Enhancing Accuracy in Molecular Weight Determination of Highly Heterogeneously Glycosylated Proteins by Native Tandem Mass Spectrometry

Guanbo Wang,†‡§ Rob N. de Jong,† Ewald T. J. van den Bremer,‖ Paul W. H. I. Parren,‖⊥ and Albert J. R. Heck†‡§

†Biomolecular Mass Spectrometry and Proteomics, Bijvoet Center for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands
‡Netherlands Proteomics Centre, Padualaan 8, 3584 CH Utrecht, The Netherlands
§School of Chemistry and Materials Science, Nanjing Normal University, 1 Weyuan Road, Nanjing, Jiangsu 210023, China
‖Genmab, Yelaan 60, 3584 CM Utrecht, The Netherlands
⊥Department of Immunohematology and Blood Transfusion, Leiden University Medical Center, Albinusdreef 2, 2333 ZA Leiden, The Netherlands

Supporting Information

ABSTRACT: The determination of molecular weights (MWs) of heavily glycosylated proteins is seriously hampered by the physicochemical characteristics and heterogeneity of the attached carbohydrates. Glycosylation impacts protein migration during sodium dodecyl sulfate (SDS)-polyacrylamide gel electrophoresis (PAGE) and size-exclusion chromatography (SEC) analysis. Standard electrospray ionization (ESI)-mass spectrometry does not provide a direct solution as this approach is hindered by extensive interferences of ion signals caused by closely spaced charge states of broadly distributed glycoforms. Here, we introduce a native tandem MS-based approach, enabling charge-state resolution and charge assignment of protein ions including those that escape mass analysis under standard MS conditions. Using this method, we determined the MW of two model glycoproteins, the extracellular domains of the highly and heterogeneously glycosylated proteins CD38 and epidermal growth factor receptor (EGFR), as well as the overall MW and binding stoichiometries of these proteins in complex with a specific antibody.

Extensively glycosylated proteins occur in almost all living organisms. The glycan chains attached to the protein backbone affect protein stability, solubility, and conformation and also modulate functional activity and biological processes including molecular recognition, signaling, immune defense, inflammation, viral invasion, parasitic infections, etc. Owing to the high number of glycosylation sites and the various co-occurring oligosaccharide patterns at each site (e.g., composition, sequence, and linkage of monosaccharide residues), glycoproteins exist naturally in a microheterogeneous mixture of a number of glycoforms/proteoforms, exhibiting a spectrum of biological activities and therapeutic efficacies and a wide distribution of the overall molecular weight (MW) of glycoproteins. The average MW of the ensemble of glycoforms is an indispensable parameter for determination of oligomeric states, binding stoichiometries, and actual mass concentration (based on, e.g., UV absorbance) and for derivation or estimation of various physical properties (e.g., density, viscosity, diffusion, sedimentation, electrophoretic mobility, specific heat capacity, etc.). Moreover, MW is helpful for comparison of different batches of glycoproteins and assessment of the efficiency of glycan trimming or removal. Unfortunately, conventional biophysical approaches (see below) to measure MW dependent on biochemical properties of these glycoproteins largely suffer from insufficient accuracy.

We seek to enhance the accuracy in the assessment of the average MW by direct mass measurement using mass spectrometry (MS), which has already been used as a powerful tool to localize the glycosylation sites and profile site-specific glycosylation of glycoproteins. To avoid disrupting the integrity of proteins and provide the mass information with higher precision under near-physiological conditions, we performed measurements at the intact protein level using native tandem MS. The model systems we focus on are the soluble extra-cellular domains of the CD38 and epidermal growth factor receptor (EGFR), two highly heterogeneously glycosylated proteins. For method validation, we estimate the carbohydrate contents of both proteins based on the number of glycosylation sites present (4 and 9) and the MW range of the individual N-glycan chains (1.4–2.7 kDa) determined by analysis at the glycopeptide level (Table S1).

Received: December 27, 2016
Accepted: April 1, 2017
Published: April 6, 2017

DOI: 10.1021/acs.analchem.6b05129
Anal. Chem. 2017, 89, 4793−4797
EXPERIMENTAL SECTION

Materials. Soluble His-tagged extracellular domains of EGFR (sEGFR) and CD38 (sCD38) and anti-CD38-IgG1 and anti-EGFR-IgG1 monoclonal antibodies (mAbs) were expressed and purified as previously described.17 sCD38 was also produced in a nonglycosylated format where the four residues N100, N164, N209, and N219 that are glycosylated in the natural form were replaced with aspartic acid. All other chemicals were of analytical grade or higher. The deglycosylated formats of sCD38, sEGFR, and mAbs were produced through PNGase F (Roche, Germany) treatment following the standard procedures under denaturing or native (for mAb only) conditions.

Native Mass Spectrometry. Native MS experiments were carried out with a Q-TOF-2 (Waters) mass spectrometer modified for enhanced CID and equipped with a standard static nano-electrospray ionization (ESI) source.18,19 Samples were electrosprayed from an aqueous 150 mM ammonium acetate solution (pH adjusted to 7.5 with ammonium hydroxide). 20% (mol/mol) triethylammonium acetate (TEAA) was added to the sample buffers for charge reduction.20,21 sCD38, sEGFR, and their antibody-bound complexes were assembled through incubation at 21 °C for at least 10 min. Tandem MS was performed by mass selection of precursor ions using a quadrupole mass analyzer with an effective width of 140 Th at half-maximum of precursor peak, followed by their collision with xenon molecules. The details of other measurements are presented in the Supporting Information.

RESULTS AND DISCUSSION

Using standard proteins as MW markers, our sodium dodecyl sulfate (SDS)-polyacrylamide gel electrophoresis (PAGE)
measurements of sCD38 and sEGFR resulted in deviated MW values which hinted at a carbohydrate content of about 14 and 40 kDa, respectively (Figure 1A, Tables 1 and S1), close to the carbohydrate content of the full-length proteins previously determined by SDS-PAGE.22,23 If these numbers are correct, the MW of each glycan chain would be 3.6 and 4.5 kDa on average, far exceeding the upper MW limit of the identified glycans. Such deviations, which are largely due to the decrease in migration of glycoproteins resulting from poor glycan–SDS interactions, are frequently encountered in MW determination of glycoproteins using SDS-PAGE.24 Notably, MW determined by size-exclusion chromatography (SEC) deviates even further from the estimated values (Figure 1B, Table 1). Since the SEC elution of a protein is directly determined by the hydrodynamic radius rather than the mass, the MW determination is influenced by factors such as geometric shape,25 which cannot be corrected for by calibrating with protein standards. Furthermore, compared to the contribution of the protein content of the glycoprotein to the hydrodynamic radius, the contribution of the attached glycan may be disproportionate, leading to an exaggeration of the glycoprotein MW.26 Coupling multiangle light-scattering (MALS) with SEC enables one to determine to some extent the MW of separated species based on their light-scattering properties.26,27 However, the accuracy in MW determination of glycoproteins by MALS is hampered by multiple assumptions and approximations.27,28

As an approach that directly determines MW, mass spectrometry (MS) is generally capable of measurement of the mass of nonglycosylated and slightly glycosylated proteins. Since proteins become multiply charged in ESI and m/z (mass-to-charge ratio) values are measured, the MW calculation by multiple assumptions and approximations.27,28

To improve the accuracy in MW determination using native ESI-MS, we introduce here a 3-step “dilution-tandem MS” strategy (Figure 2A), aiming to resolve and assign the charge states of the glycoprotein ions. Step 1 “dilutes” the contribution of the heterogeneous carbohydrate chains to the overall mass dispersion by attaching the glycoprotein to a high MW protein. Therefore, sCD38 was incubated with a specific and high-affinity human IgG1 monoclonal antibody (mAb) against CD38. With excess sCD38, mAb/sCD38 complexes were assembled at a predominant stoichiometry of 1:2 as assessed by native MS,29 showing visibly better resolved charge states than for the uncomplexed sCD38 (Figure 2B). In Step 2, we used

Figure 2. 3-step approach enhancing the accuracy in MW determination of highly glycosylated proteins using native tandem ESI-MS. (A) Conceptual illustration with essential steps indicated by the color-coded arrows. 1: Attaching the glycoprotein (green) of interest to a high-MW-homogeneous protein (e.g., mAb; gray) reduces the relative heterogeneity of the resulting high MW complex (M_p). 2: Addition of TEAA reduces the charge state of M_p to +i. 3: In tandem MS of M_p, the release of fragments (M_i) removes additional charges due to overlapping ion signals from all diverse glycoforms/proteoforms carrying different charges. Utilizing charge reduction in the native ESI process with triethylammonium acetate (TEAA)20,29 did not improve the spectral resolution substantially (Figure 1C). Although a rough estimation of the MW can be made using the centroid m/z of the unresolved signal envelope and the average charge state (z_{avg}) estimated on the basis of an empirical correlation between MW and z_{avg} detected in native MS20,31 (Table 1; see the Supporting Information for details), the precision of such estimations is influenced by the physical properties (e.g., density) of the analytes and the experimental parameters adopted for data acquisition which affect z_{avg}.32 In fact, it has been argued that the average charge state is correlated more precisely to the mass of nonglycosylated and slightly glycosylated proteins.32 In tandem MS of M_p, the release of fragments (M_i) removes additional charges from the residual (M_e) and provides constraints useful for charge and MW determination. (B) Native mass spectra of sCD38 showing the effects of Steps 1–3. In gray, the initial native MS spectrum of both the sCD38 and mAb/sCD38 complex is shown in the upper panel and in blue its charge-reduced mass spectrum is displayed; in the bottom panel, the tandem MS spectrum of mass-selected sCD38 ions is shown. (C) Heat maps presenting \( \sigma_M \) (green scale) and \( \Delta M \) (blue–red scale). The underlined numbers represent \( i_0 \) and \( j_0 \) and the bold, italic numbers represent \( i_c \) and \( j_c \). \( \sigma_M \) and \( \Delta M \) values calculated using \( i_0 \) and \( j_0 \) are boxed with thin lines, and those calculated using \( i_c \) and \( j_c \) are boxed with thick lines.
TEAA as an additive in the spray solution to reduce charging of the ion, further increasing the spectral spacing between adjacent charge states (Figure 2B). Although the resulting charge-state resolution appears to allow direct spectrum deconvolution, determination of charge states of these ions is not straightforward since conventional algorithms frequently results in incorrect charge assignments for such high MW glycoproteins (see ref 37 for a detailed explanation on the origin of deviation). To solve this problem, in Step 3, we applied mass selection of a subpopulation of ions with limited mass dispersion. When subjecting these ions to tandem MS, we used collision-induced dissociation (CID), which yielded not only charge reduced species useful for the MW calculation but also additional constraints enabling more accurate charge calculations. As illustrated in Figure 2B, CID of the mass-selected (mAb)1/(sCD38)1 ions (where “i” stands for the unknown charge state) results in (i − 1)+ and (i − 2)+ charged species (see Figure S1 for detailed analysis) and release of N-terminal segments of the mAb heavy chain, with MW − 2). Following the conventional approach, we first determined and , i.e., values of and that lead to the least standard deviations of the corresponding MW (Figure 2C). Because the mass profiles of the isolated ions and the charge-reduced ions may be inconsistent, the proper assignments of and may not be per se but rather their neighboring integer values, despite the adequate resolution of the corresponding signal peaks. Accordingly, we determined and using an additional constraint defined by the mass balance rule for the CID reaction, i.e.,

\[
M_p = M_i + M_f
\]

where is the MW of precursor ions calculated using , is the average MW of the fragment ions, and is the MW of the residual ions calculated using (see the Supporting Information). Then and should give the least absolute value of , which is expressed as

\[
\Delta M = M_p - (M_i + M_f)
\]

As illustrated by the heat maps presenting and derived from a range of combinations (Figure 2C), and (25 and 18) resulted in a of 7.8 kDa which is significantly larger than the experimental error defined by (25) and (18), satisfying the mass profile constraint but violating the mass balance constraint. 23 and 17 that satisfy both constraints are determined as and . Thus, the MW of sCD38 can be calculated as

\[
M_{\text{CD38}} = \frac{(M_{\text{complex}} - M_{\text{mAb}})}{2}
\]

where is known, and is calculated using and the corresponding centroid m/z values exhibited by the spectra of ion ensembles (Figure 2B). With the MW of the polypeptide backbone (calculated on the basis of the amino acid sequence) subtracted from , the carbohydrate content was determined as 7.5 kDa, within the coarse range estimated on the basis of the data obtained at the glycopeptide level (Table 1, Table S1). In addition to the accuracy evidenced by such agreement, this approach was also validated in terms of self-consistency (see Figure S2B for detailed analysis).

Following the aforementioned procedure, we next focused on the even more extensively glycosylated sEGFR protein. After incubating sEGFR with a specific high affinity human IgG1 mAb at a molar ratio of 2:1, we detected two different mAb/sEGFR complexes in native MS (denoted as “I” and “II” in Figure 3). Although the suboptimal resolution of the signals hindered the direct mapping of the charge states of isolated proteoforms to those of the ensemble, we nevertheless determined by averaging the derived from multiple isolation windows spanning multiple charge states (see Figure S3). The content of sEGFR in Complex II was determined to be twice that in Complex I (Figure S3B,C), indicating incorporation of 1 and 2 copies of sEGFR in Complexes I and II, respectively. Such stoichiometries agree with the previous observation that this mAb binds to EGF-expressing cells at least partly in a bivalent manner. However, derived from the two complex species were consistent (Figure S3), and the accordingly calculated carbohydrate content lies within the range of expected values (Table S1).

Conclusions

In summary, we developed an MS-based approach to measure heterogeneously glycosylated proteins. With this calibrate-free approach, we determined the average MW of two glycoproteins whose carbohydrate content can contribute more than 20% to the total MW. The MW determination is based on the mass of the analytes per se and is thus independent of the factors (e.g., shape, radius, electrophoretic behavior) that typically lead to significant systematic errors. In comparison to the MS analysis at the glycopeptide level, which provides a rough estimate of the mass, our approach directly measures the mass distribution of intact glycoproteins. Since the “dilution” of heterogeneity is performed regardless of the nature of the intermolecular interactions, this approach can be applied to various non-covalently binding and covalent conjugation systems, allowing a broad range of glycoproteins to be analyzed. In addition, the compatibility of this approach with native conditions enabled MW determination of intact glycoproteins simultaneously with determination of binding stoichiometry.
Detailed experimental procedures; MW estimation using empirical equations; supplementary table of carbohydrate content of model proteins; supplementary figures including isolation window of ions, sequence information derived from the CID measurements and self-consistency of the method demonstrated by the results of charge assignment and MW determination, and simultaneous determination of MW of sEGFR and its binding stoichiometry with the anti-EGFR mAb (PDF)

Notes
The authors declare the following competing financial interest(s): As disclosed also in the manuscript, G.W. and A.J.R.H. received Genmab funding. R.N.d.J., E.T.J.v.d.B., and P.W.H.I.P. are Genmab employees and have stock and/or warrants.

ACKNOWLEDGMENTS
This work was supported by The Netherlands Organisation for Scientific Research (NWO) via the Roadmap Initiative Proteins@Work (project number 184.032.201) and the TOP-Punt Grant 718.015.003 for A.J.R.H. G.W. and A.J.R.H. received Genmab funding. A.J.R.H. acknowledges additional support through the European Union Horizon 2020 programme FET-OPEN project MSmed, project number 686547.

REFERENCES
(1) Spiro, R. G. Glycobiology 2002, 12, 43R–56R.
(2) Notthaft, H.; Szymanski, C. M. J. Biol. Chem. 2013, 288, 6912–6920.
(3) Geyer, H.; Geyer, R. Biochim. Biophys. Acta, Proteins Proteomics 2006, 1764, 1853–1869.
(4) Rudd, P. M.; Dwek, R. A. Crit. Rev. Biochem. Mol. Biol. 1997, 32, 1–100.
(5) Rademacher, T. W.; Parekh, R. B.; Dwek, R. A. Annu. Rev. Biochem. 1998, 57, 785–838.
(6) Yang, Y.; Liu, F.; Franc, V.; Halim, L. A.; Schellekens, H.; Heck, A. J. Nat. Commun. 2016, 7, 13397.
(7) Rudd, P. M.; Joao, H. C.; Coghill, E.; Fiten, P.; Saunders, M. R.; Opdenakker, G.; Dwek, R. A. Biochemistry 1994, 33, 17–22.
(8) Chen, W.; Smeekens, J. M.; Wu, R. J. Proteome Res. 2014, 13, 1466–1473.
(9) Faridmoyayer, A.; Fentabli, M. A.; Mills, D. C.; Klassen, J. S.; Feldman, M. F. J. Bacteriol. 2007, 189, 8088–8098.
(10) Han, L.; Costello, C. E. Biochemistry (Moscow) 2013, 78, 710–720.
(11) Thyssen-Andersen, M.; Packer, N. H. Biochim. Biophys. Acta, Proteins Proteomics 2014, 1844, 1437–1452.
(12) Mehta, K.; Shahid, U.; Malavasi, F. FASEB J. 1996, 10, 1408–1417.
(13) Lurje, G.; Lenz, H. J. Oncology 2009, 77, 400–410.
(14) Gao, Y.; Mehta, K. Mol. Cell. Biochem. 2007, 295, 1–7.
(15) Zhen, Y. J.; Caprioli, R. M.; Staros, J. V. Biochemistry 2003, 42, 5478–5492.
(16) Hara-Yokoyama, M.; Kikimoto-Niino, M.; Terasawa, K.; Harumiya, S.; Podymina-Inoue, K. A.; Hino, N.; Sakamoto, K.; Itoh, S.; Hashii, N.; Hiruta, Y.; Kawasaki, N.; Mishima-Tsumagari, C.; Kaitou, Y.; Matsumoto, T.; Wakiyama, M.; Shirouzu, M.; Kasama, T.; Takayanagi, H.; Utsunomiya-Tate, N.; Takatsu, K.; Katada, T.; Hiramatsu, Y.; Yokoyama, S.; Yanagishita, M. Structure 2012, 20, 1585–1595.
(17) Wang, G.; de Jong, R. N.; van den Bremer, E. T. J.; Beurskens, F. J.; Labrijn, A. F.; Ugurlar, D.; Gros, P.; Schuurman, J.; Parren, P.; Heck, A. J. R. Mol. Cell 2016, 63, 135–145.
(18) van den Heuvel, R. H.; van Duijn, E.; Mason, H.; Synowsky, S. A.; Lorenzen, K.; Versluis, C.; Brouns, S. J.; Langridge, D.; van der Oost, J.; Hoyes, J.; Heck, A. J. Anal. Chem. 2006, 78, 7473–7483.
(19) Snijder, J.; Heck, A. J. Annu. Rev. Chem. 2014, 7, 43–64.
(20) van de Waterbeemd, M.; Snijder, J.; Tsvetkova, O. B.; Dragnea, B. G.; Cornelissen, J. J.; Heck, A. J. R. Am. Soc. Mass Spectrom. 2016, 27, 1000–1009.
(21) Pacholzar, K. J.; Barran, P. E. EuPA Open Proteomics 2016, 11, 23–27.
(22) Jackson, D. G.; Bell, J. I. J. Immunol. 1990, 144, 2811–2815.
(23) Soderquist, A. M.; Carpenter, G. J. Biol. Chem. 1984, 259, 12586–12594.
(24) Segrest, J. P.; Jackson, R. L.; Andrews, E. P.; Marchesi, V. T. Biochem. Biophys. Res. Commun. 1971, 44, 390–395.
(25) Hong, P.; Koza, S.; Bouvier, E. S. P. J. Liq. Chromatogr. Relat. Technol. 2012, 35, 2923–2928.
(26) Wen, J.; Arakawa, T.; Phojo, J. S. Anal. Biochem. 1996, 240, 155–166.
(27) Oliva, A.; Llabres, M.; Farina, J. B. Curr. Drug Discovery Technol. 2004, 1, 229–242.
(28) Arakawa, T.; Wen, J. Anal. Biochem. 2001, 299, 158–161.
(29) Pacholzar, K. J.; Barran, P. E. EuPA Open Proteomics 2016, 11, 23–27.
(30) Tolic, L. P.; Anderson, G. A.; Smith, R. D.; Brothers, H. M.; Spindler, R.; Tomalia, D. A. Int. J. Mass Spectrom. Ion Processes 1997, 165, 405–418.
(31) Fernandez de la Mora, J. Anal. Chem. Acta 2000, 406, 93–104.
(32) Samalikova, M.; Matecko, I.; Muller, N.; Grandori, R. Anal. Bioanal. Chem. 2004, 378, 1112–1123.
(33) Kaltashov, I. A.; Mohimen, A. Anal. Chem. 2005, 77, 5370–5379.
(34) Dyachenko, A.; Wang, G. B.; Belov, M.; Makarov, A.; de Jong, R. N.; van den Bremer, E. T. J.; Parren, P.; Heck, A. J. R. Anal. Chem. 2015, 87, 6095–6102.
(35) Mann, M.; Meng, C. K.; Fenn, J. B. Anal. Chem. 1989, 61, 1700–1708.
(36) Tito, M. A.; Tars, K.; Vajegard, K.; Hajdu, J.; Robinson, C. V. J. Am. Chem. Soc. 2000, 122, 3550–3551.
(37) Abzalimov, R. R.; Kaltashov, I. A. Anal. Chem. 2010, 82, 7523–7526.
(38) Wang, G.; Johnson, A. J.; Kaltashov, I. A. Anal. Chem. 2012, 84, 1718–1724.
(39) Smith, R. D.; Barina, C. J.; Udhay, H. R. J. Phys. Chem. 1989, 93, 5019–5022.
(40) Bleeker, W. K.; van Bueren, J. J. L.; van Ojik, H. H.; Gerritsen, A. F.; Pluyter, M.; Houtkamp, M.; Halk, E.; Goldstein, J.; Schuurman, J.; van den Dijk, M. A.; van de Winkel, J. G. J.; Parren, P. J. Immunol. 2004, 173, 4699–4707.