Parsimonious Charge Deconvolution for Native Mass Spectrometry

Marshall Bern,*,[†] Tomislav Caval,[‡] Yong J. Kil,[†] Wilfred Tang,[†] Christopher Becker,[†] Eric Carlson,[†] Doron Kletter,[†] K. Ilker Sen,[†] Nicolas Galy,[‡] Dominique Hagemans,[‡] Vojtech Franc,[‡] and Albert J. R. Heck*,[‡]
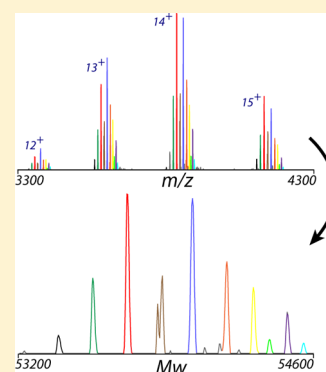
[†]Protein Metrics, Inc., San Carlos, California 94070, United States

[‡]Biomolecular Mass Spectrometry and Proteomics, Bijvoet Center for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, Science4Life, Utrecht University and Netherlands Proteomics Centre, Padualaan 8, 3584 CH Utrecht, The Netherlands

S *Supporting Information*

ABSTRACT: Charge deconvolution infers the mass from mass over charge (m/z) measurements in electrospray ionization mass spectra. When applied over a wide input m/z or broad target mass range, charge-deconvolution algorithms can produce artifacts, such as false masses at one-half or one-third of the correct mass. Indeed, a maximum entropy term in the objective function of MaxEnt, the most commonly used charge deconvolution algorithm, favors a deconvolved spectrum with many peaks over one with fewer peaks. Here we describe a new "parsimonious" charge deconvolution algorithm that produces fewer artifacts. The algorithm is especially well-suited to high-resolution native mass spectrometry of intact glycoproteins and protein complexes. Deconvolution of native mass spectra poses special challenges due to salt and small molecule adducts, multimers, wide mass ranges, and fewer and lower charge states. We demonstrate the performance of the new deconvolution algorithm on a range of samples. On the heavily glycosylated plasma properdin glycoprotein, the new algorithm could deconvolve monomer and dimer simultaneously and, when focused on the m/z range of the monomer, gave accurate and interpretable masses for glycoforms that had previously been analyzed manually using m/z peaks rather than deconvolved masses. On therapeutic antibodies, the new algorithm facilitated the analysis of extensions, truncations, and Fab glycosylation. The algorithm facilitates the use of native mass spectrometry for the qualitative and quantitative analysis of protein and protein assemblies.

KEYWORDS: *monoclonal antibody, glycoprotein, daclizumab, infliximab, cetuximab, factor P, properdin, algorithm, maximum entropy, parsimony, intact mass, high-resolution native mass spectrometry*

INTRODUCTION

Electrospray ionization mass spectra of biological macromolecules and protein complexes contain series of ion signals corresponding to the same chemical species in a sequence of charge states. The masses and intensities (ion currents) of the analyzed chemical species, as represented by an entire neutral-mass spectrum, can be inferred from the mass over charge measurements by computational deconvolution.

All charge deconvolution algorithms in use today are iterative algorithms that converge to a deconvolved neutral mass spectrum along with charge distributions for the neutral masses that together explain the observed m/z (mass over charge) spectrum. The most widely used deconvolution algorithm, with implementations called MaxEnt and ReSpect, was developed about 25 years ago[1,2] and licensed to most of the mass spectrometry (MS) instrument manufacturers. This algorithm converges to a deconvolved neutral mass spectrum that optimizes an objective function that measures the quality of the result using criteria such as fit to the observed data, peak width, correlation between neighboring charge states, and—its defining characteristic—the Shannon entropy of the neutral-mass spectrum. A more recent algorithm, UniDec,[3] leaves out the entropy term, and builds in expected correlation between neighboring charge states by blending them with a smoothing filter. UniDec also includes specific support for ion mobility data and nanodisk analysis. Other recent work has focused on peak enhancement of m/z spectra[4] to improve the performance of maximum entropy charge deconvolution for native mass spectrometry.

Regardless of the algorithmic details, the deconvolution iteration generally converges to a local rather than a global optimum. Two important user-controlled parameters for deconvolution are the input m/z range and the output mass range. Deconvolution algorithms usually assume that all of the ions (except perhaps some low-charge m/z peaks, recognizable by resolved isotopes) in the input range represent chemical species in the mass range. This assumption allows deconvolution of lower signal-to-noise spectra by limiting the number of masses and charges that the algorithm must consider, but it runs the risk that chemical species outside the mass range may be undetected or give false additional masses within the user-set target mass range. A practical solution entails deconvolution of the m/z range onto a wide mass range to

survey the masses, followed by deconvolution of selected m/z ranges onto narrow mass ranges to capture more detailed information.

With a wide target mass range, deconvolution can produce "harmonic" artifacts, for example, false mass peaks at one-half or twice the true mass, due to coincidences of the m/z series for masses with ratio relationships. Even with relatively narrow target mass ranges, off-by-one charge assignments produce another type of artifact, side lobes on either side of the true masses, for example 3000 Da too low and high if the strongest m/z signal is around m/z 3000. Both harmonic and off-by-one artifacts increase entropy of the deconvolution, so the entropy term in the objective function, which helps the algorithm resolve closely spaced masses, has the undesired side effect of promoting artifacts. Artifacts are a minor problem in some scenarios, but they can be quite misleading in other practical applications:

(1) Automated workflows that forego expert human inspection

(2) Analysis of antibodies, including bispecifics, where harmonic artifacts may be mistaken for half-mAbs, aggregations, or mispairings

(3) Antibody–drug conjugates (ADCs), where off-by-one artifacts may bias quantitation of drug loading

(4) Heavily glycosylated or other highly modified proteins

To be fair, note that MS automation, bispecifics, and ADCs barely existed when the maximum entropy algorithm was developed in the early 1990s.

Perhaps the most important development in intact MS since the early 1990s is native MS,[5−9] made possible by methods and instrument innovations including the introduction of Orbitrap mass analyzers optimized for the transmission of high m/z ions.[10,11] Native MS has enabled the measurement of the micro- and macro-heterogeneity in proteins and complexes bound to multiple cofactors[12] or harboring multiple post-translational modifications (PTMs)[13−18] and in large endogenous protein assemblies, such as the ribosome[11] and intact viruses.[13,19] Complex native MS spectra, sometimes exhibiting ion signals of several hundred species of different molecular weight, require sophisticated algorithms and software to extract qualitative and quantitative information on co-occurring proteoforms or protein–ligand stoichiometries.

To address these issues, we present an improved charge deconvolution algorithm that divides the process into two stages: charge inference and peak sharpening. The charge inference stage aims for an artifact-free neutral mass spectrum with a "parsimonious" set of mass peaks that explains the observed m/z spectrum. The optional peak sharpening stage uses point-spread-function deconvolution on the neutral mass spectrum to resolve closely spaced peaks. Post-deconvolution peak sharpening on the neutral mass spectrum has practical advantages over coupled charge inference and peak sharpening, including speed of processing, visual inspection of before and after spectra, and compatibility with a variety of well-developed super-resolution algorithms, such as Richardson–Lucy,[20,21] maximum entropy,[22] and convolutional neural networks. This design choice imposes some restrictions on the super-resolution algorithm's underlying physical model; for example, the point-spread function may depend upon mass, for example, broadening at higher mass, but not upon charge or m/z.

We focus on the charge inference stage because charge inference is central and unique to ESI mass spectrometry, and it is also the source of the most misleading deconvolution artifacts, meaning false masses far removed from all true masses. (The super-resolution stage can produce minor artifacts such as "ringing" around true masses.) We demonstrate parsimonious charge inference on complex glycosylated therapeutic antibodies and a heavily glycosylated plasma glycoprotein, all analyzed under native conditions. We reveal on several therapeutic antibodies a variety of interesting causes of species microheterogeneity,[23] including N-terminal extensions and truncations, abundant C-terminal lysine retention, and multiple glycosylation sites. We argue that this improved parsimonious charge deconvolution tool will benefit the qualitative and quantitative analysis of protein therapeutics, including biosimilar testing, drug load quantification in ADCs, and glycoproteoform analysis.

■ MATERIALS AND METHODS

Chemicals and Materials

The three therapeutic mAbs, namely, cetuximab (lot number 7663503, expiration date 3/2010), daclizumab (lot number B0035, expiration date unknown), and infliximab/Remicade (lot number and expiration date unknown) used in this work are all commercially available and were kind gifts from Genmab (Utrecht, The Netherlands). All mAb samples given to us likely represent expired batches (see Table S2). Properdin, also known as Factor P (Uniprot code: P27918), purified from human blood plasma, was obtained from Complement Technology (Tyler, TX). We obtained amino acid sequences from literature[24] and Web searches (www.commonchemistry.org). All amino acid sequences lacked the N-terminal signal peptides (except daclizumab, for which we used the sequence with signal peptide obtained from its European patent application: EP 2 527 429 A2), and specifications of the samples are listed in Tables S1 and S2. Dithiothreitol (DTT), iodoacetamide (IAA), and ammonium acetate (AMAC) were purchased from Sigma-Aldrich (Steinheim, Germany). Phosphate buffer was from Lonza (Verviers, Belgium). Formic acid (FA) was from Merck (Darmstadt, Germany). Acetonitrile (ACN) was purchased from Biosolve (Valkenswaard, The Netherlands). Sequencing-grade trypsin was obtained from Promega (Madison, WI). Lys-C, Glu-C, and Asp-N were obtained from Roche (Indianapolis, IN). PNGase F was obtained from Asparia Glycomics (San Sebastian, Spain). The IdeS enzyme for cetuximab digestion was purchased from Genovis (Lund, Sweden).

Sample Preparation for Native MS

The powder of the therapeutic mAbs was reconstituted in Milli-Q water. The aqueous mAbs samples and unprocessed protein solution (phosphate buffer at pH 7.2) containing ∼30−40 μg of properdin were buffer-exchanged with 150 mM aqueous AMAC (pH 7.5) by centrifugation using a 10 kDa cutoff filter (Merck Millipore, Germany). The resulting protein concentration was measured by UV absorbance at 280 nm and adjusted to 2 to 3 μM prior to native MS analysis. PNGase F was used to cleave the N-glycans of mAbs and properdin using protocols previously described.[25] Cetuximab was used to demonstrate the processing of native spectra of mAb treated by IdeS enzyme. The aqueous cetuximab (30 ug) was incubated with IdeS enzyme (30 units) in phosphate buffer at pH 7.5 for 30 min at 37 °C. This sample was either submitted to the native MS measurements or further treated with 20 mM DTT and incubated for 30 min at 37 °C. All samples were buffer-exchanged to 150 mM AMAC (pH 7.5) prior to native MS measurements.

Native MS Analysis

Samples were analyzed on a modified Exactive Plus Orbitrap instrument with extended mass range (EMR) (Thermo Fisher Scientific, Bremen) using a standard m/z range of 500−10 000, as previously described in detail.[25] The voltage offsets on the transport multipoles and ion lenses were manually tuned to

achieve optimal transmission of protein ions at elevated m/z. Nitrogen was used in the higher energy collisional dissociation (HCD) cell at a gas pressure of (6 to 8) $\times$ 10^{-10} bar. MS parameters used: spray voltage 1.2 to 1.3 V, source temperature 250 °C, source fragmentation, and collision energy were varied from 30 to 100 V, and resolution (at m/z 200) 35 000 for properdin and 70 000 for mAbs. The instrument was mass calibrated as previously described using a solution of CsI.[25]

Proteolytic Digestion for Bottom-up Proteomics

The mAb daclizumab (5 μg) was reduced using 10 mM DTT at 56 °C for 30 min and alkylated with 30 mM IAA at room temperature for 30 min in the dark. The excess of IAA was quenched by using 10 mM DTT. The protein solution was first digested with Lys-C (or AspN, or GluC) at an enzyme-to-protein ratio of 1:50 (w/w) for 4 h at 37 °C and then overnight with trypsin at an enzyme-to-protein ratio of 1:100 (w/w) at 37 °C. The proteolytic digest was desalted by Oasis μElution plate,[26] dried, and dissolved in 40 uL of 0.1% FA prior liquid chromatography (LC)−MS and MS/MS analysis.

LC−MS and MS/MS Analysis

Proteolytic peptides from daclizumab (typically 300 fmol) were separated and analyzed using an Agilent 1290 Infinity HPLC system (Agilent Technologies, Waldbronn, Germany) coupled online to an Orbitrap Fusion Lumos Tribrid mass spectrometer (Thermo Fisher Scientific, Bremen, Germany). Reversed-phase separation was accomplished using a 100 μm inner diameter 2 cm trap column (in-housed packed with ReproSil-Pur C18-AQ, 3 μm) (Dr. Maisch, Ammerbuch-Entringen, Germany) coupled to a 50 μm inner diameter 50 cm analytical column (in-house packed with Poroshell 120 EC-C18, 2.7 μm) (Agilent Technologies, Amstelveen, The Netherlands). Mobile-phase solvent A consisted of 0.1% FA in water, and mobile-phase solvent B consisted of 0.1% FA in ACN. The flow rate was set to 300 nL/min. A 45 min gradient was used as followed: 0−10 min, 100% solvent A; 10.1−35 min 10% solvent B; 35−38 min 45% solvent B; 38−40 min 100% solvent B; 40−45 min 100% solvent A. Nanospray was achieved using a coated fused silica emitter (New Objective, Cambridge, MA) (outer diameter, 360 μm; inner diameter, 20 μm; tip inner diameter, 10 μm) biased to 2 kV. The mass spectrometer was operated in positive ion mode, and the spectra were acquired in the data-dependent acquisition mode. For the MS scans the scan range was set from 300 to 2000 m/z at a resolution of 60 000 and the AGC target was set to 4 $\times$ 10^5. For the MS/MS measurements HCD and electron-transfer and higher-energy collision dissociation (EThcD) were used. HCD was performed with normalized collision energy of 35%. A supplementary activation energy of 20% was used for EThcD. For the MS/MS scans the scan range was set from 100 to 2000 m/z and the resolution was set to 30 000, the AGC target was set to 5 $\times$ 10^5, the precursor isolation width was 1.6 Th, and the maximum injection time was set to 300 ms.

LC−MS/MS Data Analysis

Raw LC−MS/MS data on the digest of daclizumab were interpreted using Byonic software (Protein Metrics).[27] The following parameters were used for data searches: precursor ion mass tolerance, 10 ppm; product ion mass tolerance, 20 ppm; fixed modification, Cys carbamidomethyl; variable modification, Met oxidation. A semitryptic specificity search was chosen for all samples. The protein database contained the daclizumab protein amino acid sequence (Table S1).

Description of Algorithm

An m/z spectrum is a sequence of pairs $m_i = (x_i, y_i)$, where x_i is the m/z value and y_i is the intensity value. Most often the intensity

y_i represents a single species of ions, but, in general, the intensity represents a mix of ions of various charges, and we let $c_k(m_i)$ denote the fraction of the intensity that has charge k for $k = 1, 2, ...,$ up to some maximum charge. For each i the sum of $c_k(m_i)$ values over all k is one. The $c_k(m_i)$ values are initially unknown and set to be equal, but the algorithm iteratively learns these values as it learns the neutral mass spectrum.

An observed m/z value m_i maps to a sequence of neutral masses, $k \cdot x_i - k \cdot 1.00728$, with intensities $c_k(m_i) \cdot y_i$ for $k = 1, 2, ...$. Here we are assuming positive-mode MS; for negative mode the neutral mass is $k \cdot x_i + k \cdot 1.00728$, where 1.00728 is the mass of a proton in Daltons. We can compute a full neutral mass spectrum by accumulating, over all m_i, the intensities $c_k(m_i) \cdot y_i$ into a vector at the appropriate x values, $k \cdot x_i -$ (or for negative mode +) $k \cdot 1.00728$. The result of this m/z-to-mass "backward" mapping is a sequence of points, $M_j = (X_j, Y_j)$. For each point M_j in the neutral mass spectrum, we can also keep a record of the intensity contributions $C_k(M_j)$ from each charge k and normalize these contributions so that for each j the $C_k(M_j)$ values sum to one. The M_j points and $C_k(M_j)$ values can be used in a mass-to-m/z "forward" mapping to give a modeled m/z spectrum. Alternation of backward and forward mappings improves the values of the unobserved $c_k(m_i)$, $C_k(M_j)$, and Y_j variables. The computation stops after a predefined number of iterations or when the neutral mass spectrum converges, meaning that it changes very little between iterations.

The quality of a deconvolution can be evaluated by various criteria, and deconvolution algorithms either implicitly or explicitly aim to optimize an objective function that combines the criteria. To our knowledge, none of the maximum entropy algorithms disclose their objective functions or optimization algorithms; however, the primary criterion is always goodness of fit, which can be measured by forward mapping the neutral mass spectrum to an m/z spectrum and then evaluating, for example, the sum of the squares of the differences between the observed and computed values. A second criterion is smoothness of charge distributions $C_k(M_j)$. Maximum entropy methods add into the objective function a weighting factor times the Shannon entropy of the neutral mass spectrum regarded as a probability distribution, that is, the sum over j of $-\dot{Y}_j \log_2 \dot{Y}_j$, where $\dot{Y}_j = Y_j / \sum Y_j$. The entropy criterion tends to split broad peaks into multiple sharper peaks.

In the algorithm used here, we introduce a new criterion based on the assumption that m/z coincidences are rare, especially in highly resolved mass spectra, so that 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 criterion tends to drive the iteration to a "parsimonious" neutral mass spectrum that contains a minimal set of mass peaks to explain the m/z spectrum. Notice that if the sample does contain a problem pair of masses, say a monomer and a dimer, then each m_i point may still be fairly pure if there is some separation in m/z, for example, if the dimer cannot carry twice the charge of the monomer. Separation in m/z is less reliable in mass spectra taken under "standard" denaturing conditions than in native mass spectra, in which different oligomers tend to claim distinct m/z ranges. If there is no separation in m/z, then the dimer explains every m/z peak explained by the monomer, and the evidence of the monomer is merely taller m/z peaks at even charges of the dimer. In this case, the monomer's intensity in the computed neutral mass spectrum depends on the relative weighting of the parsimony and charging smoothness criteria.

We implemented the new charge inference algorithm in C++ in a commercial product called Protein Metrics Intact or PMI Intact, shown in Figure 1. Input data from almost any type of MS
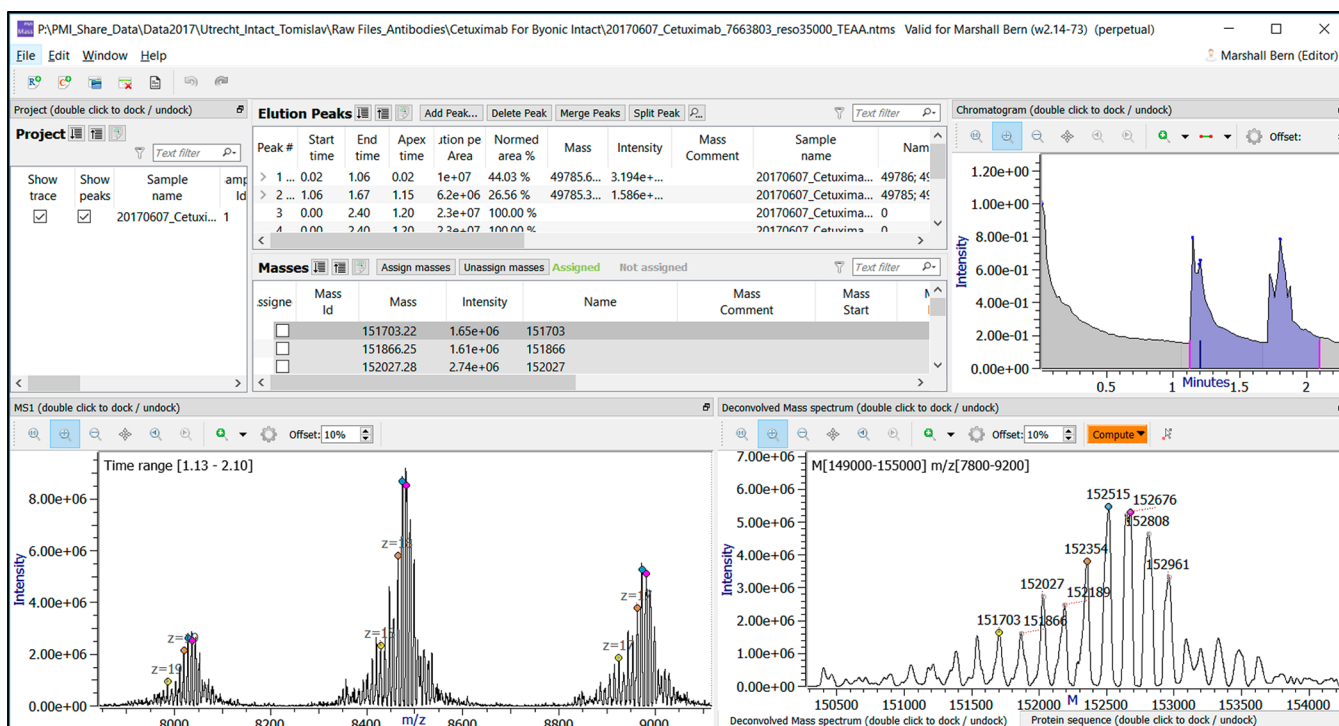
Figure 1. Screenshot of Protein Metrics Intact software interface. The software applied to the mAb cetuximab provides tables of input files, elution peaks, and detected masses (upper left); total ion chromatogram (upper right); *m/z* spectra (lower left) summed over the selected time window; and deconvolved neutral mass spectrum (lower right). Mass peaks are interactively connected to *m/z* peaks by colored dots. The mass peak at 152 354 is a good match for the calculated average isotope mass of 152 356 Da for cetuximab with G2FGal2 on its Fab glycosylation sites and G0F on its Fc glycosylation sites. 152 515 and 152 676 match G2FGal2 Fab glycosylation with one- and two-G1F Fc glycosylation.

instrument can be sliced by elution time into any number of possibly overlapping time windows, and summed mass spectra for each time window can be further sliced by *m/z* for separate deconvolution. Both *m/z* and mass point spacing are user-controllable; mass spacing below ~0.2 Da preserves isotope resolution. We also implemented Richardson−Lucy point spread deconvolution, which we call "peak sharpening" to avoid confusion. This iterative algorithm takes as input 1D or 2D signals (such as a time series, mass spectrum, or image), along with a point spread function F, and computes an output whose convolution with F gives a result close to the observed input. Our current version of the software (v2.15, released in December 2017) lets the user define point spread functions with Gaussian or Lorentzian, possibly asymmetric, tails. Gaussian tails approximate isotope distributions and measurement inaccuracy; heavy Lorentzian tails may approximate adducts. PMI Intact also includes interactive visualization. Peaks in the deconvolved mass spectrum may be selected interactively, and the software marks the selected peaks and the *m/z* points that map to these peaks with matching colored dots for human inspection and validation. The software also enables automatic peak assignment from protein sequences, masses, or mass deltas as well as automatic graphical report generation.

PMI Intact is currently in use for a diverse set of applications including analysis of both reduced and intact monoclonal antibodies, IdeS-digested and intact bispecific antibodies, antibody-drug conjugates,[28] DNA oligos, heavily glycosylated glycoproteins, protein−ligand binding, and noncovalently bound protein complexes up to 1 MDa or more.

Software Tests

We tested PMI Intact on data from properdin and the three antibodies daclizumab, infliximab, and cetuximab. Experimental

high-resolution native MS data from our laboratory was already published for properdin[16] and therefore represented an ideal test-case to demonstrate the power of this new algorithm. The three antibodies were chosen because they presented interesting analytical challenges due to their complex glycosylation profiles or extensive protein processing characteristics. We benchmarked PMI Intact against Protein Deconvolution 4.0 (Thermo Fisher Scientific) on the properdin data using identical *m/z* and mass ranges for the two programs. For PTM composition analysis, data were interpreted manually and glycan structures were deduced based on known biosynthetic pathways. Average masses were used for the PTM assignments, including hexose/mannose/galactose (Hex/Man/Gal, 162.1424 Da), *N*-acetylhexosamine/*N*-acetylglucosamine (HexNAc/GlcNAc/GalNAc, 203.1950 Da), and *N*-acetylneuraminic acid (NeuAc, 291.2579 Da). All used symbols and text nomenclature are according to recommendations of the Consortium for Functional Glycomics.

■ RESULTS

As a first demonstration of the value of parsimony in the deconvolution of ESI mass spectra, we reanalyzed published high-resolution mass spectra on the plasma protein properdin. This protein may exist in various oligomeric states and harbors a diversity of modifications on various sites, including *N*- and *O*- glycosylation, as well as *C*-mannosylation, making properdin a challenging target for structural analysis. Our initial native MS measurements revealed monomer and dimer of properdin. We first tested whether Protein Metrics Intact and Thermo Protein Deconvolution 4.0 could find both monomer and dimer using *m/z* and mass ranges large enough to accommodate both forms; this is a challenging problem for charge deconvolution algorithms due to coincidences of *m/z* peaks. As seen in Figure S1 of the Supporting Information,
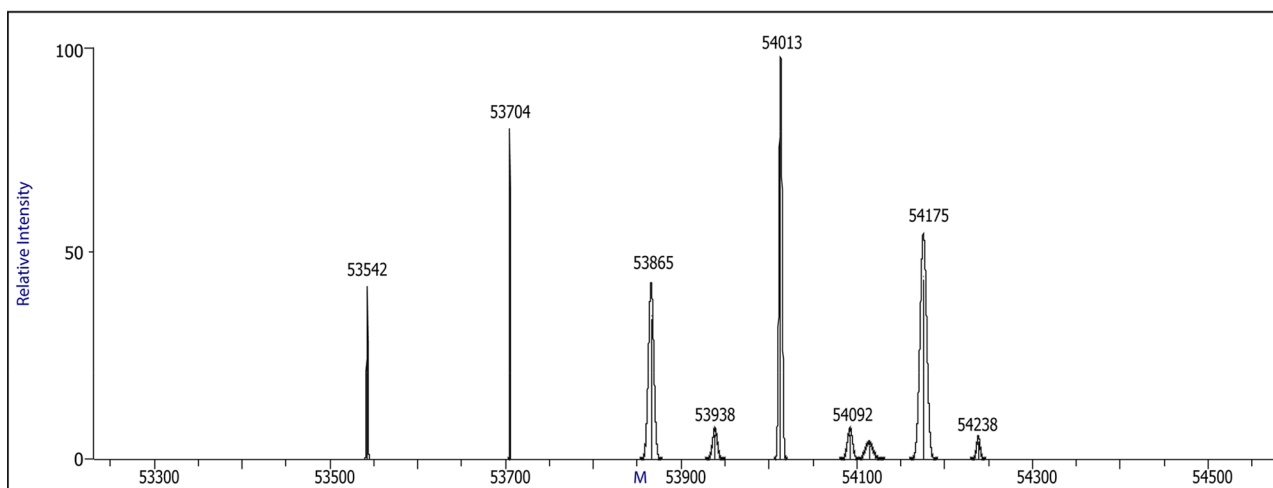
(a)
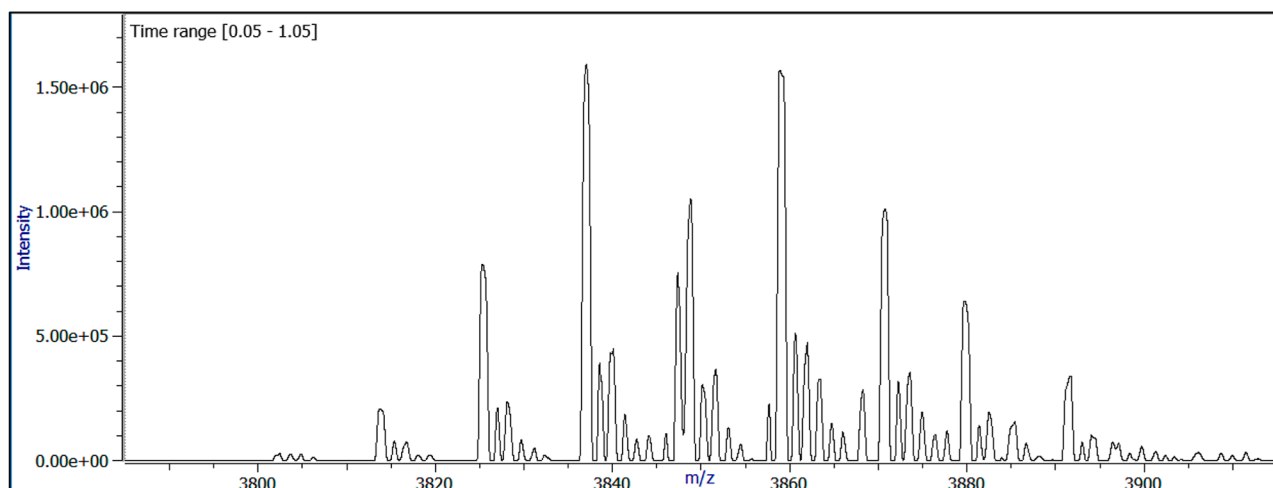
(b)
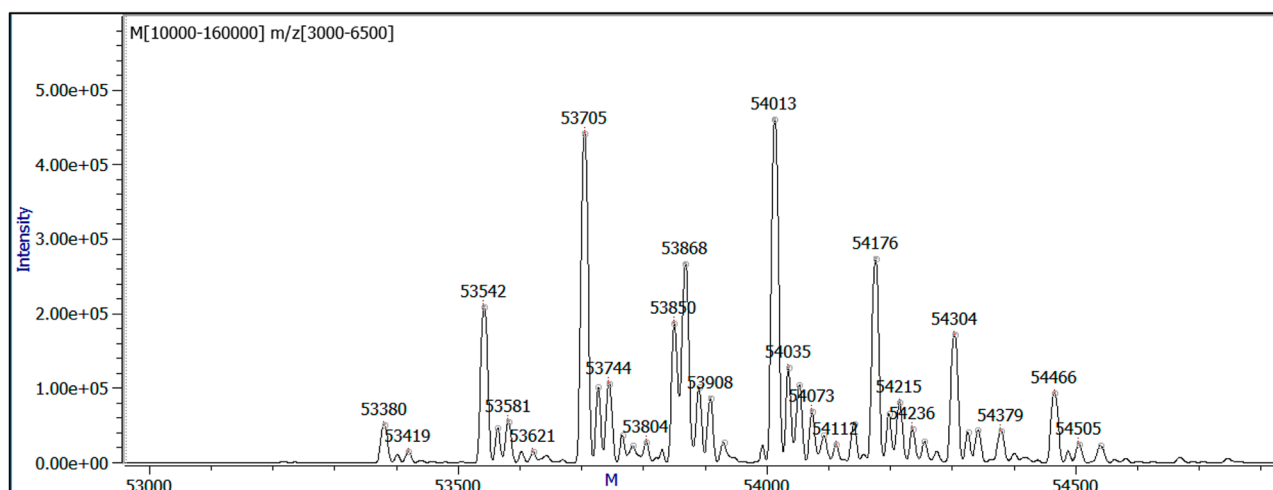
(c)

Figure 2. Proteoform profile of monomeric properdin. (a) Zoom of the 3800−4000 *m/z* range of properdin monomer mass spectrum. (b) Zoom of PMI Intact's deconvolution computed on *m/z* range 3000−6500 and *m* range 10 000−160 000. (c) Zoom of thermo deconvolution computed on the same *m/z* and *m* ranges. Thermo deconvolution misses a number of proteoforms, including abundant forms at 53 380, 54 304, and 54 466, most likely due to interference from the dimer.

Protein Metrics Intact gives an accurate deconvolution, but depending upon input parameter settings, Thermo Protein Deconvolution 4.0 either gives numerous large artifact peaks or loses the dimer form altogether, and it was impossible to find a setting that gave an accurate deconvolution. Figure 2 shows a more detailed comparison of Protein Metrics and Thermo deconvolutions, alongside the major charge state from the m/z spectrum previously used for manual analysis.[16] When deconvolved with wide m/z and mass ranges, Thermo software, along with losing the dimer form, loses many of the medium abundance monomer proteoforms yet finds some of the lower abundance proteoforms, possibly because they are at half the mass of dimer forms. Thermo also gives highly variable peak widths. A wide mass peak in a deconvolved mass spectrum generally indicates mass uncertainty, caused by m/z peaks with different charges mapping to slightly different m values, but in this case the wide mass peaks at 53 866 and 54 176 Da seem to be caused by dimer m/z peaks mistaken for monomer. PMI Intact returns a deconvolution in good visual agreement with the major charge states of the m/z spectrum and mass agreement within ±2 Da of calculated masses of correct proteoform assignments.[16] The previous work made assignments by manual inspection of individual m/z peaks, which have poorer accuracy than the deconvolved mass spectrum peaks. The previous analysis also made several assignment errors that are now apparent from the improved resolution and mass accuracy of Intact's deconvolved spectrum. PMI Intact gave about 25 interpretable species in this analysis (Supporting Information Figure S2). PMI Intact also revealed relatively high abundance of salt adducts (i.e., Na$^+$ and K$^+$) to some of the ion species. On the basis of this knowledge, we also analyzed a further desalted properdin sample by native MS, for which we obtained spectra nearly free of salt adducts, enabling us to find evidence of a low abundance of triantennary N-glycans (Figure S5), whose assignments could be confirmed by bottom-up glycopeptide analysis (Figure S6). Interestingly, the triantennary N-glycans were found on proteoforms with 15 C-mannosylations but not on those with fewer C-mannosylations, not even those with 14 C-mannosylations, which are most abundant in this sample. This is evidence of whole-protein correlation between PTMs that could not easily be obtained from bottom-up, middle-down, or top-down fragmentation spectra of a 54 kDa protein with 20 labile PTMs.

As a further demonstration of the utility of the new deconvolution algorithm to target protein therapeutics, we analyzed three clinically approved and used mAbs. As a first example, Figure 3 shows results on the PNGaseF-treated deglycosylated mAb daclizumab. Somewhat surprisingly, we observed three quite distinct masses in the deconvolved spectrum, namely, at 11 057, 132 792, and 143 831 Da, along with +340 Da masses for each of these peaks and 2 × 340 Da for the 143 831 Da species only. The calculated mass for deglycosylated daclizumab is 143 832 Da = 2 × 48 717 (heavy chain) + 2 × 23 215 (light chain) − 32 (for the 16 disulfide bonds). See Supporting Information S1 for protein sequences. The 11 057 Da species is an exact integer match to the average isotope mass of the heavy chain initial sequence QVQLVQ-SGAEVKKPGSSVKVSCKASGYTFTSYRMHWVRQAPGQGL-EWIGYINPSTGYTEYNQKFKDKATITADESTNTAYMELS-SLRSEDTAVYYCARGG.G (where. denotes the cleaved bond) with N-terminal pyro-Glu and the expected single disulfide bond. G.G is a well known clipping site for monoclonal antibodies,[29] attributed to the flexibility of GG, and in this case the even-more-flexible GGG sequence occurs in the heavy chain CDR3, making it solvent-accessible. The mass 132 792 Da corresponds to the full-length mAb minus the initial sequence ending in GG. The

fact that the mass of the observed fragments minus the mass of the intact mAb, (132 792 + 11 057) − 143 831 = 18 Da, gives the mass of water, reveals that hydrolysis is causing the cleavage rather than gas-phase fragmentation inside the mass analyzer. The extra +340 Da peaks are consistent with an N-terminal extension of VHS (part of the signal peptide). A small peak for S, with measured mass delta (104.041 Da ≈ 87.032 for S + 17.027 for pyroQ) correct to < 0.02 Da, which is 2 ppm, in isotope-resolved Figure 3c supports this interpretation. FWHM (full width at half-maximum) peak widths at m/z 1900 are ∼0.08, sufficient to resolve isotopes of 11 kDa masses. FWHM of the full mAb peaks at m/z 6000 are ∼0.9, limited by the isotope distribution of the molecule (calculated FWHM of 1 at m/z 6000) rather than by instrument resolution, which should be below 0.2 at 6000 m/z as Orbitrap resolution decreases with the square root of m/z.

We based the interpretation of GG clipping and VHS extension only on the deconvolved mass spectra and protein sequence; this inference would be difficult without high-resolution mass spectrometry and accurate artifact-free deconvolution. We then searched our bottom-up proteomics data for nonspecific peptides and peptides with N-terminal extensions, and the search results confirmed our interpretation (Supporting Information Figures S5 and S6). The information from the native MS data prompted us to look for these features in the LC−MS/MS peptide data.

Next, we targeted the mAb infliximab. We first analyzed deglycosylated infliximab, because the spectrum of the deglycosylated antibody, displayed in Figure 4, helps to interpret the more complicated spectrum of nondeglycosylated infliximab. The peak at 145 623 Da is an exact match for the calculated deglycosylated mass of 145 623 Da, and the mass deltas of +128 Da for the other two large peaks in the deglycosylated infliximab are exact integer matches for C-terminal lysines, a modification known to occur frequently in recombinant mAbs. The presence of this triplet of mAb species harboring zero, one, and two C-terminal lysines leads to a denser and more complicated spectrum for nondeglycosylated infliximab. The peaks at 148 511, 148 638, and 148 768 Da in the glycosylated infliximab spectra can be assigned as matches to proteoforms with two N-glycans with composition G0F (= HexNAc(4)Hex(3)Fuc(1)) (with average-isotope additional mass of 2891), along with zero, one, and two C-terminal lysines.

Extending the complexity of the targeted mAb still further, we next analyzed cetuximab, as far as we know the only therapeutic antibody in current clinical use that has, along with the usual Fc glycosylation site, an additional glycosylation site in the Fab region. Therefore, we chose to digest cetuximab with IdeS[30] to separate Fab and Fc. IdeS digestion produces a F(ab′)2 component. Reduction with DTT then reduces the F(ab′)2 into Fd subunits, that is, the heavy chain from the N-terminus up to ...PAPELLG, but often leaves disulfide bonds within subunits intact. After IdeS digestion, the Fc may appear as either ∼50 kDa Fc species held together noncovalently or ∼25 kDa Fc/2. High-resolution native MS data acquired for this whole mixture of species, that is, the light chain LC (∼23 kDa), the glycosylated Fc/2 (∼25 kDa), the glycosylated Fd (∼27 kDa), and the glycosylated Fc (∼50 kDa), processed by Protein Metrics Intact deconvolution, gave results in close agreement with a previous detailed analysis of cetuximab Fab glycosylation,[31] except that we noted now that the previous analysis misidentified peaks at 27 688, 27 832, and 28 216 Da as glycans with somewhat unusual GlcNAc-Gal-GlcNAc antennas. These misidentifications may stem from arithmetic mistakes as the masses are each off by ∼100 Da. We interpret the peak at 27 688 Da as HexNAc(4)Hex(7)Fuc(2), that is, a glycan with antennal
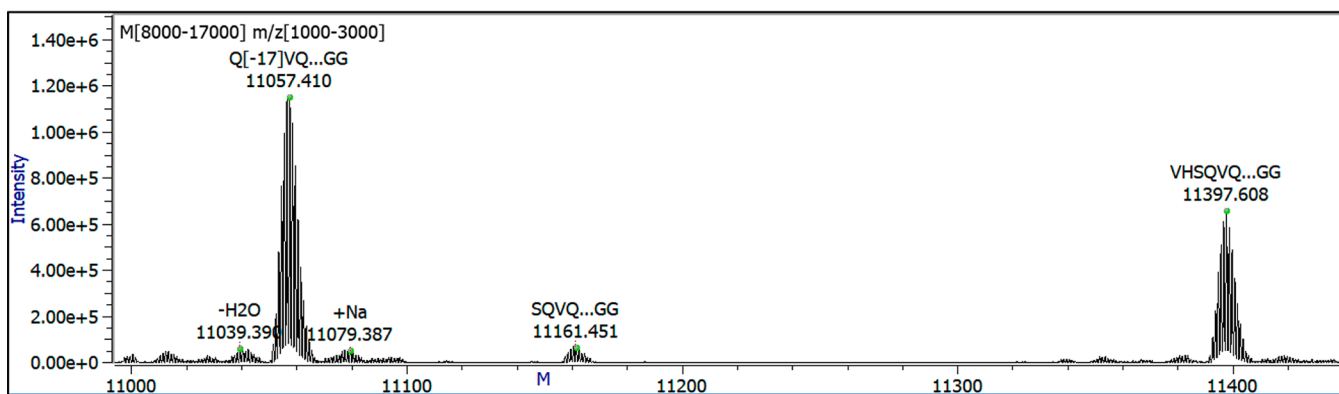
(a)
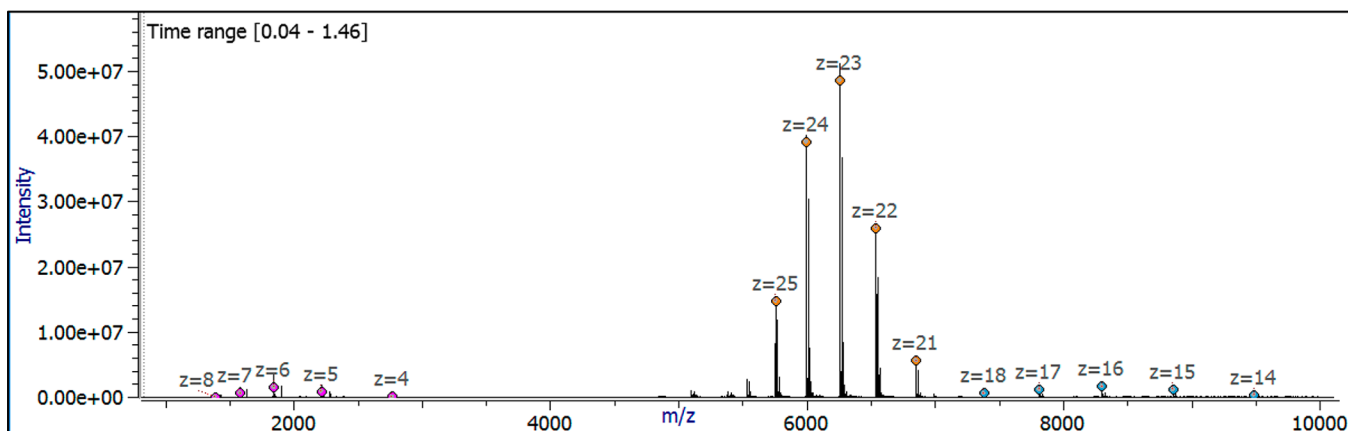
(b)
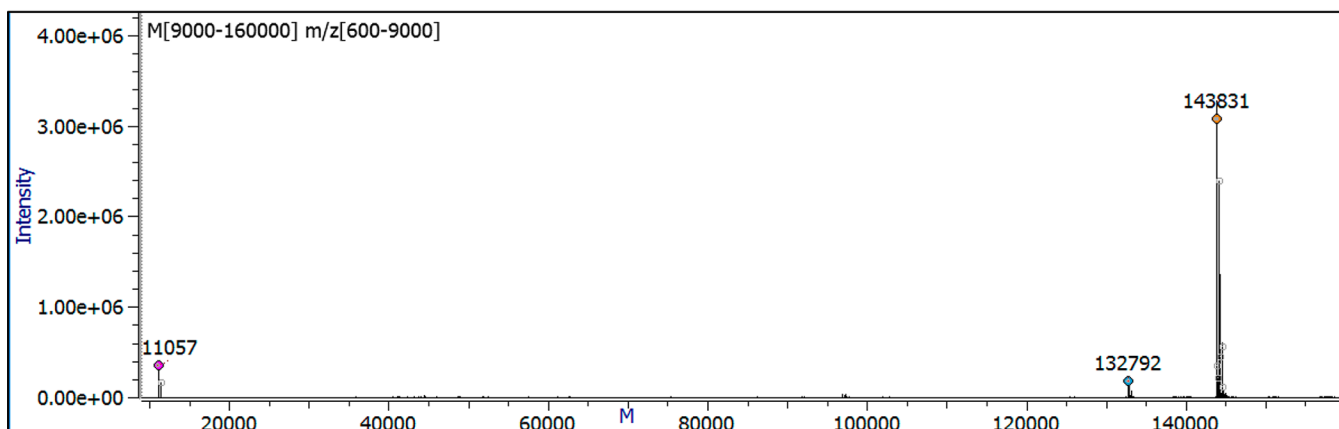
(c)

Figure 3. Full range m/z and deconvolved native ESI mass spectra of the deglycosylated mAb daclizumab. The m/z spectrum (a) shows three distinct charge series. In the mass spectrum the peak at 143 831 Da represents the mass of the full mAb without glycans or C-terminal Lys. 11 057 and 132 792 Da (which sum to 143 849) reveal the occurrence of two fragments formed via a GG clip from the heavy chain N-terminus. 143 831 is accompanied by two smaller peaks at a ΔM_w of +340 and +680 Da. The fragments of 132 792 and 11 057 Da each have only one +340 peak. These molecules originate from N-terminal extensions of the amino acid residues VHS (part of the signal peptide). (c) An isotope-resolved deconvolved mass spectrum. The small peak at 11 161.451 ($\approx$ 11 057.410 + 87.032 + 17.027) fits the GG clip along with N-terminal S, which prevents the formation of a pyro-Glu at the most abundant N-terminal Q. Thus three distinct N-termini coexist in this mAb product; the most abundant is pyroQVQLV..., the less abundant is VHSQVQLV.., and least abundant is SQVQLV....

Fuc, which gives an exact mass match to the closest integer and connects biosynthetically to the most abundant glycoform in the deconvolved spectrum, HexNAc(4)Hex(7)Fuc(1). Figure 5 includes small unlabeled peaks at 27 834 and 28 215, which are within 2 Da of the misidentified peaks in the previous analysis and also within 2 Da of the theoretical masses for the Fd with HexNAc(4)Hex(7)Fuc(3) and HexNAc(5)Hex(9)Fuc(2), respectively. As shown in Table S3 in the Supporting Information, the
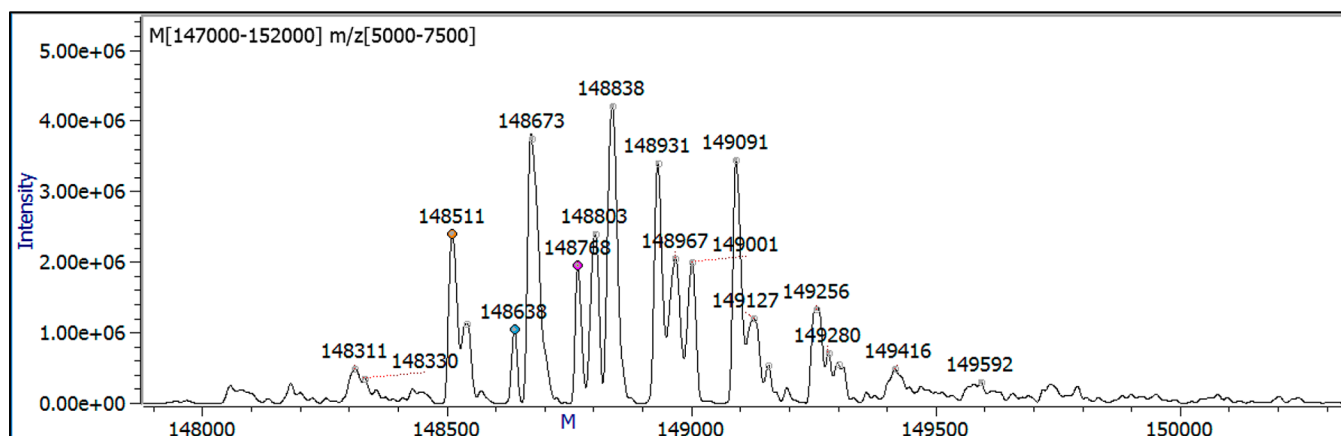
(a)
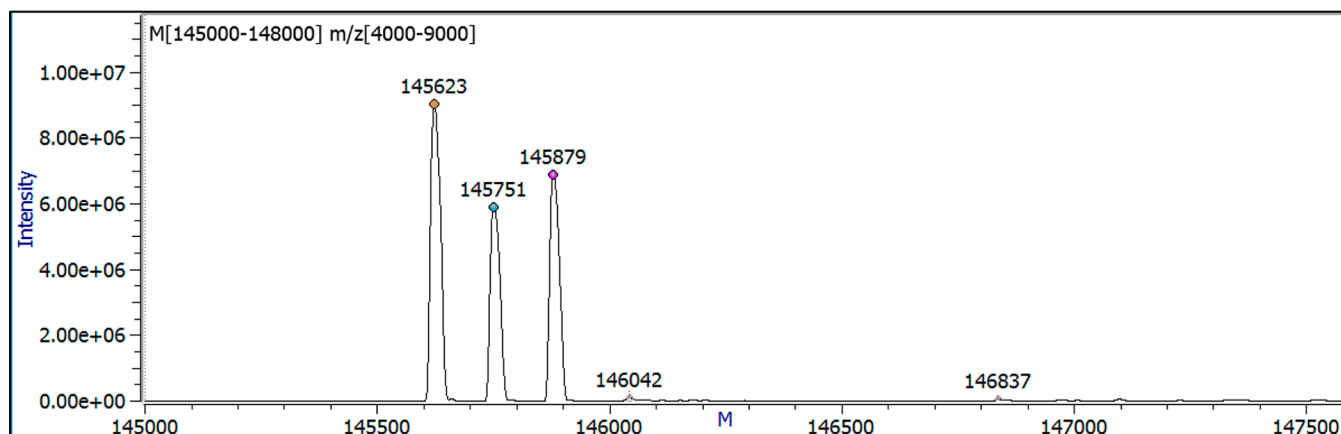
(b)

Figure 4. Deconvolved high-resolution native mass spectra of the deglycosylated and glycosylated mAb infliximab. Deglycosylated infliximab (a) shows three abundant species with masses in agreement with the amino acid sequence of the full mAb, along with species from which one or two C-terminal lysines had been clipped. The small peaks at 14 042 and 146 837 Da most likely represent, respectively, glycation on 146 042 and Man5 on 145 623. In the deconvolved mass spectrum of the glycosylated infliximab, the marked peaks exhibit the same triplets originating from the mAb with zero, one, or two C-terminal lysines, along with two N-glycans with G0F (= HexNAc(4)Hex(3)Fuc(1)). Each marked peak begins a chain of peaks with ~162 Da spacing, showing glycosylation heterogeneity. For example, the peaks at masses 148 511, 148 673, 148 838, 149 091, 149 256, and 149 416 Da correspond to the mAb with no C-terminal lysine and zero to five Gal monosaccharides.

deconvolved spectrum includes at least 14 recognizable Fd glycoproteoforms over a 100-fold dynamic range. In native MS on intact proteins, glycoproteoforms with and without sialic acids have similar ionization propensities and gas-phase stabilities, and hence peak intensities in the deconvolved mass spectrum should give accurate relative quantification.[25]

On the basis of the detailed analysis of the IdeS induced fragments of cetuximab, we were also able to annotate many of the abundant ion signals in the complicated intact cetuximab spectrum. Summing 23 422 (LC), 27 543 (Fd + G2FGal2), and 25 232 (Fc/2 + G0F) from Figure 5c and then multiplying by 2 and subtracting 36 Da for gain of water from IdeS digestion along with 4 for interchain disulfide bonds gives a mAb proteoform at 152 354, a perfect match for the peak with the orange dot in Figure 1. The peak at 152 515 then represents a proteoform with G1F on one of the Fc sites; this peak is taller than 152 354 because G1F has almost equal abundance as G0F in Figure 5b,c, and there are two chances for an extra Gal. The peaks at 151 866, 152 027, 152 189, and 152 676 are interpretable as proteoforms differing in number of galactose monosaccharides. The peaks at 152 808

and 152 961 probably contain unresolved proteoforms, including multiple fucosylation on the Fd.

■ DISCUSSION

For the past 25 years, charge deconvolution of protein ESI−MS data has almost exclusively been performed by some implementation of the maximum entropy algorithm. During this time period, MS instruments and associated technologies such as chromatography and sample handling have improved in speed, resolution, and sensitivity, and partially as a consequence of technology improvements, the variety, complexity, and masses of target molecules for intact and native MS have increased significantly. Therefore, high-resolution native MS is now widely adopted by the pharmaceutical industry to characterize some of their most important protein therapeutics, such as the mAbs analyzed here. These developments motivate the development of accurate, automated, and user-friendly deconvolution programs that can handle more difficult data with less user intervention and validation.

A primary contribution of the work presented here is the use of parsimony in charge deconvolution. Parsimony is a guiding principle
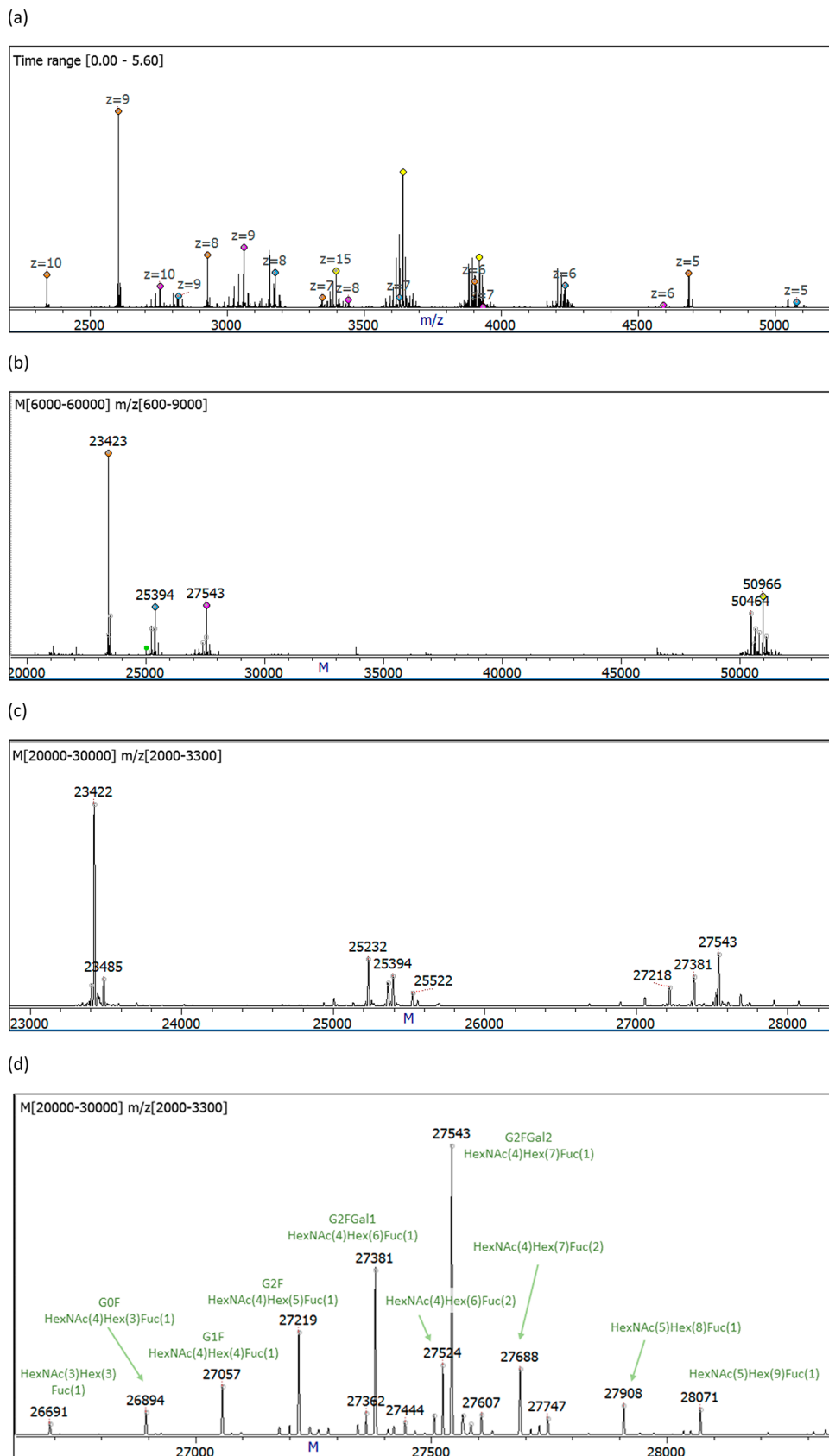
(a)

(b)

(c)

(d)

Figure 5. High-resolution native mass spectra and deconvoluted masses of the IdeS-digested and reduced mAb cetuximab. Deconvolution of the full *m/z* range (a) of cetuximab shows mass clusters (b) at about 23.4, 25.4, 27.5, and 50.5 kDa, corresponding to the light-chain LC, the glycosylated Fc/2, the glycosylated Fd, and the glycosylated Fc, respectively. A zoom of the 23−28 kDa range (c) shows good agreement with the theoretical masses of 23 423 Da for the LC with intrachain disulfide bonds, 25 233 Da for Fc/2 + G0F, and 27 543 for Fd + G2FGal(2). A further zoom in of the 26−29 kDa range (d) shows the more complicated Fab-arm/Fd glycosylation, including Gal-*α*-Gal and antennal fucosylation.

in other inverse problems arising in bioinformatics including phylogeny reconstruction from genomic data and protein inference from proteomics data. Because of its use of parsimony, Protein Metrics Intact gave fewer and smaller artifact peaks than Protein Deconvolution 4.0 on a complicated monomer/dimer example. Artifact reduction is important whenever the sample contains, or could possibly contain, molecules spanning a wide mass range, for example, light and heavy chains, monomers and dimers, or full proteins and clips.

Another contribution of the work presented here may seem obvious and unimportant, but we believe it is fundamental and far-reaching. This contribution is the "factorization" of charge deconvolution into two subproblems: charge inference and super-resolution. The two subproblems are not closely connected, even though they can both be solved by iterative algorithms. In the case of maximum entropy methods, the two subproblems are actually antagonistic, as accurate charge inference tends to decrease entropy and super-resolution explicitly aims to increase entropy. Decoupling the two problems will enable mass spectrometrists to work on charge inference, a problem unique to the field, while borrowing and adapting well-developed super-resolution algorithms from astronomy, geophysics, and so forth.

Although we chose the samples primarily as demonstrations of the new algorithm, our studies did reveal some unexpected characteristics of the targeted mAbs and properdin. For properdin we identified several novel low abundance proteoforms harboring triantennary N-glycans, seemingly exclusively on proteoforms with 15 C-mannosylations. These new proteoforms went unnoticed in our previous work due to the presence of salt adducts and the lack of a charge deconvolution program that could handle such difficult data. For daclizumab, we found both N-terminal extension and GG clipping, which, to our knowledge, have not been previously published. Such information is important to drug manufacturers because the clipped proteoform may have completely different therapeutic effects than the intact monoclonal antibody.[29] Our daclizumab sample, however, was already quite old and possibly past its expiration date, so the clipping may be due to extended storage. For cetuximab, we showed an analysis of a mAb with both Fc and Fd glycosylation using a combination of native MS with the new deconvolution algorithm, along with IdeS digestion to separate subunits, and bottom-up proteomics to confirm identified glycoforms, including glycans on the Fd site with antennal fucose.

As demonstrated in other studies,[16−18,25,32,33] native MS proves to be advantageous for analysis of mAbs and plasma glycoproteins. Native MS gives greater separation between charge states.[34] Without this separation, properdin and cetuximab would most likely give overlapping m/z states, which would seriously hamper deconvolution and visual validation. Another advantage of native MS for these target molecules is improved dynamic range; fewer charge states and lower charge means that there is more trap capacity available for minor species, such as the clipped N-terminal sequence in daclizumab. On the contrary, native MS generally requires more starting material than intact MS on denatured proteins, and native MS can lose resolution on FTICR and Orbitrap instruments by shifting the signal to higher m/z. Neither of these disadvantages, however, applies to typical analyses of therapeutic mAbs because sample is usually abundant, and FTMS resolution is more often limited by isotopic spread or adducts than by instrument resolution.

Finally, intact MS in either native or denaturing conditions provides a clear qualitative and quantitative survey of all of the proteoforms distinguishable by mass, thereby helping to identify which modifications need to be looked for in complementary bottom-up or middle-down data. The future analysis of protein therapeutics and plasma proteins is likely to rely on hybrid MS methods, complemented by advanced bioinformatics methods to analyze and integrate the data from each of the information channels. We look forward to equally rapid progress in bioinformatics to keep pace with the rapid development in instruments and experimental methods.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jproteome.7b00839.

> Table S1. Amino acid sequences of the analyzed proteins. Table S2. Therapeutic antibodies. Figure S1. Simultaneous deconvolution of monomer and dimer. Figure S2. Properdin qualitative and quantitative proteoform profile. Figure S3. Improved properdin qualitative and quantitative proteoform profile. Figure S4. Annotated glycopeptide fragmentation spectra of properdin peptides with triantennary glycans. Figure S5. Annotated peptide fragmentation of daclizumab revealing the GG clipping. Figure S6. Annotated peptide fragmentation spectra of daclizumab showing part of the signal peptide. Table S3. Observed glycoforms of the cetuximab Fd fragment. (PDF)

AUTHOR INFORMATION

Corresponding Authors

*M.B.: E-mail: bern@proteinmetrics.com. Tel: +1 650 412-4210.
*A.J.R.H.: E-mail: a.j.r.heck@uu.nl.

ORCID

Albert J. R. Heck: 0000-0002-2405-4404

Notes

M.B., Y.J.K., and C.B. are founders and part owners of Protein Metrics, Inc.
The authors declare the following competing financial interest(s): M.B., Y.J.K., W.T., C.B., E.C., D.K., and K. I.S. are employees of Protein Metrics, Inc., who intend to commercialize this software package.

REFERENCES

(1) Ferrige, A. G.; Seddon, M. J.; Green, B. N.; Jarvis, S. A.; Skilling, J.; Staunton, J. Disentangling electrospray spectra with maximum entropy. *Rapid Commun. Mass Spectrom.* **1992**, *6* (11), 707−711.
(2) Ferrige, A. G.; Seddon, M. J.; Jarvis, S.; Skilling, J.; Welch, J. The Application of Maxent to Electrospray Mass-Spectrometry. In *Maximum*

Entropy and Bayesian Methods. Fundamental Theories of Physics (An International Book Series on The Fundamental Theories of Physics: Their Clarification, Development and Application); Smith, C. R., Erickson, G. J., Neudorfer, P. O., Eds.; Springer: Dordrecht, The Netherlands, 1992; Vol. *50*, pp 327−335.

(3) Marty, M. T.; Baldwin, A. J.; Marklund, E. G.; Hochberg, G. K. A.; Benesch, J. L. P.; Robinson, C. V. Bayesian deconvolution of mass and ion mobility spectra: From binary interactions to polydisperse ensembles. *Anal. Chem.* **2015**, *87* (8), 4370−4376.

(4) Lu, J.; Trnka, M. J.; Roh, S. H.; Robinson, P. J. J.; Shiau, C.; Fujimori, D. G.; Chiu, W.; Burlingame, A. L.; Guan, S. Improved Peak Detection and Deconvolution of Native Electrospray Mass Spectra from Large Protein Complexes. *J. Am. Soc. Mass Spectrom.* **2015**, *26* (12), 2141−2151.

(5) Chowdhury, S. K.; Katta, V.; Chait, B. T. Probing conformational changes in proteins by mass spectrometry. *J. Am. Chem. Soc.* **1990**, *112* (24), 9012−9013.

(6) Katta, V.; Chait, B. T. Observation of the heme-globin complex in native myoglobin by electrospray-ionization mass spectrometry. *J. Am. Chem. Soc.* **1991**, *113* (22), 8534−8535.

(7) Robinson, C. V.; Radford, S. E. Weighing the evidence for structure: electrospray ionization mass spectrometry of proteins. *Structure* **1995**, *3* (9), 861−865.

(8) Loo, J. a. Studying noncovalent protein complexes by electrospray ionization mass spectrometry. *Mass Spectrom. Rev.* **1997**, *16* (1), 1−23.

(9) Krutchinsky, A. N.; Ayed, A.; Donald, L. J.; Ens, W.; Duckworth, H. W.; Standing, K. G. Studies of Noncovalent Complexes in an Electrospray Ionization/Time-of-Flight Mass Spectrometer. In *Mass Spectrometry of Proteins and Peptides: Mass Spectrometry of Proteins and Peptides*; Chapman, J. R., Ed.; Humana Press: Totowa, NJ, 2000; pp 239−249.

(10) Rose, R. J.; Damoc, E.; Denisov, E.; Makarov, A.; Heck, A. J. High-sensitivity Orbitrap mass analysis of intact macromolecular assemblies. *Nat. Methods* **2012**, *9* (11), 1084−1086.

(11) van de Waterbeemd, M.; Fort, K. L.; Boll, D.; Reinhardt-Szyba, M.; Routh, A.; Makarov, A.; Heck, A. J. R. High-fidelity mass analysis unveils heterogeneity in intact ribosomal particles. *Nat. Methods* **2017**, *14* (3), 283−286.

(12) Gault, J.; Donlan, J. A. C.; Liko, I.; Hopper, J. T. S.; Gupta, K.; Housden, N. G.; Struwe, W. B.; Marty, M. T.; Mize, T.; Bechara, C.; et al. High-resolution mass spectrometry of small molecules bound to membrane proteins. *Nat. Methods* **2016**, *13* (4), 333−336.

(13) Snijder, J.; Van De Waterbeemd, M.; Damoc, E.; Denisov, E.; Grinfeld, D.; Bennett, A.; Agbandje-Mckenna, M.; Makarov, A.; Heck, A. J. R. Defining the stoichiometry and cargo load of viral and bacterial nanoparticles by orbitrap mass spectrometry. *J. Am. Chem. Soc.* **2014**, *136* (20), 7295−7299.

(14) Franc, V.; Yang, Y.; Heck, A. J. R. Proteoform Profile Mapping of the Human Serum Complement Component C9 Revealing Unexpected New Features of N-, O-, and C-Glycosylation. *Anal. Chem.* **2017**, *89* (6), 3483−3491.

(15) Lössl, P.; Brunner, A. M.; Liu, F.; Leney, A. C.; Yamashita, M.; Scheltema, R. A.; Heck, A. J. R. Deciphering the interplay among multisite phosphorylation, interaction dynamics, and conformational transitions in a tripartite protein system. *ACS Cent. Sci.* **2016**, *2* (7), 445−455.

(16) Yang, Y.; Liu, F.; Franc, V.; Halim, L. A.; Schellekens, H.; Heck, A. J. R. Hybrid mass spectrometry approaches in glycoprotein analysis and their usage in scoring biosimilarity. *Nat. Commun.* **2016**, *7*, 13397.

(17) Yang, Y.; Wang, G.; Song, T.; Lebrilla, C. B.; Heck, A. J. R. Resolving the micro-heterogeneity and structural integrity of monoclonal antibodies by hybrid mass spectrometric approaches. *MAbs* **2017**, *9* (4), 638−645.

(18) Marcoux, J.; Champion, T.; Colas, O.; Wagner-Rousset, E.; Corvaia, N.; Van Dorsselaer, A.; Beck, A.; Cianferani, S. Native mass spectrometry and ion mobility characterization of trastuzumab emtansine, a lysine-linked antibody drug conjugate. *Protein Sci.* **2015**, *24* (8), 1210−1223.

(19) van de Waterbeemd, M.; Snijder, J.; Tsvetkova, I. B.; Dragnea, B. G.; Cornelissen, J. J.; Heck, A. J. R. Examining the Heterogeneous Genome Content of Multipartite Viruses BMV and CCMV by Native Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **2016**, *27* (6), 1000−1009.

(20) Richardson, W. H. Bayesian-Based Iterative Method of Image Restoration. *J. Opt. Soc. Am.* **1972**, *62* (1), 55−59.

(21) Lucy, L. B. An iterative technique for the rectification of observed distributions. *Astron. J.* **1974**, *79* (6), 745.

(22) Berger, A. L.; Pietra, V. J. D.; Pietra, S. a. D. A maximum entropy approach to natural language processing. *Comput. Linguist.* **1996**, *22* (1), 39−71.

(23) Beck, A.; Wagner-Rousset, E.; Ayoub, D.; Van Dorsselaer, A.; Sanglier-Cianférani, S. Characterization of therapeutic antibodies and related products. *Anal. Chem.* **2013**, *85* (2), 715−736.

(24) Ayoub, D.; Jabs, W.; Resemann, A.; Evers, W.; Evans, C.; Main, L.; Baessmann, C.; Wagner-Rousset, E.; Suckau, D.; Beck, A. Correct primary structure assessment and extensive glyco-profiling of cetuximab by a combination of intact, middle-up, middle-down and bottom-up ESI and MALDI mass spectrometry techniques. *MAbs* **2013**, *5* (5), 699−710.

(25) Rosati, S.; Yang, Y.; Barendregt, A.; Heck, A. J. R. Detailed mass analysis of structural heterogeneity in monoclonal antibodies using native mass spectrometry. *Nat. Protoc.* **2014**, *9*, 967.

(26) Gilar, M.; Belenky, A.; Wang, B. H. High-throughput biopolymer desalting by solid-phase extraction prior to mass spectrometric analysis. *J. Chromatogr. A* **2001**, *921* (1), 3−13.

(27) Bern, M.; Kil, Y. J.; Becker, C. Byonic: Advanced peptide and protein identification software. *Curr. Protoc. Bioinforma* **2012**, No. No. SUPPL. 40, 1−17.

(28) Campuzano, I. D. G.; Netirojjanakul, C.; Nshanian, M.; Lippens, J. L.; Kilgour, D. P. A.; Van Orden, S. L.; Loo, J. Native-MS Analysis of Monoclonal Antibody Conjugates by Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. *Anal. Chem.* **2018**, *90*, 745−751.

(29) Vlasak, J.; Ionescu, R. Fragmentation of monoclonal antibodies. *MAbs* **2011**, *3* (3), 253−263.

(30) Chevreux, G.; Tilly, N.; Bihoreau, N. Fast analysis of recombinant monoclonal antibodies using IdeS proteolytic digestion and electrospray mass spectrometry. *Anal. Biochem.* **2011**, *415* (2), 212−214.

(31) Janin-Bussat, M.-C.; Tonini, L.; Huillet, C.; Colas, O.; Klinguer-Hamour, C.; Corvaïa, N.; Beck, A. Cetuximab Fab and Fc N-Glycan Fast Characterization Using IdeS Digestion and Liquid Chromatography Coupled to Electrospray Ionization Mass Spectrometry. In *Glycosylation Engineering of Biopharmaceuticals: Methods and Protocols*; Beck, A., Ed.; Humana Press: Totowa, NJ, 2013; pp 93−113.

(32) Rosati, S.; Rose, R. J.; Thompson, N. J.; Van Duijn, E.; Damoc, E.; Denisov, E.; Makarov, A.; Heck, A. J. R. Exploring an orbitrap analyzer for the characterization of intact antibodies by native mass spectrometry. *Angew. Chem., Int. Ed.* **2012**, *51* (52), 12992−12996.

(33) Parsons, T. B.; Struwe, W. B.; Gault, J.; Yamamoto, K.; Taylor, T. A.; Raj, R.; Wals, K.; Mohammed, S.; Robinson, C. V.; Benesch, J. L. P.; et al. Optimal synthetic glycosylation of a therapeutic antibody. *Angew. Chem., Int. Ed.* **2016**, *55* (7), 2361−2367.

(34) Yang, Y.; Franc, V.; Heck, A. J. R. Glycoproteomics: A Balance between High-Throughput and In-Depth Analysis. *Trends Biotechnol.* **2017**, *35* (7), 598−609.