[35]

Computation. A detailed algorithm description can be found in the Supporting Information. Briefly, PMI Intact uses an iterative algorithm to deduce the mix of charges in each small interval of the m/z spectrum. All charge values are set equally likely for the first deconvolved mass spectrum; new charge values are then computed from the previous deconvolved mass spectrum, and the process is repeated. PMI Intact applies a small "parsimony" bias against m/z intervals with many different charges, because multiple true masses mapping to the same m/z bin are less common than deconvolution artifacts caused by charge uncertainty. On each iteration, the algorithm updates the charge vectors, which provide probabilities for each charge at each point of the observed m/z spectrum. New charge vectors are determined by the last deconvolved mass spectrum along with a priori assumptions about smoothness of charging and likelihood of

DOI: 10.1021/acs.analchem.9b00062
Anal. Chem. 2019, 91, 9472−9480

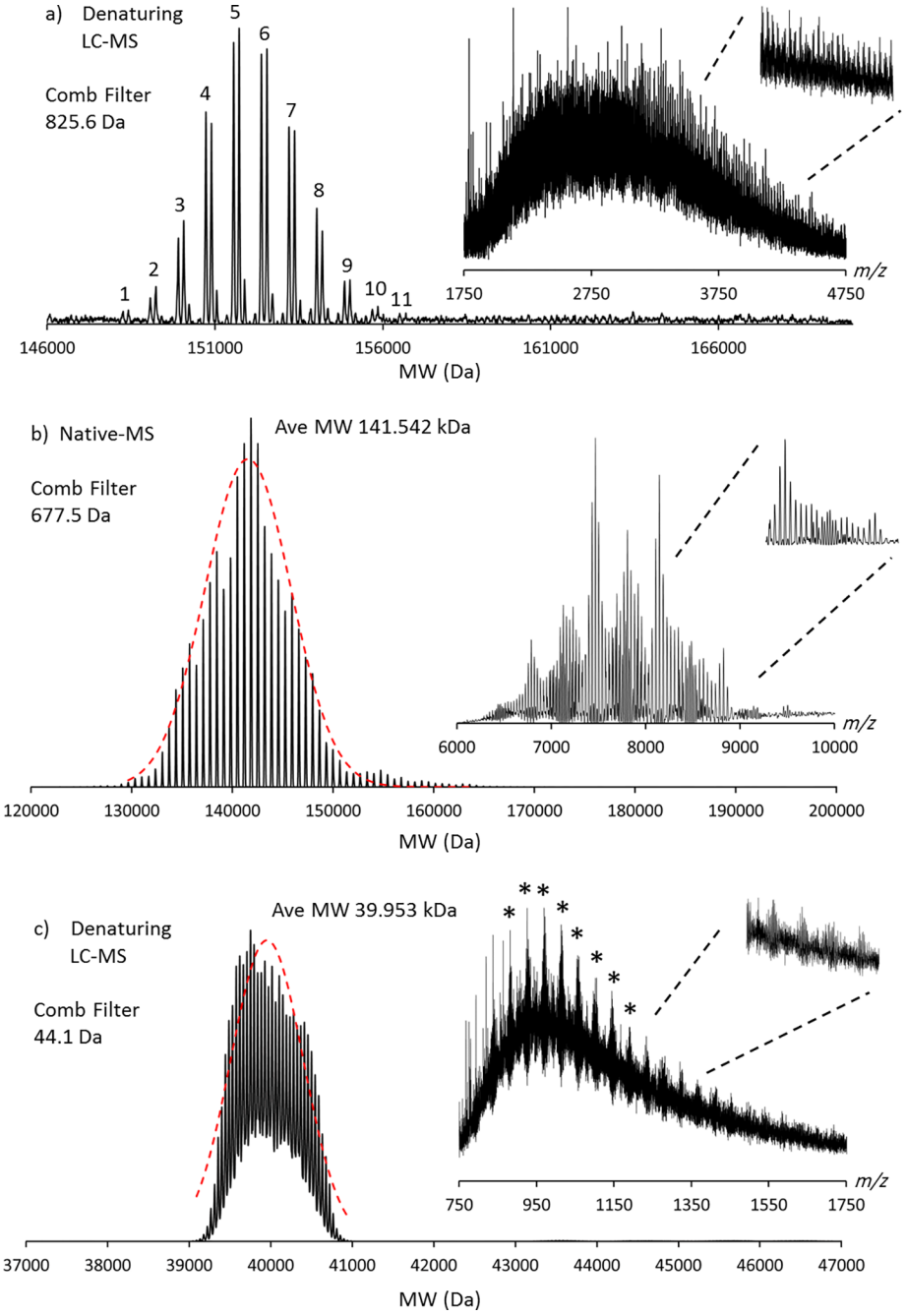


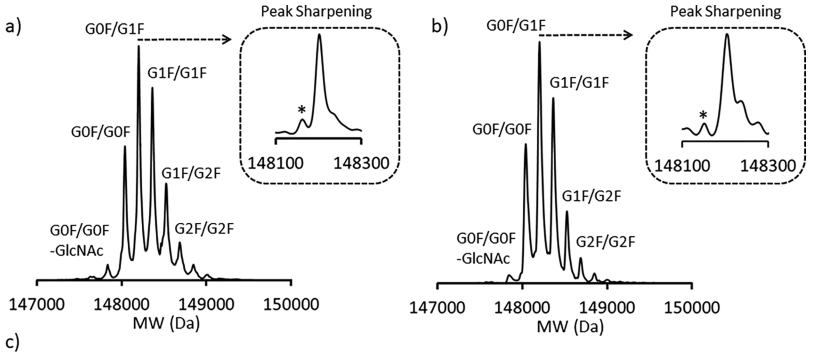
Figure 2. Effect of the comb filter (comb filter = 1) on the deconvolution of highly polydisperse MS spectral data: (a) an IgG2 PEG_{12}-biotin conjugate analyzed by denaturing LC-ToF MS; (b) an empty MSP1D1 nanodisc[17] containing the phospholipid DMPC analyzed by nESI native-MS (Orbitrap-EMR); (c) PEGylated protein analyzed by denaturing LC-ToF MS; * indicates free PEG differing by 44.1 Da. The MW (ave) is calculated from the superimposed (red hashed line) normal distribution. All insets display the unprocessed m/z data. The comb filter delta mass is also annotated. Figure S15a–c display the deconvolved spectra without the use of the comb filter.

mass coincidences. The new charge vectors give a new deconvolved mass spectrum, and each iteration reduces the sum of the squares of the differences between the observed m/z spectrum and the m/z spectrum computed from the last set of charge vectors and deconvolved mass spectrum. For polydisperse targets such as nanodiscs, the algorithm can incorporate a user defined comb filter. For example, 677.5 Da would be used to describe the delta mass for a nanodisc lipid containing dimyristoylphosphocholine (DMPC; Figure S3). Native and denaturing MS deconvolution was performed using PMI Intact (w2.15-584 develop; Protein Metrics Inc.). Raw unprocessed MS data files are dragged directly into the Create Project User Interface (Figures S4 and S5). More detailed discussions of the "Advanced" deconvolution parameters can be found in Figure S6. For additional algorithm information, please refer to the Supporting Information or Bern et al.[33]

RESULTS AND DISCUSSION

Native-MS Deconvolution. It is important to note that the PMI software is vendor neutral and accepts spectral data directly from the raw, unprocessed data files; therefore, they do not need to be converted to text format (typically m/z versus

DOI: 10.1021/acs.analchem.9b00062
Anal. Chem. 2019, 91, 9472−9480



c)

Theoretical MW* (Da)	Glycoform	Denatured LC-MS MW (Da)	ppm	Native MW (Da)	ppm
147,834.0	G0F/G0F-GlcNAc	147,836.5	16.9	147,842.1	54.8
148,037.2	G0F/G0F	148,039.3	14.2	148,039.2	13.5
148,199.3	G0F/G1F	148,201.6	15.5	148,201.3	13.5
148,361.4	G1F/G1F	148,363.5	14.2	148,362.2	5.4
148,523.6	G1F/G2F	148,525.4	12.1	148,523.4	-1.3
148,685.7	G2F/G2F	148,686.4	4.7	148,684.4	-8.7

Figure 3. Comparative deconvolved spectra of the NIST IgG1*k* mAb analyzed under: (a) denaturing LC-ToF MS; (b) native-MS and solution conditions by FTICR. The insets in all cases display a zoom-in of the glycoform G0F/G1F post-peak sharpened processing; * represents a species previously postulated to be a low level unprocessed heavy chain lysine residue (+128.1 Da).[40] Minor differences in charge state distributions are observed when analyzed under native and denaturing MS conditions, consistent with other groups observations.[34,57,64] (c) Deconvolved MWs of denatured and native-MS analyzed NIST mAb glycoforms and their respective errors (in ppm) were calculated from $n = 1$ experiments. The major glycoform theoretical MWs are reproduced from Formolo et al.[40] MW measurements were derived directly from the software centered peak.

intensity) prior to deconvolution. The full deconvolution process ranges from approximately 0.5 to 2.0 min per data file (nESI infusion and LC-MS; number of iteration and processor dependent) based on a HP Z620 Workstation (Intel Xeon 3.7 GHz, 16GB RAM, 12 cores), and the files described herein were processed across a network (data files not stored locally on the processing PC).

Figure 1a–d displays the deconvoluted spectral data for multiple proteins and complexes ranging in MW (97.1 to 802.4 kDa), stoichiometry (up to a tetradecamer), and polydispersity, all measured under native MS and solution conditions (200 mM ammonium acetate) by nESI using different MS instruments. See Figures S7 and S8 for a discussion regarding denatured and native MS theoretical versus instrument derived peak widths.

Figure 1a,b represents typical monoclonal antibodies (mAbs) and antibody–drug conjugates (ADCs), which are highly specific and potent modalities used to treat multiple disease indications.[38,39] Figure 1a displays the NIST IgG1*k* mAb standard analyzed under positive ion native nESI mode on an Orbitrap-EMR MS. As can be seen, the main glycoforms are easily identified (G0F/G0F, G0F/G1F, G1F/G1F, and G1F/G2F). Lower intensity and lower MW species are also identified as the aglycosylated G0F, G1F, and G2F (146.5 to 146.9 kDa[40]). Figure 1b represents a biotin conjugated IgG1 (Scheme 1a; 10 molar biotin equivalents) of relatively simple spectral complexity and MWs ranging from 145.9 to 147.7 kDa, analyzed under positive ion native nESI mode on a Q-ToF MS. Adjacent mass differences correspond to 338.8 ±10.1 Da (theoretical difference is 339.5 Da), representing biotin covalent conjugation to native lysine residues consistent to those previously described by FTICR.[34]

Membrane proteins constitute over 50% of current druggable targets;[41,42] therefore, their characterization by pharma using MS is of high importance.[17,35,43] The analysis of AqpZ, acquired on an FTICR (Figure 1c), represents a membrane protein homotetrameric complex with a low level of polydispersity ($n = 0$ to 4). For AqpZ, the observed MW differences in the deconvolved spectrum are small, ranging from 97.1 to 97.2 kDa, representing a previously described N-terminal formylation (theoretical MW addition of a formylation is 28.01 Da[35]). On the basis of the observed MW differences of adjacent formylation proteoforms (32.8 ±1.9 Da), it would be challenging to positively determine which post-translational modification is present. One would require either ultrahigh resolution mass measurements[44] or proteolytic digestion.[35] An additional larger MW is observed, corresponding to an approximate 182 Da increase, which is not observed when AqpZ is analyzed under denaturing LC-MS conditions (Figure S9); therefore, we attribute this species to an unidentified noncovalent adduct.

Figure 1d represents the deconvolved spectrum of GroEL, the simplest in terms of spectral complexity presented herein. However, the measured charge states (negative nESI; $z = 50-$ to $58-$) are detected far higher in the m/z scale (>14 000) and represent a higher level of salt and buffer adduction than the NIST IgG1*k* sample, the IgG1-biotin conjugate, or the AqpZ complex, therefore representing a different challenge for spectral deconvolution. Upon deconvolution, a major species with MW of 802.4 kDa is detected under negative nESI mode (consistent with the positive nESI mode data; Figure S10). The raised baseline and partially resolved charge states in Figure 1d are indicative of additional species, close in MW. In both deconvolved spectra (positive and negative nESI), there is evidence of a lower MW species (791 kDa). Previously described GroEL spectra[45,46] have also displayed additional low level species present at similar m/z values to that of GroEL. The characterization of these species is beyond the scope of this manuscript; however, they are likely to be either truncated constructs of GroEL or additional protein complexes

DOI: 10.1021/acs.analchem.9b00062
Anal. Chem. 2019, 91, 9472–9480

not removed during the purification procedure.[12] A denatured LC-MS GroEL spectrum is displayed in Figure S11 showing detection of additional lower MW species. Also note that the separation of this low intensity species from adjacent charge states is improved under negative nESI due to the charge states appearing higher in the m/z scale, emphasizing the importance of the rarely utilized negative ion nESI in native MS.[47] In all the aforementioned cases (Figure 1a–d), highly comparable basic and advanced deconvolution parameters were used (Figure S6).

Figure S12 displays a range of proteins (ubiquitin, myoglobin, NIST light and heavy chain, and BSA) of varying MW (8.1 to 66.7 kDa) analyzed by denaturing LC-MS, which have been deconvolved using comparable settings to those used for native-MS spectral processing described in Figures 1, 2, and 3. All spectra are artifact-free and are measured to an overall RMS error of 4.7 ppm to the expected theoretical values (Table S1). In all processed examples described herein, an estimate of expected charge state distribution must be input and is typically set to a wide value (z = 10+ to 100+; Figure S6). This is not the case for the original Maximum Entropy algorithm[26] or the FFT-based deconvolution algorithm[32] but is required for UniDec.[29] In rare cases such as the nanodisc,[16,17] a charge state distribution can be challenging to predict and interpret. However, on the basis of the MW versus z relationship established by de la Mora,[48] an approximate charge distribution can be predicted. Additionally, it is also common practice to perform a "scouting" deconvolution over a wide m/z, MW, and z ranges and then perform a narrow "focused" deconvolution. Ideally, a spectral deconvolution algorithm with minimal input parameters is preferred.

Comb Filter: Deconvolution of Highly Polydisperse Pharmaceutical Relevant Molecules. A comb filter sums or averages evenly spaced points in a signal. PMI Intact allows the user to specify a comb filter to average peaks corresponding to MWs with anticipated mass deltas. The comb filter was added to the "backwards step". A comb filter of width 1 is implemented as an averaging filter with weights 0.25, 0.5, and 0.25 applied to points in the last neutral MW spectrum at masses:

$$m - \Delta,\ m,\ \text{and}\ m + \Delta \tag{3}$$

where Δ is the delta mass (79.98 Da for phosphorylation, for example) and m is the neutral MW. The averaged value is then used to set the probability for charge k at m/z point m_i = 1.00728 + m/k. A comb filter of width 2 uses a weighted average of:

$$m - 2\Delta,\ m - \Delta,\ m,\ m + \Delta,\ \text{and}\ m + 2\Delta \tag{4}$$

An ADC is composed of an antibody with high affinity to a specific target and a covalently attached cytotoxic agent, via native lysine, reduced cysteine, or engineered cysteine residue.[49−51] The resultant drug-to-antibody ratio (DAR) typically ranges from 1 to 8 covalent drug conjugations, using a cleavable or noncleavable linker.[52] Depending on the level/heterogeneity of the glycosylation and the MW of the covalently attached moiety, the resultant MS spectrum can be polydisperse. Currently, there are four ADCs approved by the US FDA and more in clinical trials.[53,54] If one considers the FDA approved chimeric IgG1-based ADC Brentuximab vedotin (ADCETRIS), multiple overlaps between glycosylated charge states (G0F/G0F, G0F/G1F, G1F/G1F) and the cysteine conjugated monomethyl-aurustatin E (MMAE; MW 1316.6 Da[55]) will occur at higher DAR values, resulting in a complex MS spectrum. To reduce the spectral complexity, groups have deglycosylated (PNGaseF treated) and analyzed ADCs under native-MS conditions[56] where charge states appear higher in the m/z scale, resulting in an improved level of separation. However, by processing highly congested MS data with an effective deconvolution algorithm, regardless of where in the m/z scale the charge states appear, an accurate MW should be readily achieved. Figure 2a displays the deconvolved spectrum of an IgG2 mAb (glycans intact) covalently modified with NHS-PEG_{12}-biotin (2.5 mol equivalents) at native lysine residues (Scheme 1b), resulting in multiple covalent MW additions of 825.6 Da. Within the lower m/z regions of the spectrum, the data is highly congested; however, in higher regions of the spectrum (m/z 3750–4750), individual charge states begin to be resolved (Figure 2a, inset). When the comb filter was included (delta mass 825.6 Da), effective deconvolution was achieved. Utilizing the comb filter, the detection of lower S/N species, such as DAR 1, DAR2, DAR 10, and DAR 11, are now significantly improved (Figure S15a). The average DAR value, with and without the use of the comb filter, is also subtly different (5.63 vs 5.54, respectively). This has implications for not only which techniques (LC-MS, native MS, LC-UV, LC-HIC) are used to derive the DAR value[34,57] but also which deconvolution parameters are used to process the MS data; the use of consistent parameters and algorithms are key to optimized experimental process and consistent results.

Figure 2b represents an empty MSP1D1 nanodisc acquired on the Orbitrap-EMR instrument using intermediate activation energies (Supporting Information). Nanodiscs are enabling membrane mimetics and have been demonstrated as an effective means of immobilizing membrane proteins for further drug or fragment screening campaigns within pharma, using surface plasmon resonance.[58] Figure 2b, inset displays a broad, polydisperse spectrum with clear areas of constructive overlap.[59] Effective deconvolution is achieved using a comb filter delta mass of 677.5 Da (average MW of DMPC). An average MW of 141.542 kDa is derived of which there are approximately 143 ± 20–30 DMPC molecules (based on 2× membrane scaffold proteins of MW 22.4 kDa) constituting this empty MSP1D1 nanodisc molecule, consistent with values previously described by native-MS[16,17] and analytical ultracentrifugation.[58] The MW polydispersity, determined using PMI (130 to 160 kDa) using an applied comb filter setting of 1, is consistent with previously deconvolved MSP1D1 nanodisc spectral data.[17]

PEGylation is used to enhance the half-life of therapeutically active molecules;[60,61] however, MS analyses typically result in highly polydisperse spectra within which neutral MWs cannot be ascertained manually; algorithmic deconvolution is essential. Figure 2c shows a deconvolved average MW distribution of 39.953 kDa. This data represents an 18.8 kDa protein covalently modified with 21.9 kDa PEG similar to that described by Bagal et al.[36] The deconvolution was achieved using a comb filter delta mass of 44.1 Da, resulting in an average MW of 39.953 kDa consistent with that previously published, without the need for spectral simplification by charge reduction.[36] The LC-MS deconvolved MW (Figure 2c) is highly consistent with that derived by linear MALDI-TOF-MS (40.050 kDa; Figure S13), and contrary to recent opinion,[36] MALDI-MS is in fact an effective analytical method for MW determination of larger PEGylated protein constructs,

DOI: 10.1021/acs.analchem.9b00062
Anal. Chem. 2019, 91, 9472−9480

with the caveat that linear MALDI-TOF will result in a larger MW spread (Figure S16). Figure 2c, inset shows the unprocessed data (from LC-MS), and highlighted (*) are a series of abundant ions differing by 44.1 Da superimposed on top of a highly polydisperse series of protein-PEG conjugate charge states; in the m/z scale, where the two series overlap, constructive enhancement is observed. These PEG ions are likely to be a result of fragmentation within the MS instrument, as operating at a lower source fragmentor voltage reduced the intensity of these interferences, and subtle differences in the deconvolved ion distribution are also evident (Figure S14). Figure S15 displays the deconvolved data of IgG2-PEG_{12}-Biotin, the empty MSP1D1 nanodisc, and the PEG-GCSF molecule, all processed without the use of the comb filter. In all cases, the implementation of the comb filter improves spectral S/N. Finally, a brief but useful comparison and discussion is made between PMI Intact and five of the most commonly used protein deconvolution algorithms to process the highly polydisperse PEG-GCSF LC-MS spectral data: MaxEnt (MassLynx, Waters), MaxEnt and PMod (MassHunter, Agilent), UniDec,[29] and iFAMS[62] (Figure S16). In summary, a full algorithm comparison is well beyond the scope of this manuscript; however, qualitatively, the more recently developed algorithms such as UniDec,[29] iFAMS,[62] and PMI[33] appear to produce deconvolved spectra of superior quality.

Comparing Native and Denaturing MS Spectra. One must now consider whether deconvolution parameters can remain constant when processing denatured and native-MS spectral data and whether mass measurement parity is retained for the same protein molecule. In this case, the NIST IgG1*k* mAb is compared. Figure 3a represents the NIST IgG1k analyzed under denaturing LC-MS conditions (oa-ToF, C8 reversed phase using *n*-propanol, TFA 0.1% and 70 °C;[3] Supporting Information) and Figure 3b, the native-MS and solution condition spectrum (nESI, 15 T FTICR, 200 mM ammonium acetate, Supporting Information). The zero-charged deconvolved MW and mass measurement for the glycoforms G0F/G0F-*N*-acetylglucosamine (GlcNAc), G0F/G0F, G0F/G1F, G1F/G1F, and G1F/G2F are presented in Figure 3c. The denatured LC-MS and native-MS spectra display highly comparable MWs and respective mass measurement values. Only the glycoform G0F/G0F-GlcNAc displays a significant difference in measured mass accuracy. Less adduction is observed under LC-MS denaturing conditions; therefore, improved mass measurement is achieved (unprocessed NIST data are displayed in Figure S17). In the denatured LC-MS data (Figure 3a), a lower m/z leading edge species is detected. This species can be further resolved (Figure 3a, inset, zoom of glycoform G0F/G1F) by using the Peak Sharpening option under Advanced Settings (Figures S6 and S18). This feature can also be detected as a leading edge in the native-MS FTICR spectrum (Figure 3b). Upon Peak Sharpening, this species is further resolved (Figure 3d, inset) and is highly consistent to that presented in Figure 3a. This minor species has previously been characterized as an uncleaved C-terminal heavy chain lysine residue (+128.1 Da[40]) superimposed (but partially resolved) over the adjacent glycoform (G0F/G1F) species (*des*-K form). However, the MW difference obtained using two separate acquisitions (denaturing LC-MS-ToF and native FTICR MS) is also consistent with a loss of a GlcNAc (−203.1 Da) from an adjacent glycoform. Figure S19a shows the deconvolved LC-MS for the NIST mAb heavy chain where multiple, well resolved −GlcNAc (−203.1 Da) species are detected. In Figure S19b, a +128.1 Da addition is detected, representing a low level unprocessed heavy chain C-terminal lysine residue.[40] It is likely these leading-edge partially resolved species in Figure 3a,b are in fact a mixture of +128.1 and −203.1 Da, and relative ion intensity values appear to support this hypothesis (Figures S18 and S19).

A similar comparison was made for the IgG1-biotin 5 mol equivalents analyzed under denaturing LC-MS and native-MS (Figure S20). In both cases, the application of the peak-sharpening feature allows for the improved definition and mass measurement of the +162 Da glycation, a lysine PTM commonly observed in mAbs.[63] This improvement in glycation definition has also been demonstrated through FTICR transient apodization.[34] Additionally, the trailing edge shoulder is now partially resolved in the native-MS spectrum. The mass difference is approximately 40 Da; therefore, it is likely a noncovalent sodium, potassium, ammonium adduct, or multiples thereof. Minor differences in charge state distributions of mAb conjugates and ADCs analyzed under native and denaturing MS conditions and related analytical techniques have been previously addressed by multiple groups.[34,57,64] We assume this difference also holds true for mAb glycoforms described herein (Figure 3a,b; note the minor intensity differences between glycoforms G1F/G2F).

CONCLUSIONS

The use of a parsimonious deconvolution algorithm has been demonstrated to efficiently deconvolute spectral data acquired for proteins and complexes, both pharmaceutically relevant constructs and research grade standards, analyzed under native-MS and denaturing conditions (LC-MS) under both positive and negative modes of ionization. MS data from three different analyzers (oa-ToF, Orbitrap, and FTICR) and four different instrument vendors (Waters, ThermoScientific, Agilent, and Bruker) were successfully deconvoluted without any file format change.

The proteins and complexes analyzed varied in MW, stoichiometry, and m/z range: the NIST IgG1*k* (mAb, 148.3 kDa); an IgG1-biotin conjugate (ADC-like; 146.5 kDa); IgG1-PEG-Biotin (ADC-like; 147.5 kDa); a PEG-GCSF (39.9 kDa; up to 43 measurable PEG 20k units); an empty MSP1D1 nanodisc (141.5 kDa; two membrane scaffold proteins, approximately 124 to 170 measurable DMPC phospholipid molecules); the membrane protein AqpZ (noncovalent homotetramer, 97.5 kDa); the chaperone protein complex GroEL (homotetradecameric, 802.4 kDa). Highly comparable deconvolution parameters were used in all cases, and the resultant zero-charged spectra are artifact free (zero harmonics; third, half, double, and triple multiples of the protein MW). Additionally, when processing denatured LC-MS or native-MS spectral data (of the same constructs, NIST IgG1*k* and the IgG1-biotin conjugate), the deconvolution parameters remained constant and unchanged. In both cases, the deconvolved, zero-charged data peak widths consistently reflect those of the unprocessed data. Mass accuracy is also highly comparable.

From an industrial and biopharmaceutical perspective, this deconvolution algorithm suite is highly advantageous, as most laboratories within a research discovery and process development setting will likely use multiple MS instruments from different vendors; the ability to drag-and-drop multiple MS data files of different formats and subsequently process them is

DOI: 10.1021/acs.analchem.9b00062
Anal. Chem. 2019, 91, 9472−9480

highly attractive. Also, in certain cases, it may be required that both denaturing and native-MS analyses be performed on the same protein construct. For example, one may want to derive an accurate mAb MW through LC-MS analysis,[3] levels of specific covalent modification from high throughput screening campaign,[2] or a drug-to-antibody ratio[34,57,65] or assess the levels of degradation[66] of biotherapeutic molecules or the levels of aggregation (by SEC coupled to native-MS[67]) present in the sample. Native-MS in biopharma is also used for assessing the correct assembly of a nanodisc; it is rapid (ca. 5 min), and when combined with rapid and accurate deconvolution, one can accurately assess the level of DMPC incorporation and therefore ascertain its correct formation for further downstream manipulation of membrane proteins, for example, SPR dose dependence experiments.[58] In summary, a single algorithm can now be used for protein deconvolution within the pharmaceutical research environment, therefore removing much of the subjectivity that still exists in this most basic area of MS analytics.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.analchem.9b00062.

Traditional peak-picking MW determination, detailed mass spectrometer instrument parameters, deconvolution algorithm details, external CsI m/z calibration in both positive and negative nESI mode, advanced deconvolution algorithm parameters, comb filter parameters, theoretical versus instrument derived peak widths, deconvolved denatured LC-MS AqpZ data, GroEL nESI positive ion mode data, deconvolved denatured LC-MS GroEL data, deconvolved denatured ubiquitin, myoglobin, BSA, NIST light and heavy chain data and accompanying mass measurement errors, IgG2-PEG_{12}-biotin, empty nanodisc MSP1D1, and PEG-GCSF deconvolution without the implementation of the comb filter, PEG-GCSF intact linear MALDI analysis, denaturing LC-MS of PEG-GCSF analyzed at different source fragmentor voltages, a comparison of PMod, MaxEnt, iFAMS, UniDec, and PMI Intact, and the effects of algorithmic peak sharpening (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

*E-mail: iainc@amgen.com (I.D.G.C.).
*E-mail: bern@proteinmetrics.com (M.B.).

ORCID

Iain D. G. Campuzano: 0000-0003-4310-8540
Dhanashri Bagal: 0000-0003-1262-231X
Joseph A. Loo: 0000-0001-9989-1437

Notes

The authors declare no competing financial interest.
¶J.L.L. is an Amgen Post-Doctoral Research Fellow.

■ ACKNOWLEDGMENTS

Support from the US National Institutes of Health (R01GM103479 and S10RR028893 to J.A.L.) and the US Department of Energy for the UCLA/DOE Institute for Genomics and Proteomics (DE-FC02-02ER63421 to J.A.L.) is gratefully acknowledged. We also thank Eric Carlson, Yong Joo Kil, and Ilker Sen (Protein Metrics) for help with high MW and high m/z processing optimization. The authors would like to thank James Prell and Sean Cleary (University of Oregon) for processing the GCSF-PEG LC-MS data with iFAMS. We also thank Michael Marty (University of Arizona) for useful discussions regarding UniDec. The authors also wish to thank Ryan Holder and Nic Angell (Amgen) for their useful discussion on the GCSF-PEG LC-MS data. Peter Tieleman (University of Calgary) is thanked for the coarse-grain empty nanodisc image, which is used in the TOC.

■ REFERENCES

(1) Wanner, K., Hofner, G., Eds. *Mass Spectrometry in Medicinal Chemistry: Applications in Drug Discovery*; Methods and Principles in Medicinal Chemistry; Wiley: New York, 2007; DOI: 10.1002/9783527610907
(2) Campuzano, I. D.; San Miguel, T.; Rowe, T.; Onea, D.; Cee, V. J.; Arvedson, T.; McCarter, J. D. *J. Biomol. Screening* **2016**, *21* (2), 136−44.
(3) Dillon, T. M.; Bondarenko, P. V.; Rehder, D. S.; Pipes, G. D.; Kleemann, G. R.; Ricci, M. S. *J. Chromatogr A* **2006**, *1120* (1−2), 112−20.
(4) Marcoux, J.; Robinson, C. V. *Structure* **2013**, *21* (9), 1541−50.
(5) Katta, V.; Chait, B. T.; Carr, S. *Rapid Commun. Mass Spectrom.* **1991**, *5* (4), 214−7.
(6) Loo, J. A. *Mass Spectrom. Rev.* **1997**, *16*, 1−23.
(7) Ganem, B.; Li, Y. T.; Henion, J. D. *J. Am. Chem. Soc.* **1991**, *113*, 6294−6296.
(8) Robinson, C. V.; Radford, S. E. *Structure* **1995**, *3* (9), 861−5.
(9) Benesch, J. L.; Ruotolo, B. T.; Simmons, D. A.; Robinson, C. V. *Chem. Rev.* **2007**, *107* (8), 3544−67.
(10) Sharon, M. *Science* **2013**, *340* (6136), 1059−60.
(11) Hernandez, H.; Robinson, C. V. *Nat. Protoc.* **2007**, *2* (3), 715−26.
(12) Campuzano, I.; Giles, K. Nanospray Ion Mobility Mass Spectrometry of Selected High Mass Species. In *Nanoproteomics: Methods and Protocols, Methods in Molecular Biology*; Toms, S. A., Weil, R., Eds.; Humana Press, a part of Springer Science+Business Media, LLC: New York, 2011; Vol. *790*, pp 57−70.
(13) Loo, J. A. *J. Mass Spectrom.* **1995**, *30* (1), 180−183.
(14) Rostom, A. A.; Robinson, C. V. *J. Am. Chem. Soc.* **1999**, *121*, 4718−4719.
(15) van Berkel, W. J.; van den Heuvel, R. H.; Versluis, C.; Heck, A. J. *Protein Sci.* **2000**, *9* (3), 435−9.
(16) Marty, M. T.; Zhang, H.; Cui, W.; Gross, M. L.; Sligar, S. G. *J. Am. Soc. Mass Spectrom.* **2014**, *25* (2), 269−77.
(17) Campuzano, I. D.; Li, H.; Bagal, D.; Lippens, J. L.; Svitel, J.; Kurzeja, R. J.; Xu, H.; Schnier, P. D.; Loo, J. A. *Anal. Chem.* **2016**, *88* (24), 12427−12436.
(18) Landreh, M.; Marty, M. T.; Gault, J.; Robinson, C. V. *Curr. Opin. Struct. Biol.* **2016**, *39*, 54−60.
(19) van de Waterbeemd, M.; Snijder, J.; Tsvetkova, I. B.; Dragnea, B. G.; Cornelissen, J. J.; Heck, A. J. *J. Am. Soc. Mass Spectrom.* **2016**, *27* (6), 1000−9.
(20) Rai, D. K.; Griffiths, W. J.; Landin, B.; Wild, B. J.; Alvelius, G.; Green, B. N. *Anal. Chem.* **2003**, *75* (9), 1978−82.
(21) Schachner, L.; Han, G.; Dillon, M.; Zhou, J.; McCarty, L.; Ellerman, D.; Yin, Y.; Spiess, C.; Lill, J. R.; Carter, P. J.; Sandoval, W. *Anal. Chem.* **2016**, *88* (24), 12122−12127.
(22) He, J.; Su, D.; Ng, C.; Liu, L.; Yu, S. F.; Pillow, T. H.; Del Rosario, G.; Darwish, M.; Lee, B. C.; Ohri, R.; Zhou, H.; Wang, X.; Lu, J.; Kaur, S.; Xu, K. *Anal. Chem.* **2017**, *89* (10), 5476−5483.
(23) Mann, M.; Meng, C. K.; Fenn, J. B. *Anal. Chem.* **1989**, *61*, 1702−1708.
(24) Benesch, J. L.; Aquilina, J. A.; Ruotolo, B. T.; Sobott, F.; Robinson, C. V. *Chem. Biol.* **2006**, *13* (6), 597−605.

DOI: 10.1021/acs.analchem.9b00062
Anal. Chem. 2019, 91, 9472−9480

(25) Zhou, M.; Sandercock, A. M.; Fraser, C. S.; Ridlova, G.; Stephens, E.; Schenauer, M. R.; Yokoi-Fong, T.; Barsky, D.; Leary, J. A.; Hershey, J. W.; Doudna, J. A.; Robinson, C. V. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105* (47), 18139−44.
(26) Ferrige, A. G.; Seddon, M. J.; Green, B. N.; Jarvis, S. A.; Skilling, J.; Staunton, J. *Rapid Commun. Mass Spectrom.* **1992**, *6*, 707−711.
(27) Zhang, Z.; Marshall, A. G. *J. Am. Soc. Mass Spectrom.* **1998**, *9*, 225−233.
(28) Morgner, N.; Robinson, C. V. *Anal. Chem.* **2012**, *84* (6), 2939−48.
(29) Marty, M. T.; Baldwin, A. J.; Marklund, E. G.; Hochberg, G. K.; Benesch, J. L.; Robinson, C. V. *Anal. Chem.* **2015**, *87* (8), 4370−6.
(30) Lu, J.; Trnka, M. J.; Roh, S. H.; Robinson, P. J.; Shiau, C.; Fujimori, D. G.; Chiu, W.; Burlingame, A. L.; Guan, S. *J. Am. Soc. Mass Spectrom.* **2015**, *26* (12), 2141−51.
(31) Guan, S.; Trnka, M. J.; Bushnell, D. A.; Robinson, P. J.; Gestwicki, J. E.; Burlingame, A. L. *Anal. Chem.* **2015**, *87* (16), 8541−6.
(32) Cleary, S. P.; Thompson, A. M.; Prell, J. S. *Anal. Chem.* **2016**, *88* (12), 6205−13.
(33) Bern, M.; Caval, T.; Kil, Y. J.; Tang, W.; Becker, C.; Carlson, E.; Kletter, D.; Sen, K. I.; Galy, N.; Hagemans, D.; Franc, V.; Heck, A. J. R. *J. Proteome Res.* **2018**, *17* (3), 1216−1226.
(34) Campuzano, I. D. G.; Netirojjanakul, C.; Nshanian, M.; Lippens, J. L.; Kilgour, D. P. A.; Van Orden, S.; Loo, J. A. *Anal. Chem.* **2018**, *90* (1), 745−751.
(35) Lippens, J. L.; Nshanian, M.; Spahr, C.; Egea, P. F.; Loo, J. A.; Campuzano, I. D. G. *J. Am. Soc. Mass Spectrom.* **2018**, *29* (1), 183−193.
(36) Bagal, D.; Zhang, H.; Schnier, P. D. *Anal. Chem.* **2008**, *80* (7), 2408−18.
(37) Campuzano, I. D. G.; Larriba, C.; Bagal, D.; Schnier, P. D. *ACS Symp. Ser.* **2015**, *1202*, 75−112.
(38) Leavy, O. *Nat. Rev. Immunol.* **2010**, *10* (5), 297.
(39) Alley, S. C.; Okeley, N. M.; Senter, P. D. *Curr. Opin. Chem. Biol.* **2010**, *14* (4), 529−37.
(40) Formolo, T.; Ly, M.; Levy, M.; Kilpatrick, L.; Lute, S.; Phinney, K.; Marzilli, L.; Brorson, K.; Boyne, M.; Davis, D.; Schiel, J. Determination of the NISTmAb Primary Structure. In *State-of-the-Art and Emerging Technologies for Therapeutic Monoclonal Antibody Characterization*; ACS Symposium Series; American Chemical Society: Washington, DC: 2015 Vol. *1201*; DOI: 10.1021/bk-2015-1201.ch001.
(41) Hopkins, A. L.; Groom, C. R. *Nat. Rev. Drug Discovery* **2002**, *1* (9), 727−30.
(42) Arinaminpathy, Y.; Khurana, E.; Engelman, D. M.; Gerstein, M. B. *Drug Discovery Today* **2009**, *14* (23−24), 1130−5.
(43) Lippens, J. L.; Egea, P. F.; Spahr, C.; Vaish, A.; Keener, J. E.; Marty, M. T.; Loo, J. A.; Campuzano, I. D. G. *Anal. Chem.* **2018**, *90* (22), 13616−13623.
(44) Marshall, A. G.; Hendrickson, C. L.; Jackson, G. S. *Mass Spectrom. Rev.* **1998**, *17* (1), 1−35.
(45) Rose, R. J.; Damoc, E.; Denisov, E.; Makarov, A.; Heck, A. J. *Nat. Methods* **2012**, *9* (11), 1084−6.
(46) Belov, M. E.; Damoc, E.; Denisov, E.; Compton, P. D.; Horning, S.; Makarov, A. A.; Kelleher, N. L. *Anal. Chem.* **2013**, *85* (23), 11163−73.
(47) Liko, I.; Hopper, J. T.; Allison, T. M.; Benesch, J. L.; Robinson, C. V. *J. Am. Soc. Mass Spectrom.* **2016**, *27* (6), 1099−104.
(48) de la Mora, J. F. *Anal. Chim. Acta* **2000**, *406* (1), 93−104.
(49) Kim, M. T.; Chen, Y.; Marhoul, J.; Jacobson, F. *Bioconjugate Chem.* **2014**, *25* (7), 1223−32.
(50) Panowski, S.; Bhakta, S.; Raab, H.; Polakis, P.; Junutula, J. R. *MAbs* **2014**, *6* (1), 34−45.
(51) Valliere-Douglass, J. F.; Hengel, S. M.; Pan, L. Y. *Mol. Pharmaceutics* **2015**, *12* (6), 1774−83.
(52) McCombs, J. R.; Owen, S. C. *AAPS J.* **2015**, *17* (2), 339−51.
(53) Mullard, A. *Nat. Rev. Drug Discovery* **2013**, *12* (5), 329−32.
(54) Diamantis, N.; Banerji, U. *Br. J. Cancer* **2016**, *114* (4), 362−7.
(55) D'Atri, V.; Fekete, S.; Stoll, D.; Lauber, M.; Beck, A.; Guillarme, D. *J. Chromatogr. B: Anal. Technol. Biomed. Life Sci.* **2018**, *1080*, 37−41.
(56) Dyachenko, A.; Wang, G.; Belov, M.; Makarov, A.; de Jong, R. N.; van den Bremer, E. T.; Parren, P. W.; Heck, A. J. *Anal. Chem.* **2015**, *87* (12), 6095−102.
(57) Chen, J.; Yin, S.; Wu, Y.; Ouyang, J. *Anal. Chem.* **2013**, *85* (3), 1699−704.
(58) Xu, H.; Hill, J. J.; Michelsen, K.; Yamane, H.; Kurzeja, R. J.; Tam, T.; Isaacs, R. J.; Shen, F.; Tagari, P. *Biochim. Biophys. Acta* **2015**, *1848* (10 Pt A), 1974−1980.
(59) Marty, M. T.; Hoi, K. K.; Gault, J.; Robinson, C. V. *Angew. Chem., Int. Ed.* **2016**, *55* (2), 550−4.
(60) Molineux, G. *Pharmacotherapy* **2003**, *23* (8 Pt 2), 3S−8S.
(61) Molineux, G. *Anti-Cancer Drugs* **2003**, *14* (4), 259−64.
(62) Cleary, S. P.; Li, H.; Bagal, D.; Loo, J. A.; Campuzano, I. D. G.; Prell, J. S. *J. Am. Soc. Mass Spectrom.* **2018**, *29* (10), 2067−2080.
(63) Haberger, M.; Bomans, K.; Diepold, K.; Hook, M.; Gassner, J.; Schlothauer, T.; Zwick, A.; Spick, C.; Kepert, J. F.; Hienz, B.; Wiedmann, M.; Beck, H.; Metzger, P.; Molhoj, M.; Knoblich, C.; Grauschopf, U.; Reusch, D.; Bulau, P. *MAbs* **2014**, *6* (2), 327−39.
(64) Debaene, F.; Boeuf, A.; Wagner-Rousset, E.; Colas, O.; Ayoub, D.; Corvaia, N.; Van Dorsselaer, A.; Beck, A.; Cianferani, S. *Anal. Chem.* **2014**, *86* (21), 10674−83.
(65) Luo, Q.; Chung, H. H.; Borths, C.; Janson, M.; Wen, J.; Joubert, M. K.; Wypych, J. *Anal. Chem.* **2016**, *88* (1), 695−702.
(66) Ren, D.; Pipes, G. D.; Hambly, D. M.; Bondarenko, P. V.; Treuheit, M. J.; Brems, D. N.; Gadgil, H. S. *J. Chromatogr. A* **2007**, *1175* (1), 63−8.
(67) Haberger, M.; Leiss, M.; Heidenreich, A. K.; Pester, O.; Hafenmair, G.; Hook, M.; Bonnington, L.; Wegele, H.; Haindl, M.; Reusch, D.; Bulau, P. *MAbs* **2016**, *8* (2), 331−9.

DOI: 10.1021/acs.analchem.9b00062
Anal. Chem. 2019, 91, 9472−9480

Supporting Information

Native and Denaturing MS Protein Deconvolution for Biopharma: Monoclonal Antibodies and Antibody-Drug-Conjugates to Polydisperse Membrane Proteins and Beyond

Iain D. G. Campuzano[1]*, John H. Robinson[1], John O. Hui[1], Stone D.-H. Shi[1], Chawita Netirojjanakul[2], Michael Nshanian[3], Pascal F. Egea[5], Jennifer L. Lippens[1¶], Dhanashri Bagal[4], Joseph A. Loo[2, 5] and Marshall Bern[6]*

1. Amgen Discovery Research, Discovery Attribute Sciences, One Amgen Center Drive, Thousand Oaks, CA, 91320, USA
2. Amgen Discovery Research, Hybrid Modality Engineering, One Amgen Center Drive, Thousand Oaks, CA, 91320, USA.
3. University of California-Los Angeles, Dept. Chemistry and Biochemistry, Los Angeles, CA, 90095, USA.
4. Amgen Discovery Research, Discovery Attribute Sciences, Veterans Ways, South San Francisco, CA, 94080, USA.
5. University of California-Los Angeles, Dept. Biological Chemistry, Los Angeles, CA, USA.
6. Protein Metrics, Cupertino, CA, 95010, USA.

*Corresponding authors: Iain D. G. Campuzano, iainc@amgen.com and Marshall Bern, bern@proteinmetrics.com

¶Jennifer L. Lippens is an Amgen Post-Doctoral Research Fellow

Supporting Information Content

Figure S15. Deconvolution of IgG2-PEG_{12}-Biotin conjugate, the empty MSP1D1 nanodisc and GCSF-PEG.

Figure S16. A comparison of PEG-GCSF 1+ ion measured and detected by linear MALDI-TOF-MS and the neutral MW deconvolved spectra obtained from PMod, Maxent, MaxEnt, iFAMS, UniDec and PMI.

Figure S17. Comparison of native FT-ICR and LC-oaToF charge state distributions for the NIST IgG1*k* mAb.

Figure S18. The effect of Peak Sharpening on the NIST IgG1*k* mAb acquired under denaturing LC-MS conditions.

Figure S19. NIST IgG1*k* reduced HC data from denaturing LC-MS, displaying the -203 Da from each glycoform.

Figure S20. PNGaseF treated IgG1-biotin conjugate analyzed under denaturing LC-MS conditions and native-MS nESI and solution conditions. The effect of peak-sharpening is demonstrated.

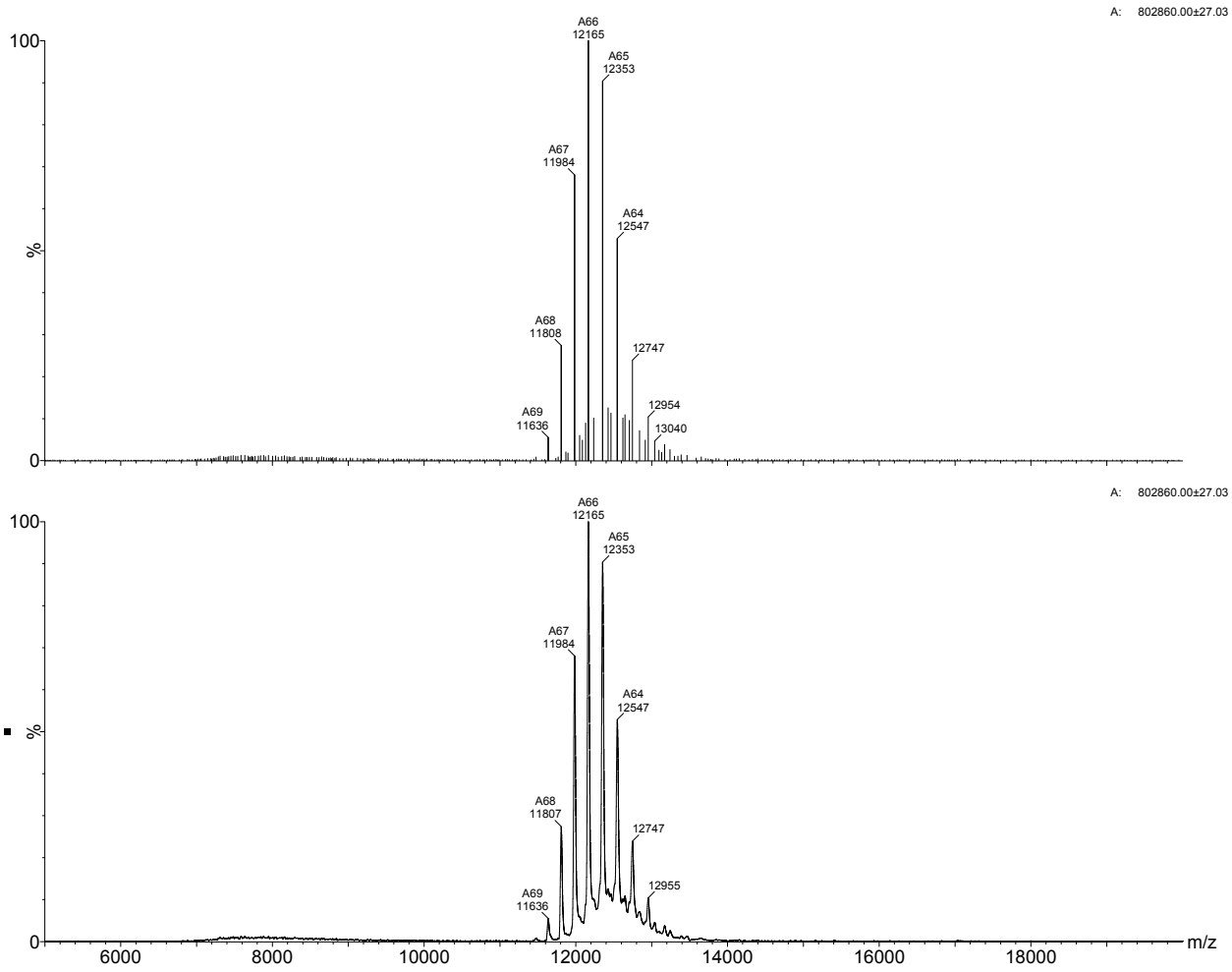


Figure S1. A native solution (200 mM ammonium acetate) nESI positive ion spectrum of 5 µM GroEL acquired on the Synapt G1 using a sample cone of 50V and a trap collision energy of 100V. The data is displayed using MassLynx 4.1. The lower spectrum is smoothed (50 channels x 2 smooths, Mean). The upper spectrum represents the smoothed spectrum which has also been centered (50 channels, 10% top). Manual peak picking of adjacent charge states can be performed through MassLynx with the subsequent automatic charge and therefore molecular weight assignment.

Detailed Mass Spectrometer Instrument Parameters

All Q-ToF experiments were performed on a Synapt G1 HDMS instrument operated in positive nanoflowESI mode. This instrument had been converted to an RF-confining drift-tube instrument, similar to that described by Bush [1]. All critical instrument voltages and pressures are as follows: capillary voltage 0.8-1.2 kV; sample cone 25-200 V, extraction cone 1 V; source block temperature 25 °C; trap collision energy 75-125 V; transfer collision energy 20 V; trap entrance 2.0 V; trap bias 5 V; trap exit 0.0 V; IMS entrance -20 V; IMS exit 21 V; transfer entrance 1.0V; transfer exit 1.0 V; transfer velocity 248 m/sec; transfer wave amplitude 3.0V; source RF-amplitude (peak-to-peak) 450V; triwave RF-amplitudes (peak-to-peak) trap 380V, IMS 250V, transfer 380V; source backing pressure 6.0 mbar; trap/transfer pressure cC_4F_8, 2.00-3.25e^{-2} mbar (pirani gauge indicated; flow rate 4.0-9.0 mL/min). Instrument control and data acquisition were carried out through MassLynx 4.1 SCN 872.

All OrbiTrap-EMR experiments were performed on a modified Orbitrap Exactive Plus instrument (ThermoFisher Scientific, Bremen, Germany) equipped with a nanoflowESI source. All critical instrument voltages and pressures are as follows: Capillary voltage 0.8-1.5kV. Ions formed by nESI are passed through a heated stainless steel capillary (6.5 cm ion transfer tube) maintained at 250^0C into an S-Lens stacked ring ion guide with an applied RF-amplitude (peak-to-peak) of 200V. Ions then travel through a transport multipole and enter the HCD cell where they were stored at a high pressure before they were returned to the C-trap. This feature allows efficient trapping and desolvation of large protein ions and dramatically improves sensitivity. Nitrogen gas was used in the C-trap as well as the HCD cell. Utilising a trapping gas pressure setting of 7.0 (software determined) the C-trap pressure is approximately 2e^{-4}mbar and the UHV pressure (OrbiTrap analyser) is 7.5e^{-10}mbar. The voltage offsets on the transport multipoles were manually tuned to increase the transmission of large complexes (C-trap entrance lens; 0V, bent flatapole DC, 4 V; inter-flatapole DC 4V; injection flatapole DC, 4 V. An in-source CID voltage of 80 V and a HCD voltages of 20 V were required to achieve efficient MSP1D1 nanodisc desolvation and transmission. The instrument was set at a nominal resolving power of 70,000 at *m/z* 200 and mass spectra were acquired for 2 minutes by averaging 10 microscans per analytical scan. Data was analyzed using Xcalibur^{TM}2.2. No additional data processing (smoothing) was performed.

All FT-ICR experiments were performed using a 15 Tesla Bruker SolariX FT-ICR-MS instrument possessing an ICR infinity cell. The nESI capillary voltage was set to 0.9~1.2 kV. The temperature of dry gas was 100 ºC and the flow rate was 2.5 L/min. The RF amplitude of the ion-funnels was 300 V_{pp}, and the applied voltages were 210 V and 6 V for funnels 1 and 2, respectively. Skimmer 1 voltage was 50 V and the skimmer 2 voltage was kept at 20 V. The lowest values of RF frequencies were used in all ion-transmission regions: multipole 1 (2 MHz), quadrupole (1.4 MHz), and transfer hexapole (1MHz). Ions were accumulated for 500 ms in the hexapole collision cell before being transmitted to the infinity ICR cell. The time-of-flight of 2.5 ms was used. Vacuum pressures for different regions were ~2 mbar for the source region, ~$2e^{-6}$ mbar for the quadrupole region, and ~2×10^{-9} mbar for the UHV-chamber pressure. Mild collisional activated dissociation (CAD) was performed in the hexapole collision cell by colliding ions with SF_6 using a voltage of 30 V. 100 scans were averaged for each spectrum and recorded at 256 k data points unless specified otherwise. The MS Control software was Compass solariXcontrol, version 1.5.0, build 103, and the data were apodized using a full-sine-bell function and presented in magnitude mode. The mass spectrometer was externally calibrated with a 50 μg/μL solution cesium iodide in 1:1 (v:v) acetonitrile:water over the *m*/*z* range 100 to 20,000.

All LC-MS data was acquired on an Agilent 6230 TOF LC/MS system with a 1290 Infinity LC system. Chromatographic separation was achieved using a Zorbax SB300-C8 3.5 μm 2.1 x 50 mm column operated at a temperature of 75 °C. The solvents used were as follows: mobile phase A was water containing 0.1% v/v TFA. Mobile phase B was 90% n-propanol containing 0.1% v/v TFA. Initial gradients conditions were 20% mobile phase B from 0.0 to 1.0 minutes; 1.0 to 9.0 minutes, 20-70% mobile phase B; 9.0-10.0 minutes, 70-100% mobile phase B, where it remains at 100% for 1 further minute. The flow rate was 0.2 mL/min. Approximately 5 ug of IgG1-biotin conjugate was loaded on to the LC-MS system for each analyses. Data was acquired over the *m/z* range 1000-7000. The source fragmentor, skimmer and octapole 1 RF values were: 460 V, 95 V and 800 V (peak-to-peak) respectively. The ESI capillary voltage was 5.9 kV. Gas temperature was 340 °C. Drying gas was 13 L/min. Nebulizer was 25 psig. Oa-ToF calibration was performed using the Agilent Tune Mix using the automated calibration procedure implemented through MassHunter Data Acquisition version B.06.01, Build 6.01.6157.

Additional Deconvolution Algorithm Details

PMI Intact resamples the input MS spectra, which typically have wider m/z spacing at higher m/z, to produce uniformly sampled MS spectra. The spacing for the uniformly sampled spectra can be set by the user, typically about equal to the finest spacing in the input spectra, for example, 0.01 Thomsons, and resampling uses linear interpolation to determine values at m/z's between input sample points. PMI Intact then uses an iterative algorithm to deduce the mix of charges (the "charge vector") in each small interval of the uniformly sampled m/z spectrum. Intervals are typically set to about 0.6 Thomson ("charge vectors spacing") to match the isotope spread of a large highly charged molecule, but generally any value from 0.2 to 2 works equally well. For each interval, all charge probabilities are set equally likely for the first deconvolved mass spectrum. On each iteration, the algorithm updates "charge vectors" c_i (z), which give the probabilities that the i-th point (x_i, y_i) in the observed m/z spectrum takes the charges z=1, 2, ..., up to some maximum user defined charge. The charge vectors give the new neutral mass spectrum by accumulating c_i(z) * y_i values into the mass spectrum at the points closest to z*x_i - z*1.0073, where 1.0073 is the mass of a proton. New charge vectors are determined by a function that blends the intensity of the latest mass spectrum at z*x_i - z*1.0073 with a bonus for smooth charging of points in the neutral mass spectrum, and a "parsimony" penalty for charge vectors with probability spread over many charges._PMI Intact applies this "parsimony" bias, because multiple true masses mapping to the same m/z bin are less common than deconvolution artifacts caused by charge uncertainty. These bias down-weights the probability for each charge, except the likeliest charge. The smooth charging bonus can also be applied directly to the charge vectors (rather than to the neutral mass spectrum) by comparing c_i(z) with c_h(z) where c_h is the charge vector for point (x_h, y_h) satisfying x_h = (z – 1)*(x_i – 1.0072)/z and also with c_j(z) where c_j is the charge vector for (x_j, y_j) satisfying x_j = (z+1)*(x_i – 1.0073)/z. To bonus for smoothness, c_i(z) is increased if c_h(z) and c_j(z) are both significantly larger than zero. After applying parsimony and/or smooth charging biases, charge vectors must be renormalized so that for each i, c_i(z) sums to one over all choices of z. Parsimony was described in the JPR article with this passage: "for each i the intensity at m/z point m_i is more likely to derive from a single mass value than from two masses, more likely to derive from two masses than from three, and so forth." This passage really does give the key idea. Many implementations of the parsimony idea seem to work well to speed up convergence and reduce artifacts relative to the same iterative algorithm without parsimony.

Here we give an implementation that works well. This implementation uses a schedule of multipliers: 1, c, c^2, c^3, c^4, … , where $c < 1$ and c^{k-1} gives the a priori probability that k distinct masses will all land at the same m/z. The k-th largest mass contributing to m_i has its charge probability adjusted by multiplying by c^{k-1}. After multiplication, charge probabilities are normalized to sum to 1. We picked c based on what we believed to be the best results on a training set.

For polydisperse targets such as nanodiscs, the software uses a comb filter to set charge probabilities for m/z value x based on the probabilities at $x \pm j \times \text{KnownMassDelta}$, for j = 0, 1, … , CombFilter, where CombFilter is a user-supplied width (number of "teeth") for the comb filter, and KnownMassDelta is a user-supplied mass delta for the repeating units, for example, 677.5 Da for a nanodisc lipid. The comb filter was added in what we called the "backwards step". A comb filter of width 1 is implemented as an averaging filter with weights 0.25, 0.5, 0.25 applied to points in the last neutral mass spectrum at masses m – Δ, m, and m + Δ. The averaged value is then used to set the probability for charge k at m/z point $m_i = 1.0073 + m/k$. A comb filter of width 2 uses a weighted average of m –2Δ, m – Δ, m, m + Δ, and m + 2Δ. The software allows multiple comb filters of various widths to accommodate multiple expected mass deltas. One set that works well for many glycoproteins is 291.3 (for NeuAc), 365.3 (for HexNAc-Hex), and 656.6 (for HexNAc-Hex-NeuAc), all with width 1.

PMI Intact Mass has only three filters: a Gaussian smoothing filter optionally applied to the input m/z spectrum, a Gaussian smoothing filter optionally applied to the m spectrum after the iterative algorithm has finished, and the comb filter described above applied within the iterations.

Deconvolution can also be performed on text (m/z versus intensity) and csv files.

Protein Metrics has shown deconvolution of synthetic and semi-synthetic spectra in posters at ASMS 2017 titled: Intact Mass Analysis and DAR Calculation for Antibody-Drug Conjugates [2] and HUPO 2018 titled: Charge Deconvolution of Hard-to-Deconvolve Mass Spectra [3]. We would point the reviewer to these poster publications. They can either be downloaded from the conference website ASMS Proceedings) or requested directly from PMI.

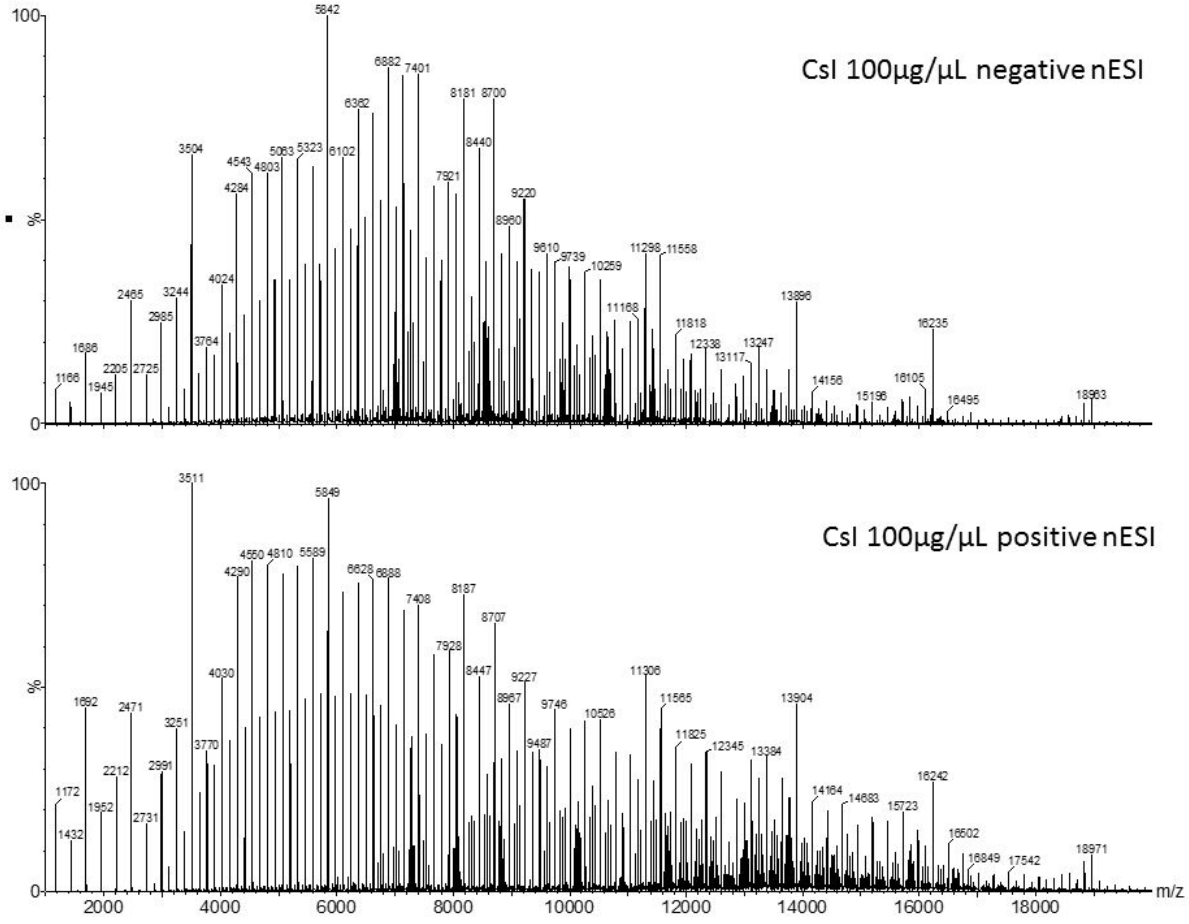


Figure S2. Typical CsI (100 µg/µL, 50% v/v acetonitrile) acquisitions used for *m/z* scale calibration, in both positive and negative modes of nESI.

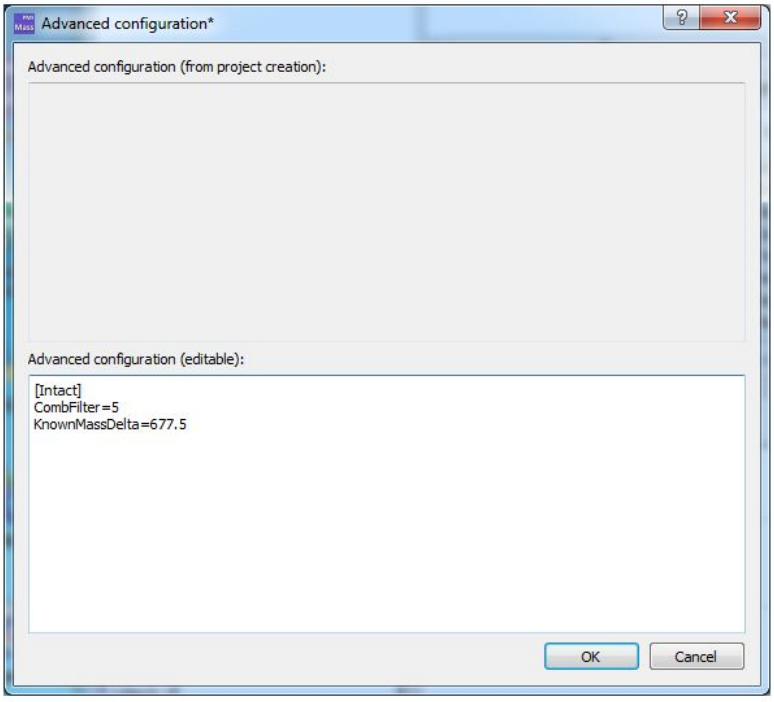


Figure S3. Configuring the Comb Filter in the Advanced Configuration settings. In order to efficiently process nanodisc native solution and mass spectrometry data, a "comb filter" step was added to the deconvolution algorithm. The comb filter “scans” for a series of *m/z* peaks with spacing corresponding to known mass deltas. If the known mass delta is d and the filter radius is set to 2, then the filtered intensity for charge z at *m/z* x is a weighted sum of the intensities at the five *m/z* bins x - 2d/z, x - d/z, x, x + d/z, and x + 2d/z. The weights sum to one in order to preserve total intensity. For example, the weights may be 0.1, 0.2, 0.4, 0.2, 0.1 for the five points. To avoid ripple artifacts, that is, masses beyond the true end of the mass peak series, a mild smoothness requirement was added for the set of *m/z* bins by up-weighting the central peak and down-weighting side peaks for lack of smoothness. A narrow comb filter with radius 2 (that is, 2 consecutive masses) worked well for glycoproteins with series of peaks spaced at 162 Da (galactose; data not shown) and a broad comb filter with radius 5 or 6 works well for highly polydisperse ions/molecules such as nanodiscs.

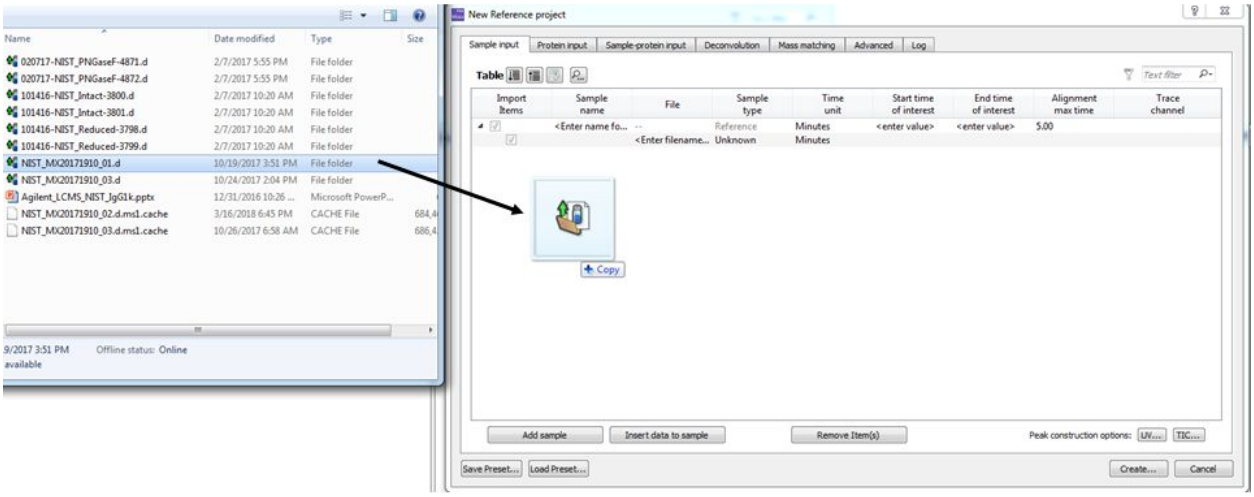


Figure S4. Native-MS infusion or LC-MS based data files for deconvolution are “dragged” in to the PMI user interface directly from any directory (server backup storage for instance).

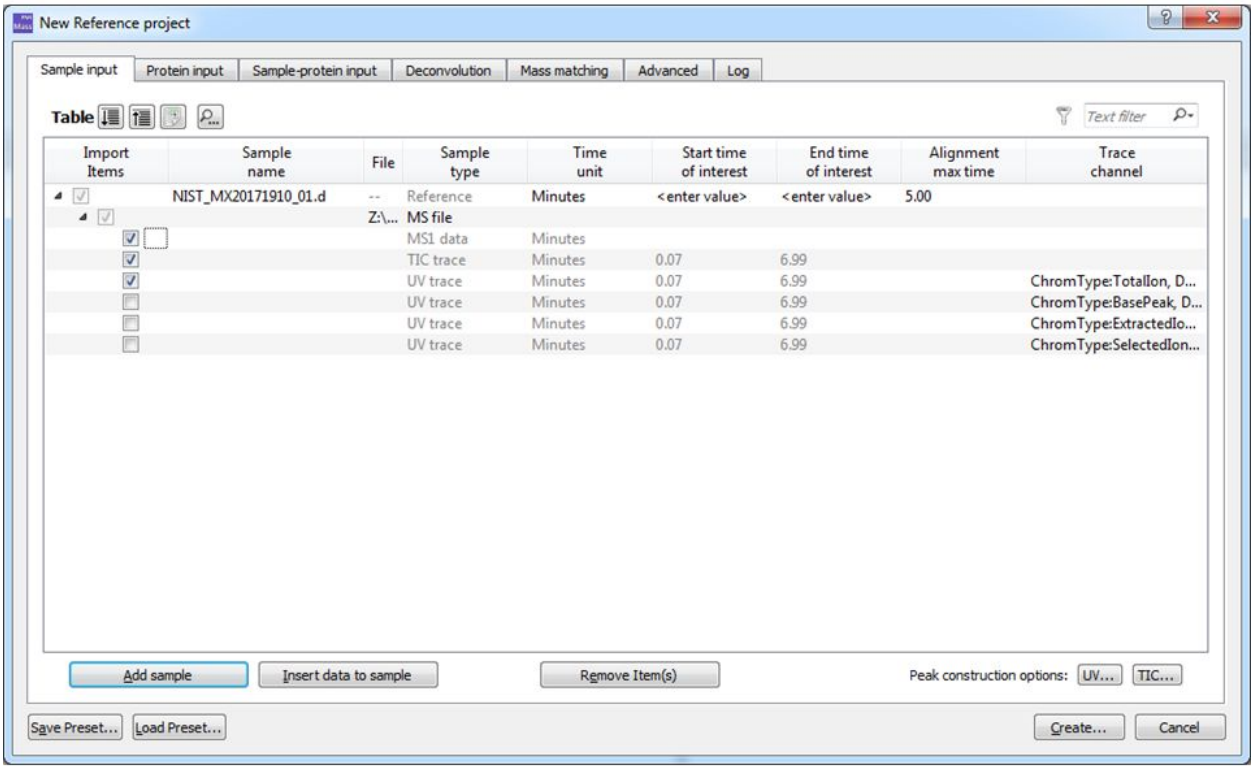


Figure S5. Once the files have been loaded into a new Reference Project, they are ready to be processed/deconvoluted.

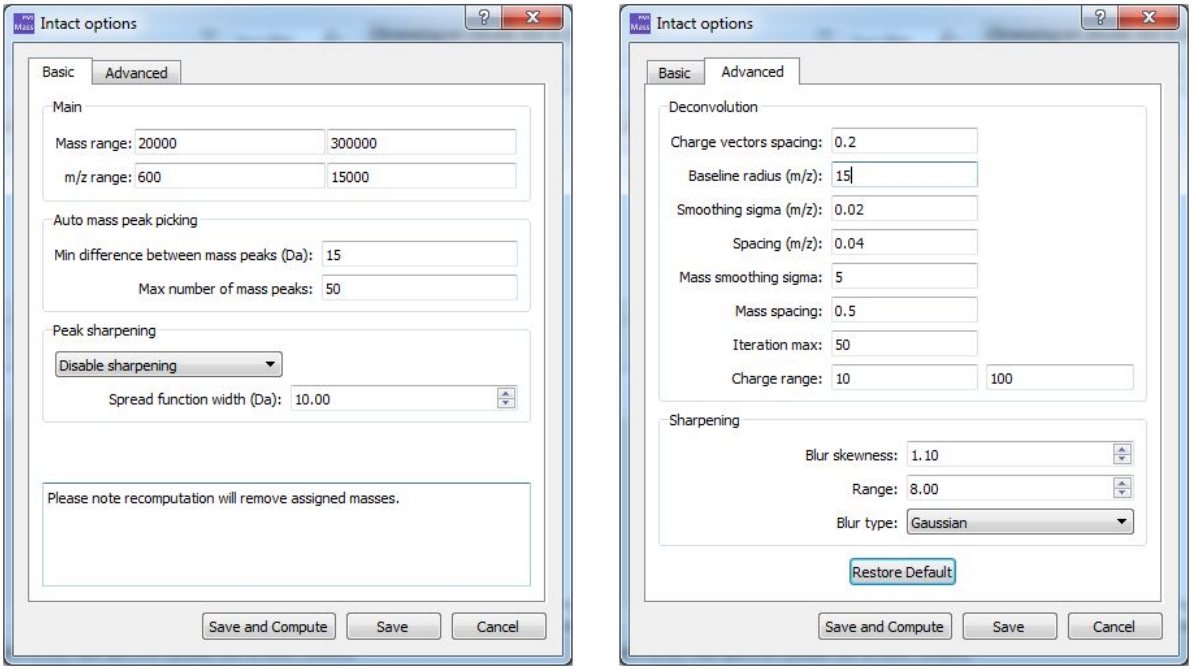


Figure S6. Specific Basic and Advanced deconvolution parameters. Typically, for native-MS nESI acquisitions when the S/N and overall signal is lower than that achieved through traditional denaturing LC-MS experiments, therefore the Mass Sigma Smoothing option is generally increased to 25-50, which was the case for the GroEL data displayed in Figure 1d and Figure S10.

Basic deconvolution values used for all spectral processing were typically: Mass Range 20,000-300,000 (and up to 1,000,000 for GroEL). The lower MW range may be reduced for smaller proteins such as those described in Figure S12; *m/z* range 600-15,000; Charge Range 10-100; Iteration Max 50.

Advanced Deconvolution Parameters

Charge vectors spacing: Typically 0.2. This parameter sets the spacing of *m/z* intervals for charge assignments. A smaller number such as 0.1 can split interleaved isotope-resolved peaks; a larger number such as 1.0 works well for isotope-unresolved peaks. For most data, any setting from 0.2 to 2 works equally well.

Baseline radius (*m/z*): Typically 15. This parameter controls the stiffness of the curve for baseline removal. A small number such as 10 will sharpen the peaks in the deconvolved spectrum but may give inaccurate quantitation. A large number such as 100 will avoid cutting into broad *m/z* peaks. This can be considered as a baseline subtraction.

Smoothing sigma (*m/z*): Typically 0.02. Allows Gaussian smoothing of the *m/z* spectrum before deconvolution. Smoothing in *m/z* translates unequally to mass due to widening by charge multiplication, so with the correct choice of parameter, isotope resolution may be retained on a 10,000 Da mass while noise fluctuations can be removed on a 150,000 Da mass.

Spacing (*m/z*): Typically 0.04. This value should closely represent the spacing of data points in the *m/z* spectrum. It should be set to approximately as fine as the *m/z* spacing of the *m/z* of the signals in the raw data; for example, 0.005 for Orbitrap with signal below *m/z* 1000; 0.01 for Orbitrap with signal in the *m/z* 1000 – 2000 range; 0.02 for Q-TOF, and 0.05 for native MS on an Orbitrap with signal around *m/z* 5000. Wider spacing can save computation time, but with the loss of resolution in the deconvolved spectrum.

Mass smoothing sigma: Typically 5.0*. Allows Gaussian smoothing of the mass spectrum. For isotope-resolved masses, 0.1 is a good choice, but for masses of 150,000 Da, a good choice is 5.0. *The mass smoothing sigma was increased to 50 for the native MS GroEL data.

Mass spacing: Typically 0.5. The spacing of points in the *m/z* spectrum. For an instrument with resolution 20,000 at *m/z* 2000 and a signal centered at 2000, a good choice would be 0.1. For masses at 150,000 Da, a good choice is 0.5.

Iteration max: Typically 25. The maximum number of iterations in the charge deconvolution algorithm. 10 is sufficient for most applications, but 25 may give finer details.

Charge range: Typically 5 to 100. The range of charges considered in charge deconvolution. These are minimum and maximum values, so these numbers must be changed in order to consider charges below 5 or above 100.

Options 2 and 3 are added directly in to the Advanced Configuration (editable; Figure S3) option.

1. Baseline removal from the m/z spectrum (Baseline Radius m/z, under Advanced Settings).

2. Optional removal of charge-one peaks. RemoveChargeOne=1

3. Optional removal of baseline from the m spectrum. EnableBaseline Removal=1

Denatured Versus Native MS Theoretical and Instrument Derived Peak Widths

Figure 1a, the NIST IgG1*k* mAb standard analysed in positive nESI mode; Figure 1b, an IgG1 mAb covalently modified with sulpho-NHS-biotin (Scheme 1a; 10-molar equivalents [4]) via native lysine residues (Scheme 1); Figure 1c, the homotetrameric membrane protein aquaporin-Z (AqpZ) displaying a polydispersity of n=0 to 4, and Figure 1d, the homotetradecameric chaperone complex GroEL, analysed in the rarely described negative nESI mode. In all cases a high quality zero-charged spectrum is obtained, highly consistent with the unprocessed data (insets) in terms of expected peak width (apparent spectral resolution), overall spectral signal-to-noise, level of baseline between charge-states and in-spectral ratio of the polydispersity distribution. For example, the peak widths (FWHM) for the deconvolved data are systematically broader than the theoretical peak widths. The theoretical peak width (FWHM) for the most intense zero-charged glycoform (NIST IgG1*k*, G0F/G1F, MW 148,199.3 Da, Figure S7, Supporting Information) is 19.0 Da and 31.9 Da modeled with the instrument resolutions of 14.6k (LC-MS oa-ToF) and 5.4k (native-MS FT-ICR at *m/z* 5459.3, based on the CsI cluster $[Cs_{21}I_{21}]Cs^+$) respectively. The measured peak widths (FWHM) following deconvolution are 31.0 Da, 19.2 Da and 40.4 Da for the LC-MS oa-ToF, LC-MS oa-ToF peak sharpened and the native FT-ICR, respectively. Interestingly, the deconvolved peak width for the fully denatured NIST mAb is still significantly broader than the theoretical peak width (31.0 Da vs 19.0 Da). This difference in peak width is also observed for the individual charge states (Figure S8, Supporting Information) and is likely caused by low level isobaric interferences such as oxidation or hydroxylysine/proline (+15.9 Da), trisulphides (+32.0 Da), hydrolysis (+18.0) for example, or near isobaric such as the additional heavy chain C-terminal lysine residue (+128.1 Da [5]) and loss of a single GlcNAc (-203.1 Da [5]), both of which become partially resolved upon peak sharpening *(vide infra*). Algorithmic peak sharpening results in a spectral peak width highly consistent with the theoretical isotopic peak width (19.2 Da vs 19.0 Da). The NIST mAb analyzed on the FT-ICR was acquired under native solution conditions (200 mM ammonium acetate), therefore a higher level of adducting is observed [4] resulting in peak asymmetry and offset to higher MW. One must also note that the FT-ICR data is apodised using a basic Full-Sine Bell function. Peak width can be further improved using different apodization functions [4, 6], or absorption mode processing [6, 7], however this will not be described further herein.

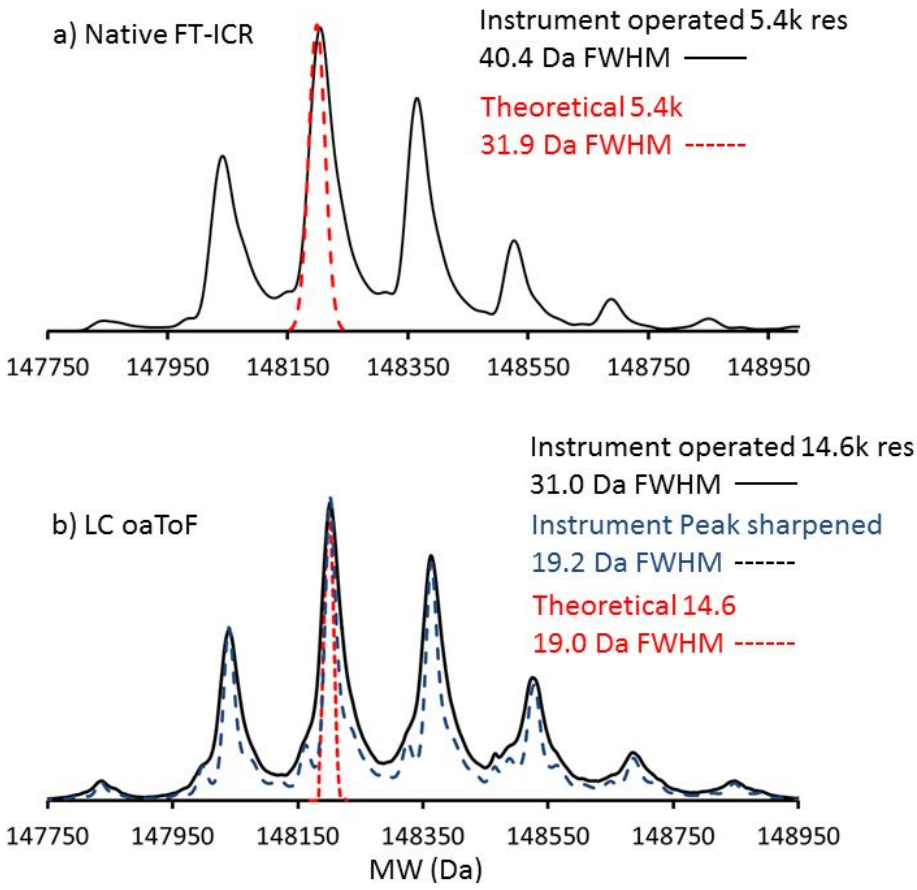


Figure S7. Overlays of theoretical versus instrument derived NIST IgG*k* peak width (FWHM) for: a) native MS acquired on the FT-ICR and: b) denaturing LC-MS acquired on the LC oaToF system. Peak widths at FWHM are annotated. The theoretical peak widths for the zero-charged G1F glycoform were calculated using IsoPro 3.1 (https://sites.google.com/site/isoproms/).

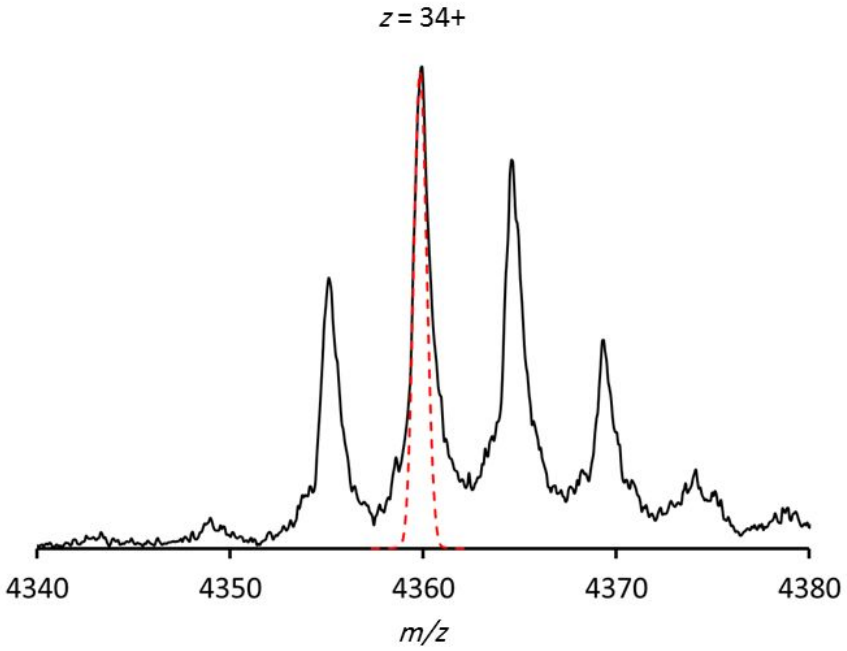


Figure S8. Comparison of the instrument derived (LC-oaToF) peak (black) widths and the theoretical NIST IgG1*k* peak widths (red hashed) for the charge state z = 34+. Instrument derived FWHM = 0.95 Th. Theoretically derived FWHM = 0.79 Th. Theoretical peak width for charge state 34+ was modelled using MassLynx 4.1.

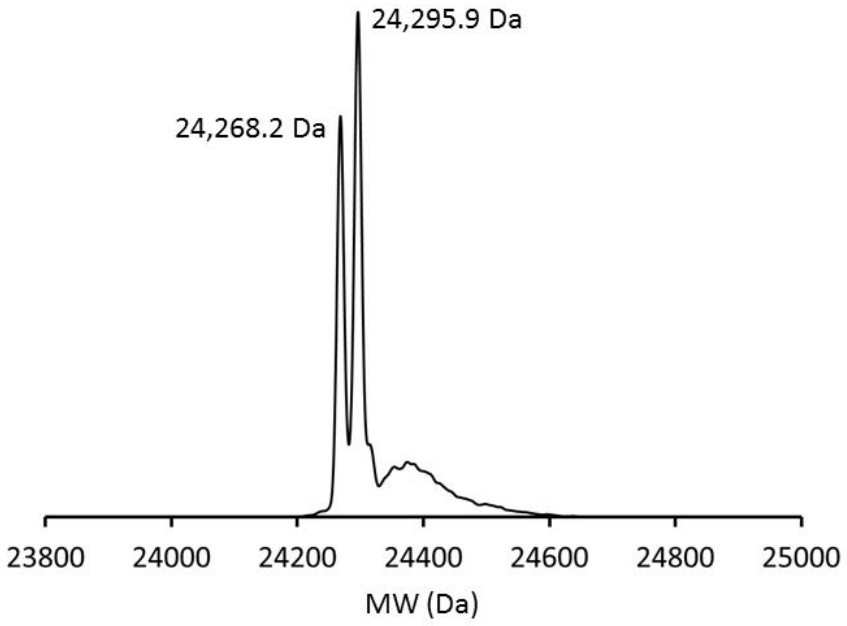


Figure S9. Deconvolved AqpZ acquired under denaturing LC-MS conditions.

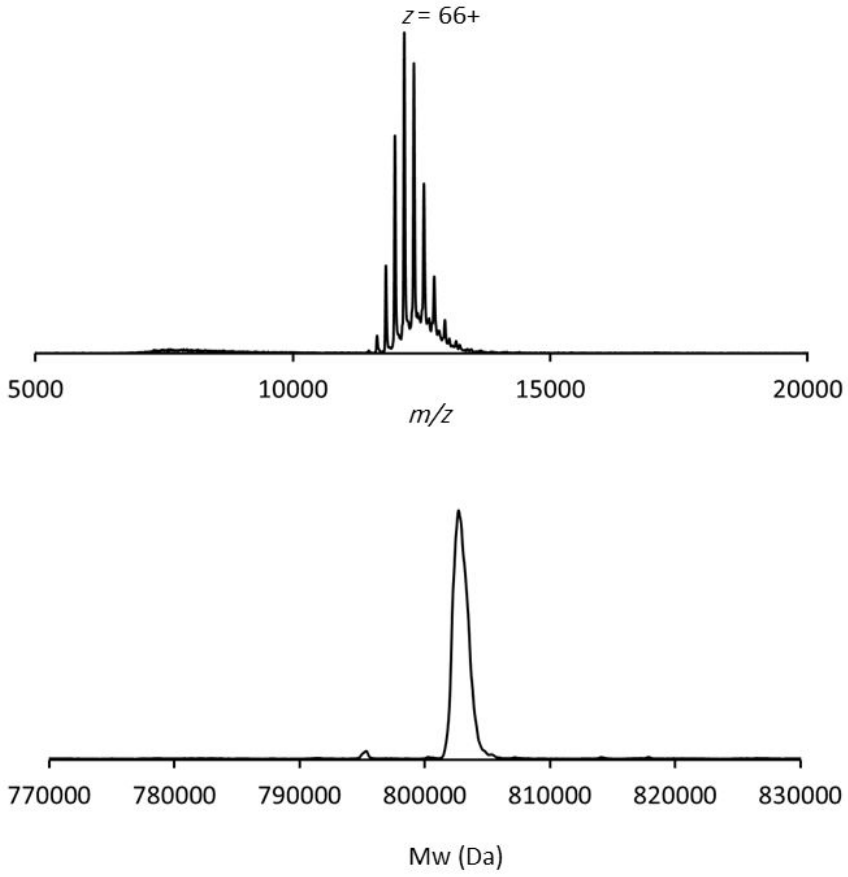


Figure S10. Native solution (200 mM ammonium acetate) and MS of GroEL analysed under positive ion nESI mode. The upper spectrum is the unprocessed charge state distribution. The lower spectrum is the PMI deconvolved data.

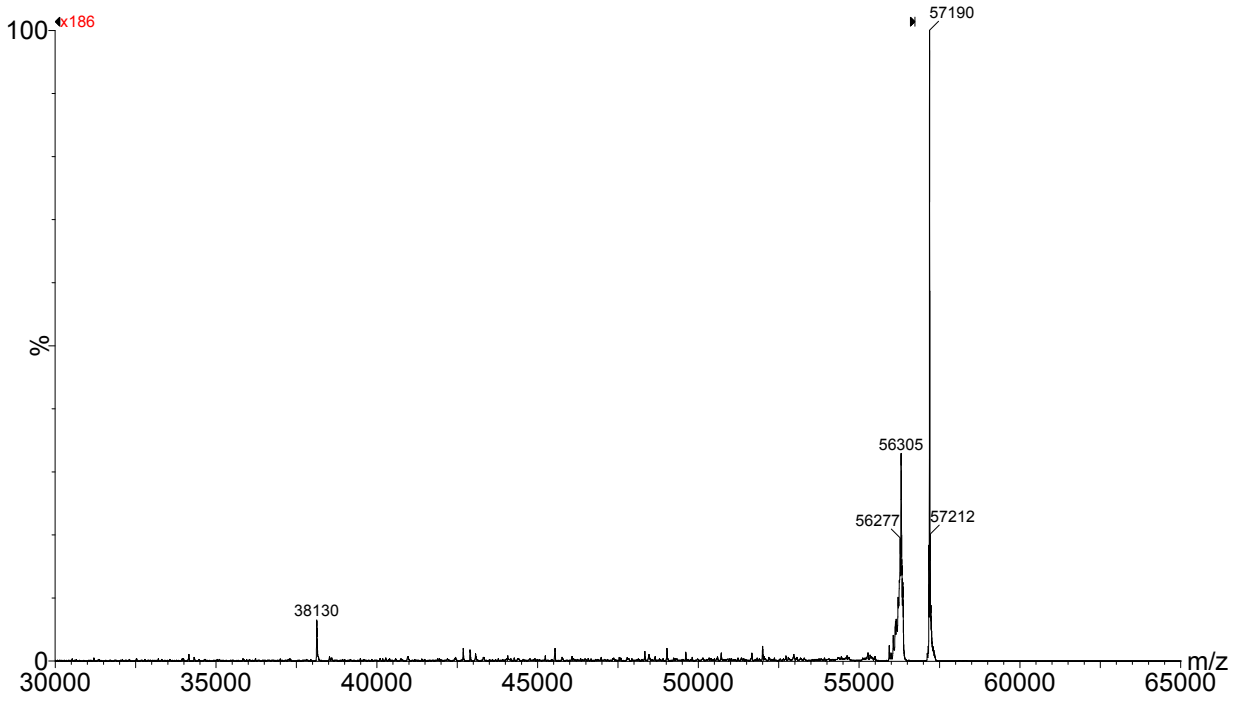


Figure S11. Deconvolved GroEL acquired under denaturing LC-MS conditions. The *m/z* range below the base-peak at 57,190 Da has been magnified by 186x. Low level species are detected at 38,130 Da and 56,305 Da. Additionally, an earlier eluting peak within the LC-MS chromatogram (data not shown) deconvolves to a MW of 52,094 Da.

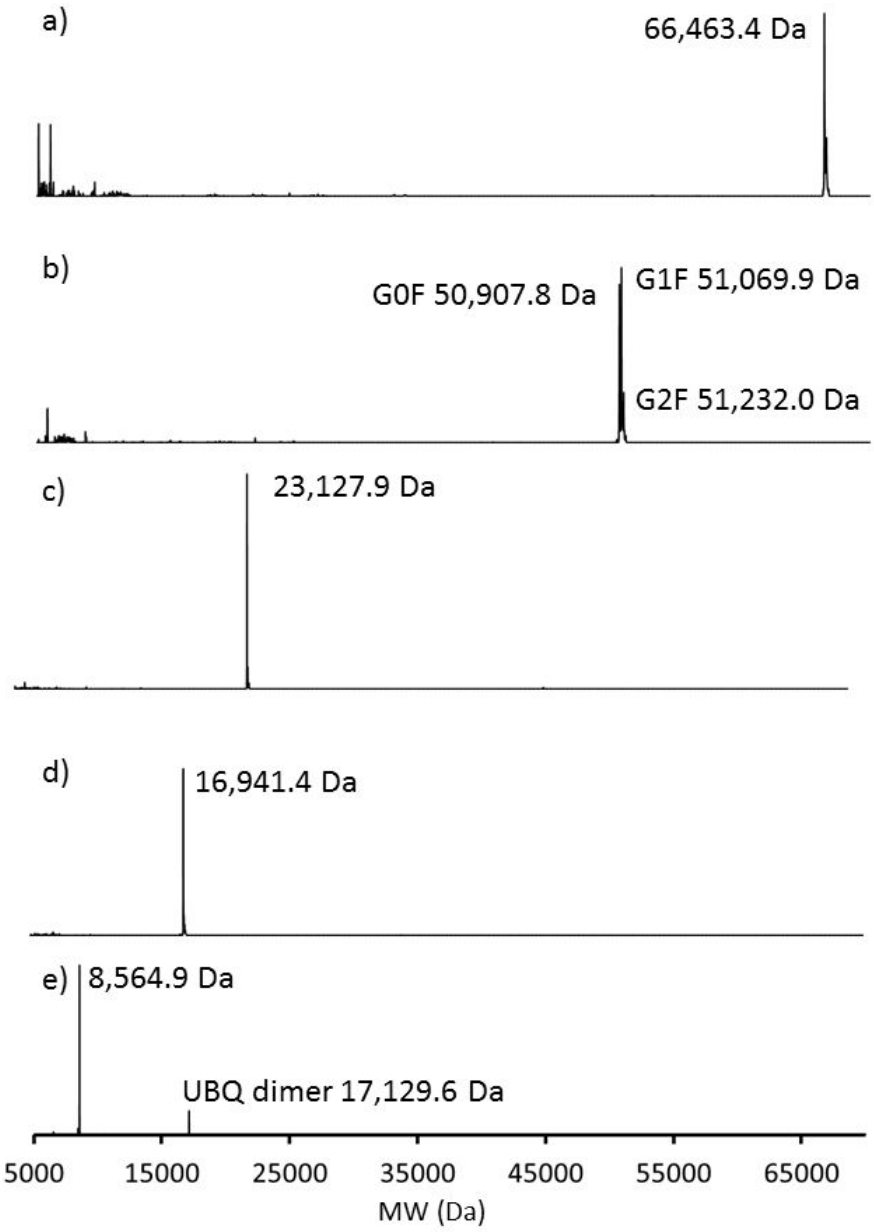


Figure S12. Deconvolved LC-MS analysed a) BSA Fraction V Calbiochem #12657; b) reduced NIST heavy chain; c) reduced NIST light chain; d) horse heart myoglobin Sigma-Aldrich M0630-1G and e) ubiquitin Sigma-Aldrich U6253-25MG. The BSA (NP_851335.1) is a natural variant containing an A to T residue swap at position 214.

	PMI Derived (Da)	Theory (Da)	Error (ppm)
BSA	66463.4	66463.2	2.4
NIST HC G0F	50907.8	50907.51	5.7
NIST HC G1F	51069.9	51069.65	4.9
NIST HC G2F	51232	51231.79	4.1
NIST LC	23127.9	23127.83	3.0
Myo	16951.4	16951.5	-5.8
UBQ	8564.9	8564.85	5.8
			RMS 4.7

Table S1. Deconvolved MWs verses theoretically calculated zero-charged MWs. Theoretical MWs were calculated through MassLynx 4.1. Note that this mass accuracy was achieved on an LC-MS oa-ToF (Agilent 6224) system operating at a modest resolution (14.5k FWHM).

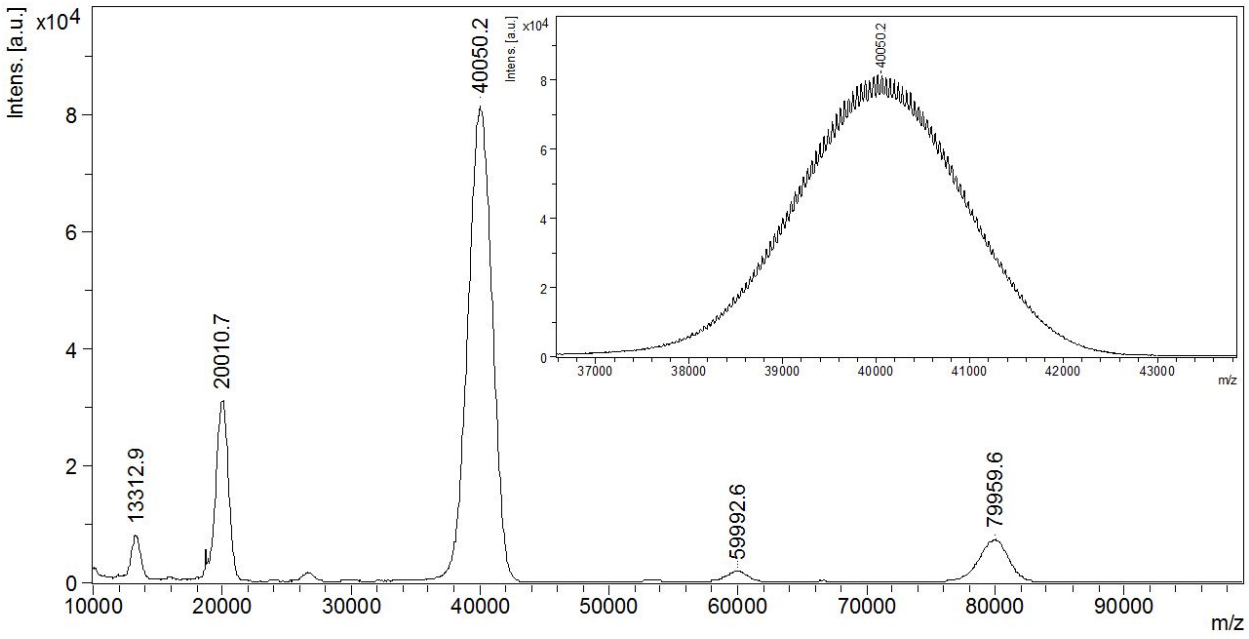


Figure S13: Linear MALDI-MS of GCSF-PEG. 2,5-Dihydroxyacetopheone (2,5-DHAP) was purchased from Bruker Daltonic. Diammonium hydrogen citrate (DAC) was obtained from Sigma. The matrix was prepared by mixing 7.6 mg of DHAP in 375 µL of ethanol with 125 µL of DAC (18 mg/mL). For MALDI mass measurement, 2 µL of PEG-GCSF (1 mg/mL) was thoroughly mixed with 2 µL of matrix and 2 µL of 2 % TFA; 1 µL was spotted onto the target plate and dried at room temperature prior to introduction into the mass spectrometer. A Bruker UltrafleXtreme TOF/TOF was employed and the instrument was calibrated externally using bovine serum albumin. Spectra were obtained at positive linear mode with an acceleration voltage of 25 kV. Eight thousand laser shots per analysis were collected usually at a laser shot rate of 2 kHz. The time of flight was converted to mass using the Flex software supplied with the instrument.

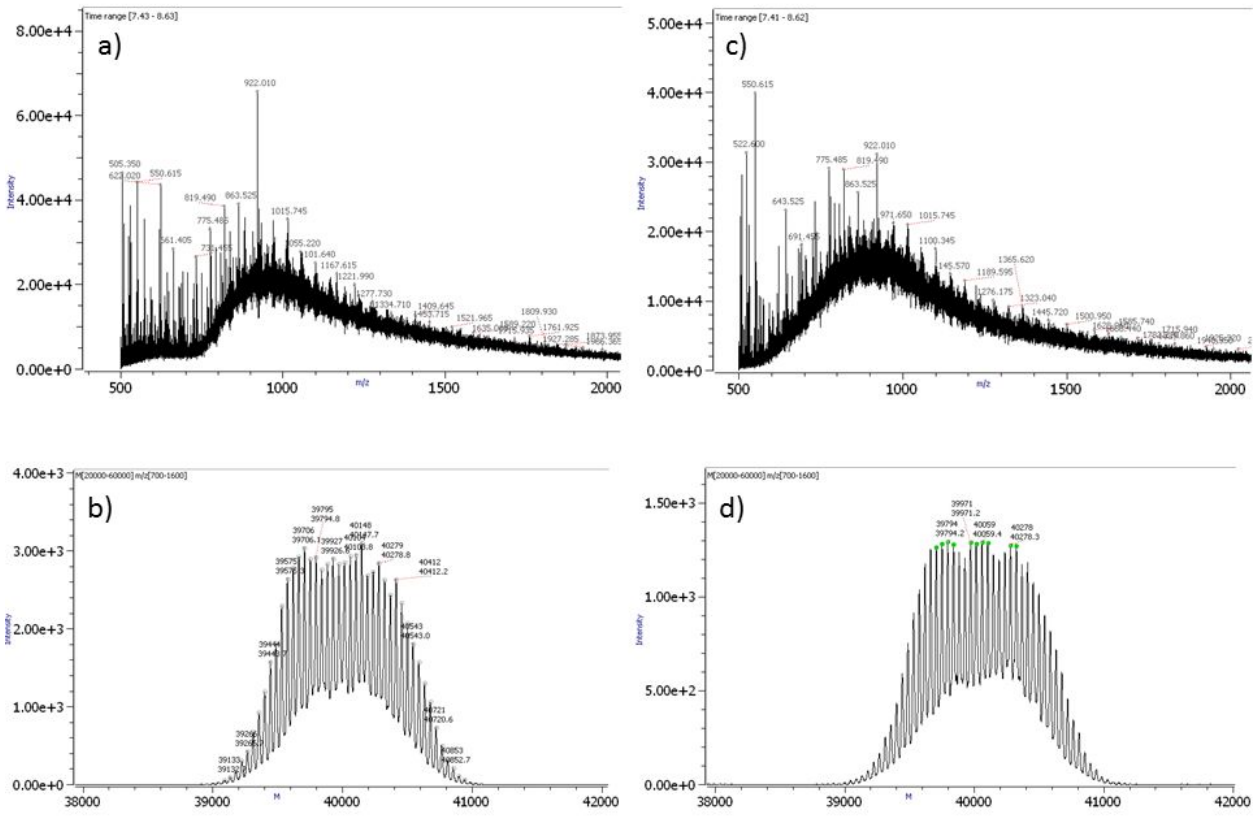


Figure S14. Denaturing LC-ToF MS of GCSF-PEG at a) fragmentor voltage of 230V; b) the deconvolved 230V spectra; c) fragmentor voltage of 180V; d) the deconvolved 180V spectra. It must be noted that the deconvolved data reflects the subtly different unprocessed data. In Figure 2c (main text) there is evidence of a higher MW shoulder (40.5 kDa). However, at lower source fragmentor voltages (Figures S14b & d) this feature is reduced. We infer that this feature is a result of low level in-source fragmentation resulting in higher populations of lower MW species. At lower source voltages, the MW distribution is more equally distributed.

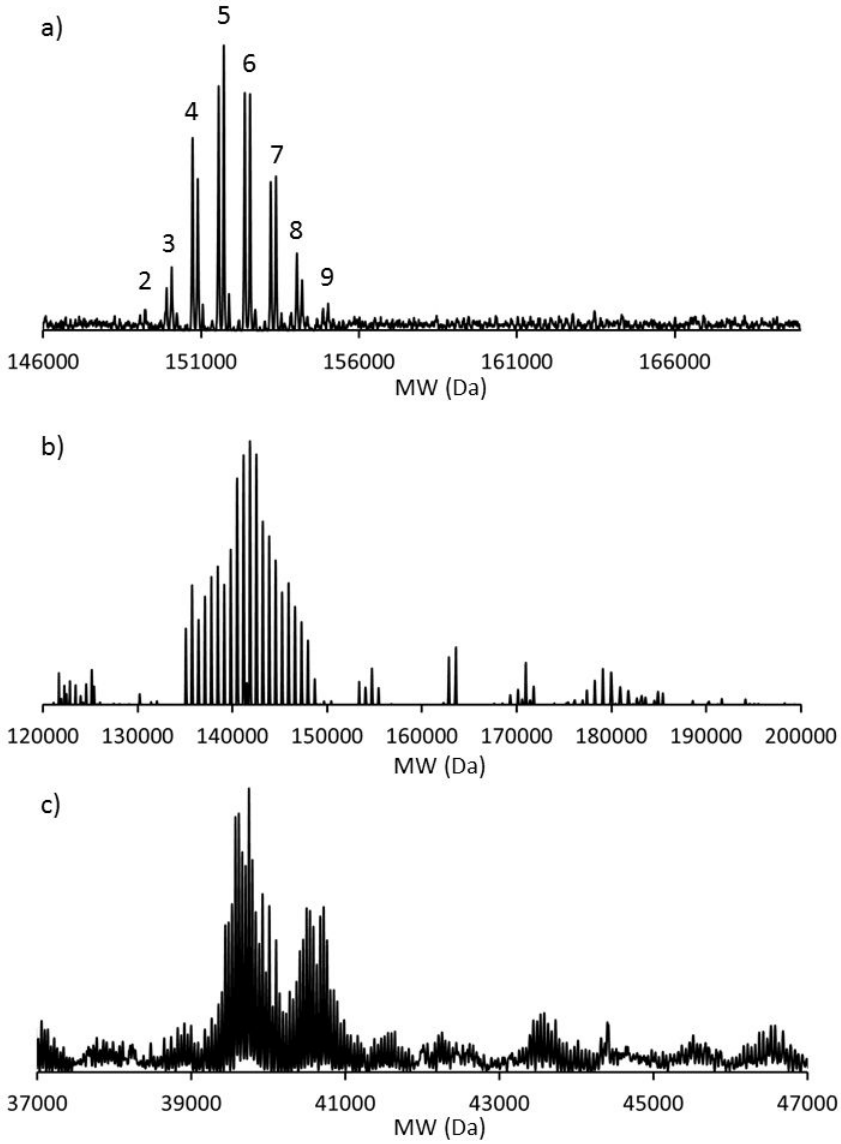


Figure S15. Deconvolution of a) IgG2-PEG_{12}-Biotin conjugate analysed by denaturing LC-ToF MS; b) empty MSP1D1 nanodisc analyzed by native-MS (Orbitrap-EMR) and c) GCSF-PEG analysed by denaturing LC-MS ToF, without the use of the Comb Filter. The DAR species are annotated for the IgG2-PEG_{12}-Biotin conjugate. In all cases the S/N is improved when using the Comb Filter: compare to Figures 2a, b & c in the main text.

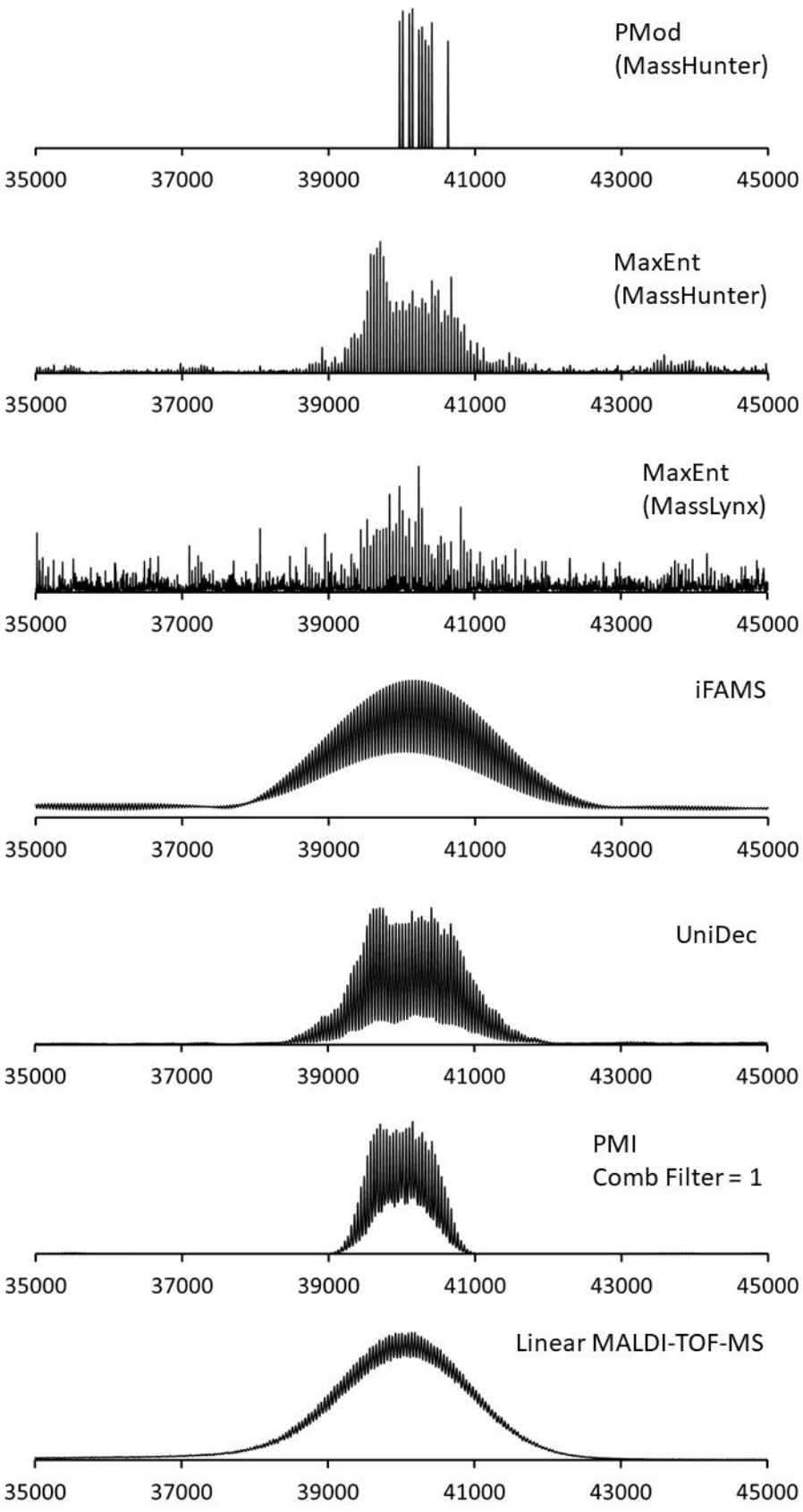

PMod
(MassHunter)
35000
37000
39000
41000
43000
45000
MaxEnt
(MassHunter)
35000
37000
39000
41000
43000
45000
MaxEnt
(MassLynx)
35000
37000
39000
41000
43000
45000
iFAMS
35000
37000
39000
41000
43000
45000
UniDec
35000
37000
39000
41000
43000
45000
PMI
Comb Filter = 1
35000
37000
39000
41000
43000
45000
Linear MALDI-TOF-MS
35000
37000
39000
41000
43000
45000

Figure S16. A comparison of PEG-GCSF 1+ ion measured and detected by linear MALDI-TOF and the neutral MW deconvolved spectra obtained from PMod (MassHunter, Build 7.0.7024.0), Maxent (MassHunter, Build 7.0.7024.0), MaxEnt (MassLynx v4.1 SCN566), iFAMS (v5), UniDec (v3.2.0) and PMI. The LC-MS spectral data for the PEG-GCSF (Skimmer 230V) molecule was selected as it represents the most complex and polydisperse molecule analyzed within this study.

From a purely qualitative perspective, iFAMS, UniDec and PMI all produce a deconvolved spectrum possessing significantly higher S/N than PMod or MaxEnt. The neutral MW spread of iFAMS, UniDec and PMI does differ. In addition to the deconvolved neutral MW, both iFAMS and UniDec output the charge state distribution of the raw data, which is highly useful for polydiscperse data where the charge state distribution cannot be manually derived. For the PEG-GCSF protein, iFAMS and UniDecss charge distribution determination is z = 27+ to 47+ and z = 30+ to 50+ , respectively.

The algorithm input parameters are as follows:

pMod (MassHunter): Deconvolution algorithm pMod; Mass Range 30,000 to 50,000; Mass Step 1.0; Use limited *m/z* range 700-2000; Subtract baseline factor 7.0; Adduct Proton; Peak Width ½ theoretical width; Peak filters all set to automatic.

MaxEnt (MassHunter): Deconvolution algorithm Maximum Entropy; Mass Range 30,000 to 50,000; Mass Step 1.0; Use limited *m/z* range 700-2000; Subtract baseline factor 7.0; Adduct Proton; Isotope width Automatic; Peak signal-to-noise >= 30.0; Maximum number of peaks 100; Minimum consecutive charge states 5; minimum protein fit score 8.

MaxEnt (MassLynx): Ranges 30000:50000; Resolution 1 Da/channel; Damage Model Uniform Gaussian Width at half height 0.5 Da; Minimum intensity ratios 75%/75%; Iterate to convergence.

iFAMS: For direct comparison to other analyses performed here on data from m/z 700-1500, a 1D Fourier transformation was performed and a charge state range of z=27+ to 47+ and subunit mass 44.05 (mass of PEG) was chosen to locate Fourier peaks (The charge state range is a means of limiting the m/z range based on initial spectral parameters i.e. neutral MW of the protein. Both

the first and second harmonics were used to reconstruct the charge state specific *m/z* functions, and build the zero charge spectrum.

UniDec: m/z range 700-1500; Subtract Curved 75; Bin every 1.0; Gaussian Smoothing 0.0; Intensity Threshold 0.0; Adduct Mass (Da) 1.0; Acceleration Voltage (kV) 0.0; Charge Range 10 to 100; Mass Range 30,000 to 50,000; Sample Mass Every (Da) 1.0; Peak FWHM 0.5; Peak Shape Function Gaussian; Charge Smooth Width 1.0; Point Smooth Width 1.0; Mass Smooth Width 1.0; Mass Difference (Da) 44.0; Maximum # of Iterations 50; m/z to Mass Transformation Inter; polation.

PMI Intact: See Figure S6 and Advanced Deconvolution Parameters section above.

MaxEnt and PMod (MassHunter) deconvolution were performed directly on the MassHunter data files. The PEG-GCSF spectral data was exported as an xy.txt file for further deconvolution using the additionally noted algorithms. A Masslynx .RAW file format was created from the data xy.txt file by using DBRIDGE.exe, that is located within the MassLynx file directory. MaxEnt deconvolution is then subsequently performed. The input for UniDec and iFAMS is the xy.txt file.

Additionally, for a comparison of complex and polydisperse nanodisc spectra, we refer this reviewer to the Supporting Information of Campuzano et. al. Analytical Chemistry 2016 88 (24) p12417-12436, where we compared UniDec to MaxEnt deconvolution and discussed the results within the accompanying Supporting Information.

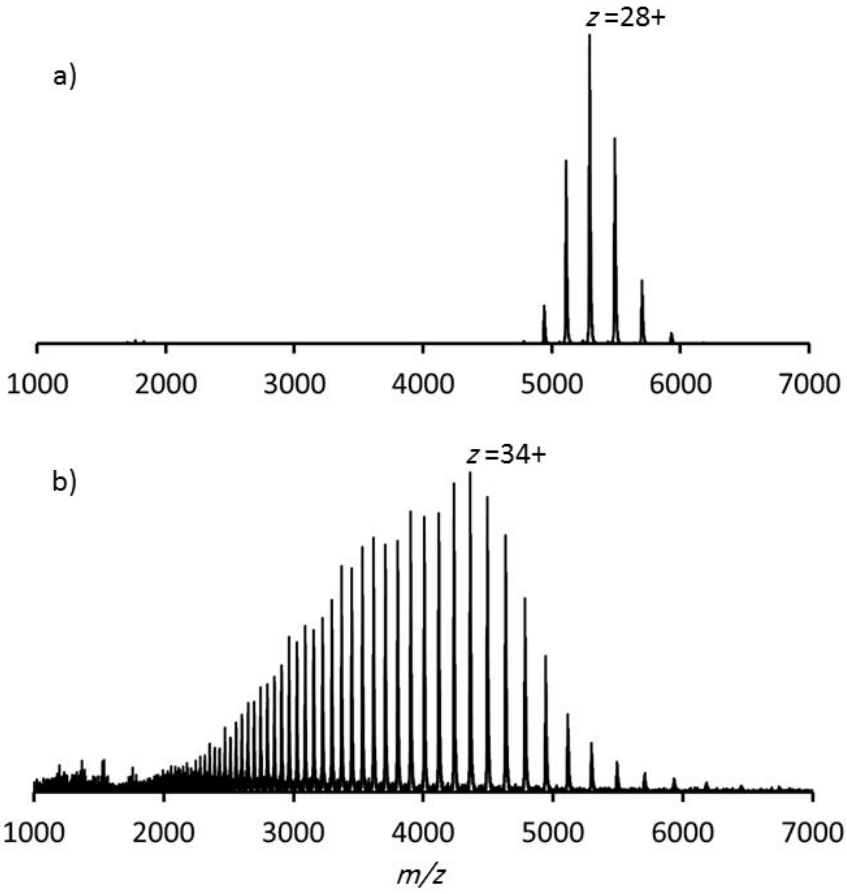


Figure S17. Comparison of the a) native FT-ICR and; b) LC-oaToF charge state distributions for the NIST IgG1*k* mAb. The charge (*z*) is also annotated on the most intense charge state.

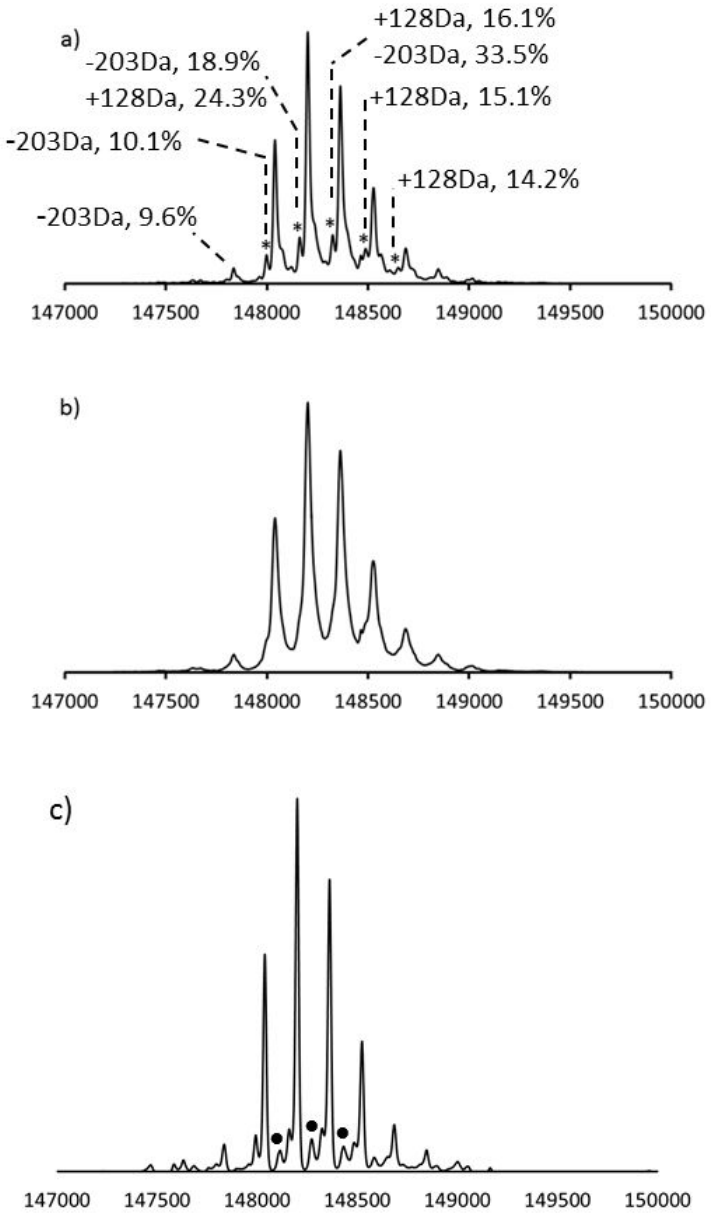


Figure S18. The effect of Peak Sharpening on the NIST IgG1*k* mAb acquired under denaturing LC-MS conditions. Enable Peak Sharpening; Spread Function Width 10 Da; Sharpening Blur Skewness 1.10; Range 8.00; Blur Type Gaussian. The uncleaved C-terminal heavy chain lysine residue [5] (+128.1 Da, annotated by *) are now partially resolved from the adjacent glycoform. Upper spectrum (a) has been peak sharpened. Middle spectrum (b) is the non-peak sharpened data; c) the effect of Peak Sharpening using a Blur Skewness value of 2.0.

The annotated peaks corresponding to the unprocessed C-terminal lysine residues (+128 Da;*) in Figure S18a present at relative quantitation values of 24.3%, 16.1%, 15.1%, and 14.2% to the adjacent glycoform. This value is comparable to the 8.7%, 6.5% and 6.1% derived from Figure 19a (reduced NIST heavy chain) and 6.9% derived from Figure S19b where the unprocessed lysine is fully resolved from the processed and reduced heavy chain, once treated with PNGaseF.

One can also consider the annotated peaks in Figure S18a as losses of GlcNAc (-203 Da) from the corresponding glycoforms, and are present at 9.6%, 10.1%, 18.9% and 33.5% relative to the corresponding glycoform, again comparable to the relative values derived from Figure S19a of 5.6%, 7.1% and 24.3% corresponding to the loss of GlcNAc (-203 Da) from the heavy chain analysed under reducing conditions.

It must be noted however, that the levels of +128 Da and -203 Da will differ when comparing levels between reduced heavy chain and the intact mAb. These comparisons serve as valid controls in to how well the Peak Sharpening function can identify and quantify (relative levels) partially resolved ion species.

The annotated peaks (●) are over represented in this deconvolved spectrum (Figure S18c) as compared to the unprocessed data which is displayed in Figure S8, and also those presented in Figure S18a. Additionally, the tailing edge/asymmetry of the glycoforms in both the Sharpened (Figure S18a) and non-Sharpened (Figure 18b) spectra are significantly reduced when the incorrect Blur Skewness value is used. Therefore, care must be taken when using the Peak Sharpening option.

PMI Intact includes an optional Richardson-Lucy peak sharpening to deconvolve a known "point spread function" from the neutral mass spectrum. Currently PMI Intact allows point spread functions with either Gaussian or Lorentzian tails, with user-definable width parameters (standard deviation for Gaussian and half-maximum for Lorentzian). PMI Intact provides another parameter called "Blur skewness", which is the ratio of the right tail to the left tail width parameter, so that skewness of 1.1 (the default) means that the right-hand tail is expected to be 10% longer and heavier than the left tail (the two tails are scaled to meet in the middle). Peak sharpening will turn shoulders into well-defined peaks, but if the point spread function is inaccurate, it can also produce artifacts such as split peaks or ringing around broad peaks. Also note that by deconvolving point

spread in the neutral mass spectrum rather than in the *m/z* spectrum, PMI Intact models instrument resolution (and adducts and isotopes) averaged over *m/z*, rather than resolution varying with *m/z*.

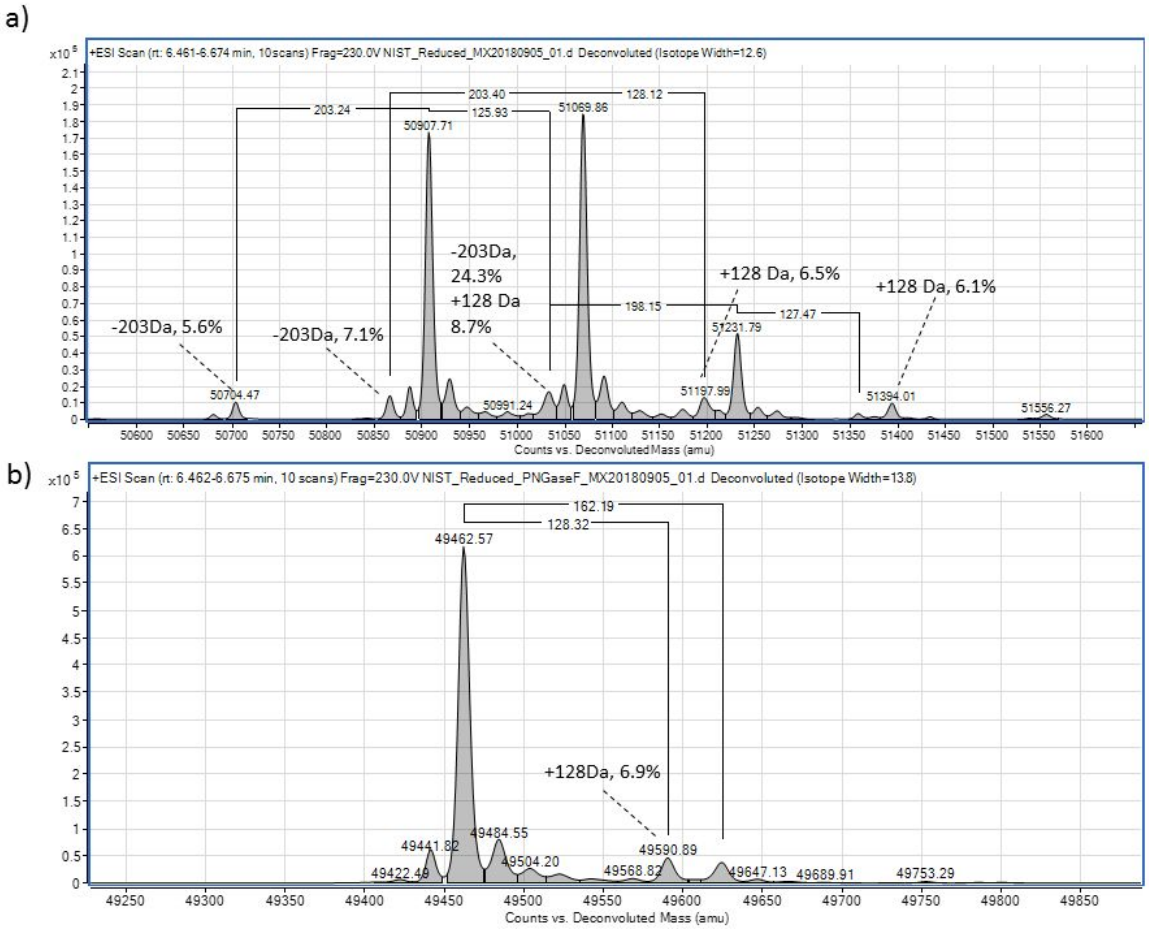


Figure S19 a) NIST IgG1*k* reduced HC data from denaturing LC-MS, displaying the -203 Da from each glycoform. G2F is lower intensity than G0 and G1F, therefore the intensity of the -203 Da is also lower and therefore the associated MW error is higher; an observed loss of -198 Da is detected; b) PNGaseF treated NIST HC data from denaturing LC-MS, displaying the +128 Da. Also detected is low level glycation (+162 Da). The specific relative percentage values for the -203 Da and +128 Da have been annotated and are consistent with those calculated/described in Figure S18.

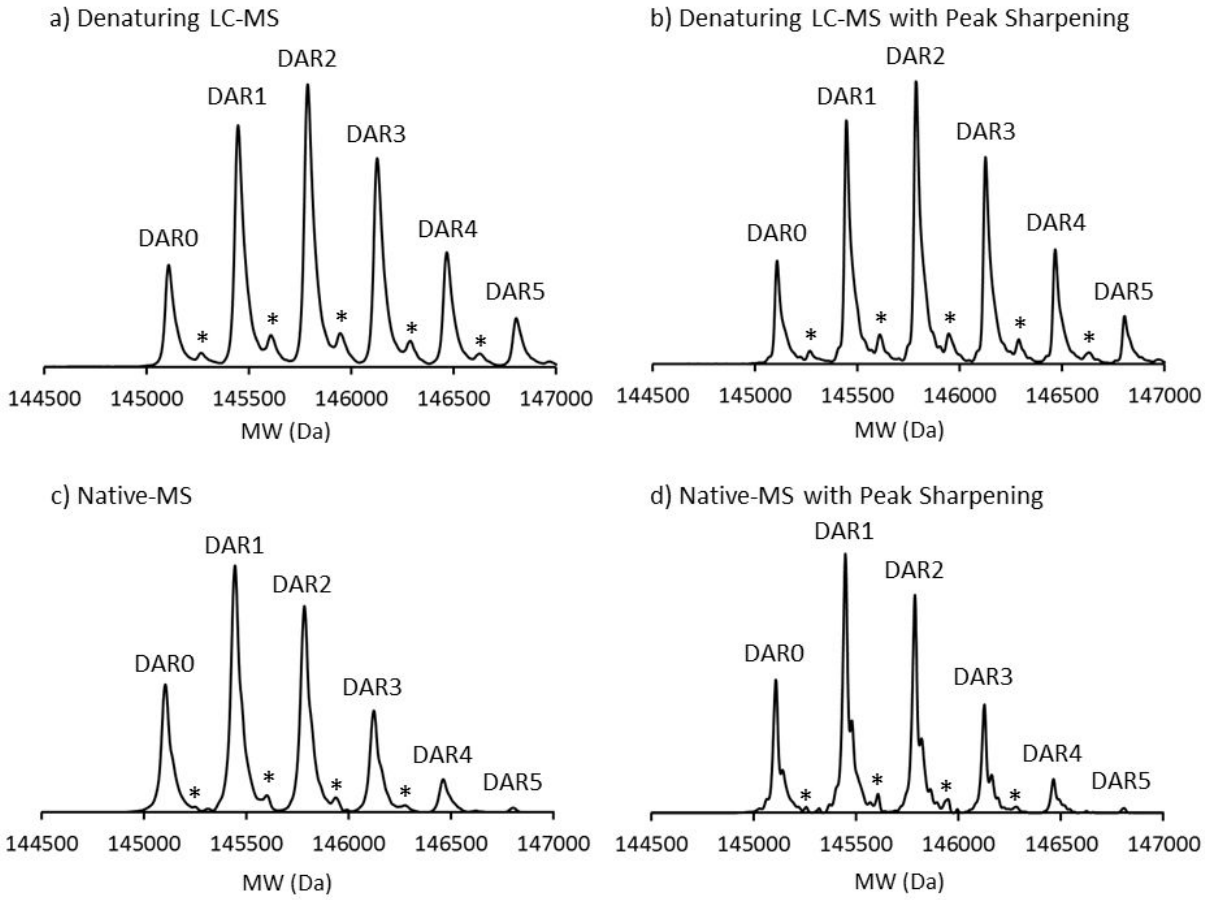


Figure S20. PNGaseF treated IgG1-biotin conjugate (5-molar biotin equivalents) analyzed under a) and b) denaturing LC-MS conditions by oa-ToF and c) and d) native-MS nESI and solution conditions (200 mM ammonium acetate) by FT-ICR. DAR values are annotated. * represents the low-level glycation (+162 Da). The effect of Peak Sharpening is most dramatic in the native-MS data set. Improved separation of the single low-level glycation (+162 Da) is achieved. An additional trailing edge is also resolved from all DAR species, with an approximate MW increment of +40 Da, which could be explained by non-covalent adducts, such as +Na and/or +K and/or $+NH_4$. The difference in DAR values observed between denaturing LC-MS and native-MS have previously been discussed by Campuzano [4].

References

1. Bush, M. F.; Hall, Z.; Giles, K.; Hoyes, J.; Robinson, C. V.; Ruotolo, B. T., Collision cross sections of proteins and their complexes: a calibration framework and database for gas-phase structural biology. *Anal Chem* **2010,** *82* (22), 9557-65.
2. Bern, M.; Sen, K. I.; Skilton, J.; Leon-Danon, L.; Qiu, D.; Peddicord, M.; Ding, W., Intact Mass Analysis and DAR Calculation for Antibody-Drug Conjugates. *Proc. 65th ASMS Conf. Mass spectrometry and Applied Topics, Indianapolis 2017* **2017**.
3. Bern, M.; Yang, E.; Kil, Y. J.; Carlson, E.; Becker, C.; Sen, K. I.; Skilton, J.; Nicholas, A., Charge Deconvolution of Hard-to-DeconvolveMass Spectra. *Proc. 17th HUPO Meeting, Orlando, FL, 2018* **2018**.
4. Campuzano, I. D. G.; Netirojjanakul, C.; Nshanian, M.; Lippens, J. L.; Kilgour, D. P. A.; Van Orden, S.; Loo, J. A., Native-MS Analysis of Monoclonal Antibody Conjugates by Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. *Anal Chem* **2018,** *90* (1), 745-751.
5. Formolo, T.; Ly, M.; Levy, M.; Kilpatrick, L.; Lute, S.; Phinney, K.; Marzilli, L.; Brorson, K.; Boyne, M.; Davis, D.; Schiel, J., Determination of the NISTmAb

Primary Structure. *State-of-the-Art and Emerging Technologies for Therapeutic Monoclonal Antibody Characterization Volume 2. ACS Symposium Series; American Chemical Society: Washington, DC, doi: 10.1021/bk-2015-1201.ch001* **2015**.
6. Kilgour, D. P.; Van Orden, S. L., Absorption mode Fourier transform mass spectrometry with no baseline correction using a novel asymmetric apodization function. *Rapid Commun Mass Spectrom* **2015,** *29* (11), 1009-18.
7. Campuzano, I. D. G.; Van Orden, S.; H., L.; Kilgour, D. P.; Lippens, J. L.; Nshanian, M.; Netirojjanakul, C.; Wongkongkathep, P.; Egea, P. F.; Loo, J. A., *Proc. 65th ASMS Conf. Mass spectrometry and Applied Topics, Indianapolis 2017 ThOB pm*.