Evaluation of the extent of haptoglobin glycosylation using orthogonal intact-mass MS approaches

Yang Yang, Jake W. Pawlowski,[¶] Ian J. Carrick,[§] and Igor A. Kaltashov*

Department of Chemistry, University of Massachusetts-Amherst, 240 Thatcher Road, Amherst, MA 01003

Abstract

Intact-mass measurements are becoming an increasingly popular in mass spectrometry (MS) based protein characterization, as they allow the entire complement of proteoforms to be evaluated within a relatively short time. However, applications of this approach are currently limited to systems exhibiting relatively modest degrees of structural diversity, as the high extent of heterogeneity frequently prevents straightforward MS measurements. Incorporation of limited charge reduction into electrospray ionization (ESI) MS measurements provides an elegant way to obtain meaningful information on most heterogeneous systems, yielding not only the average mass of the protein, but also the mass range populated by various proteoforms. Application of this approach to characterization of two different phenotypes of haptoglobin (1-1 and 2-1) provides evidence of a significant difference in their extent of glycosylation, with the glycan load of phenotype 2-1 being notably lighter. More detailed characterization of their glycosylation patterns is enabled by the recently introduced cross-path reactive chromatography (XP-RC) with on-line MS detection, a technique that combines chromatographic separation with in-line reduction of disulfide bonds to generate metastable haptoglobin subunits. Application of XP-RC to both haptoglobin phenotypes confirms that no modifications are present within their light chains, and provides a wealth of information on glycosylation patterns of the heavy chains. The haptoglobin 1-1 glycans are mature fully sialylated biantennary structures that exhibit high degrees of fucosylation. In contrast, phenotype 2-1 contains a significant fraction of incomplete biantennary structures and exhibit significantly lower levels of sialylation and fucosylation. The glycosylation patterns deduced from the XP-RC/MS measurements are in agreement with the conclusions of haptoglobin analysis by limited charge reduction, suggesting that the latter can be employed in situations when a fast assessment of a protein heterogeneity is needed (*e.g.*, comparability studies of biopharmaceutical products).

Introduction

Intact mass analysis[1,2] continues to enjoy growing popularity as a means of mass spectrometry (MS) based characterization of proteins, as it allows the structural diversity of various protein samples to be evaluated on a relatively short time scale. As such, it is increasingly used in biopharmaceutical analysis for tasks ranging from characterization of biosimilars[3] to pharmaco-kinetics and biotransformation studies,[4] as well as support of upstream process development.[4] It is also likely to make a significant impact in the field of clinical analysis, where the ability to evaluate the proteoform pool size for a particular protein will improve the accuracy of the existing (non-MS based) diagnostic methods of biomarker measurements.[5] Furthermore, the ability to profile the entire complement of proteoforms will undoubtedly be a boon for the development of new methods of clinical analysis. Glycoform assessment is particularly promising in this regard, as the glycosylation profiles of many proteins appear to have high diagnostic (and in some cases prognostic) value vis-à-vis a variety of pathologies[6] ranging from cancer[7] to inflammation.[8] While the de-novo characterization of protein glycoforms involves both localization of the glycosylation sites within the protein and determination of their structure,[9] monitoring the changes in glycosylation patterns frequently does not require such time- and labor-intensive efforts and can be accomplished by observing changes in the protein mass distribution.[10] Direct infusion electrospray ionization (ESI) MS can be used for intact glycoform analysis of systems with relatively low levels of glycosylation (such as monoclonal antibodies built on IgG templates),[10] although it can certainly benefit from incorporating an on-line separation step prior to MS analysis.[11] Hydrophilic interaction chromatography (HILIC) remains the most popular option for the glycoform separation prior to the MS step;[12,13] however, it requires the use of mobile phases with a high organic solvent content, inevitably leading to the glycoprotein denaturation. Apart from being problematic for proteins having low tolerance to organic co-solvents, this also creates challenges for the analysis of large oligomeric proteins in which variation of the number of polypeptide subunits within the protein molecule is another source of intrinsic heterogeneity (in addition to glycosylation). Furthermore, the vast majority of successful applications of HILIC in the field of intact glycoproteins encompass separation of glycoforms of relatively small (< 35 kDa) proteins, with examples of employing this separation modality to large glycoproteins remaining rare.[14]

In this work we use an alternative approach to evaluate structural heterogeneity of larger glycoproteins (90-190 kDa) using a combination of native MS complemented with limited charge reduction[15] and a recently introduced cross-path reactive chromatography (XP-RC) platform with on-line MS detection.[16] The model used to test this approach is haptoglobin (Hp), an acute phase plasma glycoprotein, which captures free Hb in circulation during intravascular hemolysis and transports it to macrophages;[17] above and beyond being a safeguard against

the hemolysis-associated oxidative damage, Hp may also act as a modulator of inflammation.[18,19] While the latter is a commonly acknowledged hallmark of severe COVID-19,[20] there is also ample evidence pointing at the involvement of hemolysis in the progression of this disease,[21] suggesting that haptoglobin is likely to be an important (although presently unacknowledged) player in the host organism's response to the SARS-CoV-2 infection. Importantly, Hp levels were observed to spike in response to respiratory coronavirus infections in both animal models[22] and humans,[23] where it was suggested to act as an important factor in lung tissue repair following the inflammatory (neutrophil-mediated) damage.[23] Furthermore, the levels of circulating Hp have prognostic value vis-à-vis the risk of in-hospital mortality among sepsis patients,[24] and certain Hp variants are linked to patients' susceptibility to acute respiratory distress syndrome during sepsis[25] (it was not until very recently that Hp expression levels have also been shown to correlate with the severity of COVID-19[26]).

As far as clinical diagnostics, Hp glycosylation profiles have shown a great promise as a means of differentiation between healthy and diseased subjects for a variety of pathologies ranging from autoimmune disorders[27] and cancer[28] to chronic alcoholism.[29] While detailed glycosylation profiling of Hp can be readily accomplished using standard glycan release assays with subsequent LC/MS analyses, intact-mass analysis (if feasible) would provide a significantly simplified approach to this task. Unfortunately, the large number of *N*-glycans present on Hp molecules (*vide infra*) make this task a very challenging undertaking, and requires highly specialized MS equipment.[30] Indeed, Hp is a multi-meric protein incorporating two types of polypeptide chains, non-glycosylated light chains (L) and extensively glycosylated heavy chains (H), which are cross-linked by disulfide bonds[31] (see ***Supplementary Material*** for more detail). In humans, Hp is expressed in two allelic forms, HP1 and HP2, which differ by size of the light chain.[32] The larger light chain (termed L* in this paper) incorporates three additional cysteine residues, one of which participates in formation of an external disulfide bond with either L or L* chain and, as a result, extensive multimerization of the protein. The L* chain is absent in only one of the three major phenotypes (termed Hp 1-1); this type of the protein has a well-defined tetrameric structure with an H-L-L-H connectivity.[17] The second phenotype (Hp 2-1) contains both L and L*-type light chains, giving rise to polymeric chains of HL* subunits (connected to each other via L*-chains), which are terminated with HL "endgroups" at both ends of the $(HL^*)_n$ oligomer. The L-chains are absent in the third major phenotype (Hp 2-2), meaning that the HL* polymerization can be terminated only via "chain cyclization," giving rise to circular $(HL^*)_n$ oligomeric structures (see ***Supplementary Material*** for more detail). Each H-chain incorporates four *N*-glycosylation sites; therefore, the minimal number of glycans per Hp molecule is eight (in Hp 1-1 and in the lowest-molecular weight isoforms of Hp 2-1, which have the same architecture as Hp 1-1, H-L-L-H, see

Supplementary Material).

The carbohydrate fraction of the Hp mass is close to 20%, making this protein highly heterogeneous. In fact, even average mass measurement for Hp is a challenging task (the protein mass spread gives rise to broad ionic peaks and complicates charge state assignment in ESI mass spectra for even the simplest form of the protein, Hp 1-1;[33] the MS analysis of other phenotypes is more difficult still[34]). In this work we use limited charge reduction[15] to determine average masses and mass ranges of Hp 1-1 and several isoforms of Hp 2-1. These measurements provide evidence of a significant difference in the extent of glycosylation of the H-chains within these two phenotypes, a notion that is validated by characterizing their glycosylation profiles using a recently introduced cross-path reactive chromatography (XP-RC) with on-line MS detection.[16] The agreement between these two orthogonal techniques suggests that ESI MS complemented with limited charge reduction can be used for tasks requiring a fast but nonetheless reliable assessment of a protein heterogeneity.

Experimental

Materials. Human Hp (phenotypes 1-1 and 2-1) were purchased from Athens Research & Technologies, Inc. (Athens, GA), and TCEP was purchased from Thermo-Fisher Scientific (Waltham, MA). All buffers, solvents and salts used in this work were of analytical grade or higher. The protein samples were reconstituted in 150 mM ammonium acetate (pH 6.9) to a final concentration of 1-5 μM for MS analyses (ESI MS and limited charge reduction) and 75 mM ammonium acetate (pH 5.5) for XP-RC/MS analyses.

Methods. All limited charge reduction measurements were carried out with a Synapt G2Si (Waters, Milford, MA) hybrid quadrupole/time-of-flight mass spectrometer equipped with a nanospray ion source and an ion mobility analyzer. The following set of parameters was used in the ESI interface region: capillary voltage, 1.5 kV; sampling cone voltage, 100 V; source offset, 80 V. Isolation of ionic populations for limited charge reduction in the trap cell was performed by setting the quadrupole LM resolution in the range of 4.3 - 4.7. Limited charge reduction of polycationic ions was triggered by introducing 1,3-dicyanobenzene anions after setting the trap wave height to 0.8-1.5 V; the discharge current was set at 8-20 μA. The MS data were analyzed by using Mass Lynx 4.1 (Waters, Milford, MA).

In-line (on-column) chemical reactions were carried uisng a TSKgel SuperSW mAb HTP (Tosoh, Tokyo, Japan) SEC column and an HP 1100 (Agilent, Santa Clara, CA) HPLC system. A flow rate of 0.15 mL/min was typically used in XP-RC/MS measurements for the analysis of Hp glycosylation patterns. The reduction plug was composed of 50 mM TCEP and 4M guanidinium chloride in the 75 mM ammonium acetate (pH 5.5) mobile phase; in the case of in-line reduction of Hp 2-1, the plug also contained acetonitrile (20% by volume). The plug was

introduced using a manual injector with a loop volume of 150 μL (Hp 1-1) or 365 μL (Hp 2-1), which was placed between the sample injector and the SEC column. The protein samples were injected with a delay time of 1 min with respect to the reagent plug injection. A flow splitter was used to direct *ca*. 10% of the eluate to MS for on-line mass measurements, which were performed with a SolariX 7 (Bruker Daltonics, Billerica, MA) Fourier transform ion cyclotron resonance (FT ICR) mass spectrometer equipped with a 7.0 T superconducting magnet and a standard ESI source. The following parameters were use in the ESI interface: capillary voltage, 4500 V; the dry gas (nitrogen) flow rate, 4.6 L/min, capillary temperature, 200 ^{O}C. Each mass spectrum was an average of five 16,000-point transients to increase the data acquisition rate. The data analysis was performed using Compass Data Analysis software (Bruker Daltonics).

Results and Discussion

ESI mass spectra of both Hp phenotypes used in this work (1-1 and 2-1) acquired under commonly used conditions (high organic solvent content and acidic pH) contain near-continuum distributions of the ionic signal spread over a wide *m/z* range (see the inset in **Figure 1**). The two mass spectra largely overlap, and neither contains significant discernable features that can be used to estimate the ionic charge states. However, the protein spectra acquired under near-native conditions (the neutral pH and the physiological ionic strength) contain abundant ionic signals with readily discernable features that can be interpreted as different charge states and (in the case of Hp 2-1) clusters of peaks corresponding to specific oligomers (see the main panel in **Figure 1**). The lowest-*m/z* cluster in the mass spectrum of Hp 2-1 (presumably representing the lowest-mass isoform, H-L-L-H) appears to overlap with the signal of Hp 1-1 (containing exclusively H-L-L-H oligomers). However, careful examination of the two distributions suggests that they are not identical to each other (indeed, while the peaks at *m/z* 4900 in these two mass spectra are almost ideally aligned with each other, noticeable deviations exist for the peaks at *m/z* 5250, and the incongruence becomes even more apparent for the peaks at *m/z* 5500).

Application of limited charge reduction to these groups of ionic peaks provides an explanation for this mismatch: the two distributions are misaligned with respect to the ionic charge states. Indeed, while the ionic species at the apex of the peak at *m/z* 5250 in the mass spectrum of Hp 1-1 have a charge state +18 (**Figure 2A**), their counterparts in the mass spectrum of Hp 2-1 have a charge state +17 (**Figure 2B**). Furthermore, limited charge reduction was applied to several ionic populations within *m/z* windows selected at 100 *m/z* unit intervals across the spectral range spanning the distance between the apexes of the two most abundant peaks in the mass spectrum of Hp 1-1 (**Figure 2A**). This allowed not only the mass of the most

abundant ions to be determined (93,620±130 Da), but also estimate the mass range of the species giving rise to the ionic signal in the mass spectrum of Hp 1-1 as 91.8-97.2 kDa. A similar analysis carried out for species giving rise to the ionic signal in the *m/z* region 5250-5550 in the mass spectrum of Hp 2-1 yields 89,450±120 Da as the mass of the most abundant species, and the mass range populated by these low-molecular weight Hp 2-1 species (having the same molecular architecture as Hp 1-1, *i.e*. H-L-L-H) as 85.8-96.0 kDa.

The noticeable difference between the masses of the protein species with identical molecular architecture present in the two Hp phenotypes is intriguing, as it points out to a significant difference in the glycosylation patterns of their H-chains. Indeed, the L-chains lack any *N*- or *O*-glycans, and although prolonged circulation in the plasma may result in protein glycation,[35] the latter cannot account for the large mass difference observed in our work. Assuming no PTMs are present within the L-chains of either Hp species, it is possible to estimate the average masses of the H-chains within Hp 1-1 molecules as 37,620±150 Da, and those within the Hp 2-1 as 35,536±150 Da. The polypeptide mass of the H-chain (calculated based on its amino acid sequence) is 27,265 Da; therefore, the carbohydrate component of this subunit in Hp 1-1 is 10,356 Da (or 2,589 per a single glycosylation site), while this number is significantly lower for Hp 2-1 (8,271 per subunit, or 2,068 per a single glycosylation site). Importantly, the latter estimate is based on the masses of the lowest-molecular weight Hp 2-1 species; however, application of limited charge reduction to more complex Hp 2-1 species (generating abundant ionic signal above *m/z* 6,000 in **Figure 1**) leads to similar estimates. Indeed, limited charge reduction applied to the second distinct set of ion peaks in the mass spectrum of Hp 2-1 (*m/z* 6,000 - 6,800, presumably corresponding to the HL-(HL*)-LH species) yields 139.8 kDa as a mass of the most abundant species and 137.6-155.1 kDa as a mass range (**Figure 3A**). The other higher-molecular weight population in the Hp 2-1 mass spectrum (m/z range 6,800-7,600, presumably corresponding to the HL-$(HL^*)_2$-LH species) yields 190.4 kDa as a mass of the most abundant species and 188.3-208.1 kDa as a mass range (**Figure 3B**). These experimentally measured masses provide 35,100±100 Da as an estimate of the H-chain average mass (which corresponds to the total glycan mass of 7,830 Da within each chain, or 1,958 Da per a single glycosylation site).

The ability to detect the differences in the extent of Hp glycosylation using limited charge reduction measurements (despite the fact that the resolution afforded by MS1 measurements is not sufficient for detection of distinct glycoforms) is remarkable. However, basing such estimates on the average masses alone may lead to erroneous conclusions should a significant number of other PTMs be present within the protein. In order to verify the conclusions based on the limited charge reduction measurements, we sought to obtain mass profiles of Hp subunits representing both Hp 1-1 and Hp 2-1 samples using an orthogonal

approach. Unfortunately, H-chains are unstable and prone to precipitation following disulfide reduction, which rules out the use of conventional methods based on disulfide reduction followed by capping the free thiol groups with alkylating reagents and MS analysis. This necessitated the use of a recently introduced technique, cross-path reactive chromatography (XP-RC) with on-line MS detection[16] as a means of obtaining mass profiles of all subunits of both phenotypes of Hp studied in this work. XP-RC is a platform that combines in-line chemical transformation (*e.g.*, disulfide reduction) with separation of the reaction products and unconsumed reagents/reaction byproducts prior to MS detection, which allows even metastable species to be characterized without sacrificing the quality of MS data.[16] Previously we have demonstrated that this technique can be used to reduce Hp 1-1 molecules inside the size exclusion column, followed by separation of the L- and H-chains.[16] In this work we used XP-RC/MS for in-line reduction and subunit characterization of both Hp 1-1 and Hp 2-1.

Consistent with the previous findings, we have not observed any detectable mass-changing PTMs within the L-chains of Hp 1-1 (data not shown). Furthermore, no detectable PTMs were present within the L- and L*-chains of Hp 2-1 (**Figure 4**). Even though the chromatographic resolution was insufficient to separate the two types of light chains, on-line MS detection allowed them to be readily distinguished from one another based on the mass difference (9,192 Da and 15,946 Da for the L- and L*-chains, respectively). In addition to abundant ions representing L and L* monomers, two series of low-intensity peaks were detected, which represented L-L* and L*-L* dimers, the apparent products of incomplete reduction of external disulfide bonds (**Figure 4**). It is worth mentioning that the mass measurements of the monomeric L-chains are consistent with the notion of the internal disulfide bonds being fully reduced, while the internal disulfide reduction within the L*-chains may be incomplete (measured mass of L* monomers is 2 Da below that of an average mass of the polypeptide calculated on the assumption that all cysteine side chains are fully reduced). Apart from that, there are no indications that any PTMs are present within the light chains of Hp 2-1 (our particular concern was the possibility of glycation, a PTM that can make a significant impact on the average mass of the protein, *vide supra*).

On-line MS analysis of the H-chains produced by XP-RC of Hp 1-1 is consistent with the previously published data,[16] indicating the presence of mature bi-antennary glycans of the complex type that are fully sialylated and extensively fucosylated (**Figure 5**). Interestingly, the most abundant ions correspond to species that have a tri-antennary structure at one (out of four) glycosylation sites; another abundant set of peaks (*m/z* 2910 - 2960 in **Figure 5**) represents H-chains with two glycans having a tri-antennary structure. All of these species are also fully sialylated (*i.e.*, they bear nine and ten sialic acid residues, respectively) and exhibit

even higher levels of fucosylation compared to the base glycoform (*i.e*., while the species carrying two and three fucose residues remain the most abundant, the presence of the fourth fucose residue – absent from the lower molecular weight glycoforms - becomes apparent). The average mass of H-chains derived from Hp 1-1 (calculated based on MS presented in **Figure 5** and taking into account uneven abundance distribution across all detected glycoforms) is 37,565 Da, and the mass range is 36.5-38.4 kDa. This allows the expected average mass of an intact Hp 1-1 molecule to be calculated as 93,514 Da (and the full mass range including all detectable glycoforms as 91,290 – 95,224 Da). This is in a good agreement with the results of limited charge reduction measurements of the intact Hp 1-1 molecules (average mass 93,620±130 Da).

A similar analysis was carried to determine the glycosylation patterns of the Hp 2-1 H-chains, yielding notably different results (**Figure 6**). A significant proportion of the signal corresponds to protein species with at least one out of four glycans being an incomplete bi-antennary structure (lacking a Gal-GlcNAc unit and exhibiting incomplete sialylation). The extent of fucosylation also appears to be significantly lower compared to the glycans of the H-chains produced by in-line reduction of Hp 1-1, with the most abundant species corresponding to a-fucosylated species (see the inset in **Figure 6**). The average mass of the Hp 2-1 H-chain is 36,510 Da (mass range 35.1 – 37.8 kDa), which corresponds to an average mass of the lowest-molecular weight intact species (having H-L-L-H architecture) being 91,404 Da and the mass range 88.7 - 93.9 kDa, which agrees reasonably well with the results of the limited charge reduction measurements (which defined the mass range as 85.8 – 96.0 kDa). Extrapolating the XP-RC/MS data to Hp 2-1 species with more complex architecture provides a 139.8-147.6 kDa mass range for the HL-(HL*)-LH species, while the results of limited charge reduction measurements suggest that the abundant species fall within the 138-144 kDa range, with several lower-abundance species found at masses as high as 155.1 kDa (*vide supra*).

Remarkably, the mass difference between the average glycan loads of the H-chains in Hp 1-1 and Hp 2-1 deduced from the limited charge reduction measurements (2.1 kDa) is nearly identical to that determined by XP-RC/MS measurements (1.9 kDa). At the same time, the mass range populated by the Hp 2-1 HL-LH species revealed by the limited charge reduction measurements partially overlaps with that of Hp 1-1, which is also consistent with the results of the XP-RC/MS measurements, which show that several glycoforms present in the higher-mass range of the Hp 2-1 distribution (**Figure 6**) are also present in the lower- and middle-mass ranges of the Hp 1-1 distribution (**Figure 5**). This indicates that limited charge reduction is not only straightforward, but also reliable way of comparing structurally heterogeneous biopolymers from different sources. As such, it might become an indispensable tool in the

analytical arsenal of comparability studies,[36] where the initial assessment of the (dis)similarity of two protein-based therapeutics can be carried out by intact-mass analysis, a task that can become extremely challenging if the proteins exhibit a high degree of structural heterogeneity. Furthermore, it will likely prove useful in certain types of clinical analyses where the ability to detect significant changes in the extent of protein glycosylation has significant diagnostic value.[27-29]

Conclusions

Intact mass analysis continues to enjoy a steady growth in popularity as a means of fast characterization of proteins, particularly in the field of biotechnology. However, the ever increasing complexity of protein-based therapeutics increasingly gives rise to macromolecular systems that exhibit extreme degrees of structural heterogeneity. This certainly presents a significant challenge to the intact-mass analysis, as the very notion of the "protein mass" loses its intuitive meaning, and a single number must be substituted with a convoluted distribution of masses. MS addresses this challenge in a variety of ways, and combination of ESI with gas-phase ion chemistry (limited charge reduction) recently become one of the robust methods of extracting meaningful information from convoluted and poorly resolved ionic signals in ESI mass spectra of highly heterogeneous biopolymers. Cross-validation of this method by the recently introduced XP-RC/MS demonstrated in this work indicates that limited charge reduction can be used as a means of analysis of highly glycosylated proteins and applied for tasks where fast but nonetheless reliable evaluation of the extent of glycosylation is required (*e.g.*, comparability studies and/or clinical analyses). Further improvements of this methodology are possible, *e.g.* by integrating this approach with front-end separations (as has bene recently demonstrated in the analysis of PEGylated protein therapeutics[37]), as well as by complementing limited charge reduction in the gas phase with solution-phase charge manipulation techniques.[38]

Supporting Information

Supporting information contains amino acid sequences of the L-, L*-, and H-chains of human Hp and schematic representation of the quaternary organization of the Hp 1-1 and Hp 2-1 isoforms.

Author Information

*Corresponding author (email to kaltashov@chem.umass.edu).

¶Current address: Analytical development, 360 Binney St, Cambridge, MA 02142.

§Current address: Department of Chemistry, Purdue University, 560 Oval Drive, West Lafayette, IN 47907.

Authors' contributions: I.K., J.P. and Y.Y. designed the study; J.P., I.C. and Y.Y. planned the experimental work and acquired the data; J.P., I.C., Y.Y. and I.K. interpreted the data; I.K. wrote the manuscript. All authors have given approval to the final version of the manuscript.

Acknowledgements

This work was supported by a grant R01 GM132673 from the National Institutes of Health. All measurements were carried out at the Mass Spectrometry Core Facility at UMass-Amherst.

Figure legends

Figure 1. ESI mass spectra of aqueous solutions of Hp 1-1 (0.43 mg/mL) and Hp 2-1 (1.0 mg/mL) in 150 mM ammonium acetate, pH 6.9 shown with a gray-shaded curve and a black trace, respectively. Charge state assignments are made based on the results of limited charge reduction measurements (refer to Figures 2 and 3). The inset shows the mass spectra of the two proteins acquired under denaturing conditions (water/acetonitrile/acetic acid, 50/49.9/0.1 by volume).

Figure 2. Limited charge reduction of ESI-generated ions of Hp 1-1 (**A**) and the lowest-molecular weight species of Hp 2-1 (**B**). The initial precursor ion selection was made in each case at the apex of one of the most abundant peaks in the distribution, and all precursor ion selection windows were spaced by 100 *m/z* unit increments.

Figure 3. Limited charge reduction of ESI-generated ions of Hp 2-1 species having HL-(HL*)-LH (**A**) and HL-$(HL^*)_2$-LH architecture (**B**). The initial precursor ion selection in each case was made at the apex of the most abundant peak in the distribution, and all precursor ion selection windows were spaced by 100 *m/z* unit increments.

Figure 4. The on-line mass spectrum averaged across an 11-13 min time window of the XP-RC/MS chromatogram of Hp 2-1 corresponding to elution of the light chains (L and L*, labeled with dark blue and teal ovals, respectively). The signals of the L*L* and LL* dimers arise due to the incomplete reduction of disulfide bonds.

Figure 5. The on-line mass spectrum averaged across a 10.5-11.5 min time window of the XP-RC/MS chromatogram of Hp 1-1 corresponding to elution of the H-chain. The inset shows a zoomed view of the ionic signal at charge state +13. The base peaks in the distribution shown in the inset (identified with the dotted lines) correspond to a-fucosylated glycoforms (the schematic structures of the glycans are shown at the top of the spectrum; the detailed explanation of the structures is provided at the bottom of the figure). The red triangles identify fucosylated glycoforms (with the number of triangles indicating the extent of fucosylation). The blue/yellow arrow identifies glycoforms related to each other by addition of a Gal-GlcNAc disaccharide unit.

Figure 6. The on-line mass spectrum averaged across a 10.5-11.5 min time window of the XP-RC/MS chromatogram of Hp 2-1 corresponding to elution of the H-chain. The inset shows a zoomed view of the ionic signal at charge state +11. The base peaks in the distribution shown in the inset (identified with the dotted lines) correspond to a-fucosylated glycoforms (the schematic structures of the glycans are shown at the top of the spectrum; the detailed explanation of the structures is provided at the bottom of the figure). The red triangles identify glycoforms that are related to the base peaks by addition of fucose residues (with the number of triangles indicating the extent of fucosylation). The dotted purple arrows identify glycoforms related to the base peaks by loss of sialic acid residues; fucosylation of such species is indicated with red arrows.

References

(1) Carini, M.; Regazzoni, L.; Aldini, G. Mass Spectrometric Strategies and Their Applications for Molecular Mass Determination of Recombinant Therapeutic Proteins. *Curr. Pharm. Biotechnol.* **2011**, *12*, 1548-1557.

(2) Tassi, M.; De Vos, J.; Chatterjee, S.; Sobott, F.; Bones, J.; Eeltink, S. Advances in native high-performance liquid chromatography and intact mass spectrometry for the characterization of biopharmaceutical products. *J. Sep. Sci.* **2018**, *41*, 125-144.

(3) Li, W.; Yang, B.; Zhou, D.; Xu, J.; Ke, Z.; Suen, W. C. Discovery and characterization of antibody variants using mass spectrometry-based comparative analysis for biosimilar candidates of monoclonal antibody drugs. *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* **2016**, *1025*, 57-67.

(4) Lanter, C.; Lev, M.; Cao, L.; Loladze, V. Rapid Intact mass based multi-attribute method in support of mAb upstream process development. *J. Biotechnol.* **2020**, *314-315*, 63-70.

(5) van der Burgt, Y. E. M.; Cobbaert, C. M. Proteoform Analysis to Fulfill Unmet Clinical Needs and Reach Global Standardization of Protein Measurands in Clinical Chemistry Proteomics. *Clin. Lab. Med.* **2018**, *38*, 487-497.

(6) Dotz, V.; Wuhrer, M. N-glycome signatures in human plasma: associations with physiology and major diseases. *FEBS Lett.* **2019**, *593*, 2966-2976.

(7) Thomas, D.; Rathinavel, A. K.; Radhakrishnan, P. Altered glycosylation in cancer: A promising target for biomarkers and therapeutics. *Biochim. Biophys. Acta Rev. Cancer* **2020**, *1875*, 188464.

(8) Verhelst, X.; Dias, A. M.; Colombel, J. F.; Vermeire, S.; Van Vlierberghe, H.; Callewaert, N.; Pinho, S. S. Protein Glycosylation as a Diagnostic and Prognostic Marker of Chronic Inflammatory Gastrointestinal and Liver Diseases. *Gastroenterology* **2020**, *158*, 95-110.

(9) Zaia, J. Mass spectrometry and glycomics. *OMICS* **2010**, *14*, 401-418.

(10) Pawlowski, J. W.; Bajardi-Taccioli, A.; Houde, D.; Feschenko, M.; Carlage, T.; Kaltashov, I. A. Influence of glycan modification on IgG1 biochemical and biophysical properties. *J. Pharm. Biomed. Anal.* **2018**, *151*, 133-144.

(11) Camperi, J.; Pichon, V.; Delaunay, N. Separation methods hyphenated to mass spectrometry for the characterization of the protein glycosylation at the intact level. *J. Pharm. Biomed. Anal.* **2020**, *178*, 112921.

(12) Domínguez-Vega, E.; Tengattini, S.; Peintner, C.; van Angeren, J.; Temporini, C.; Haselberg, R.; Massolini, G.; Somsen, G. W. High-resolution glycoform profiling of intact therapeutic proteins by hydrophilic interaction chromatography-mass spectrometry. *Talanta* **2018**, *184*, 375-381.

(13) Camperi, J.; Combès, A.; Fournier, T.; Pichon, V.; Delaunay, N. Analysis of the human

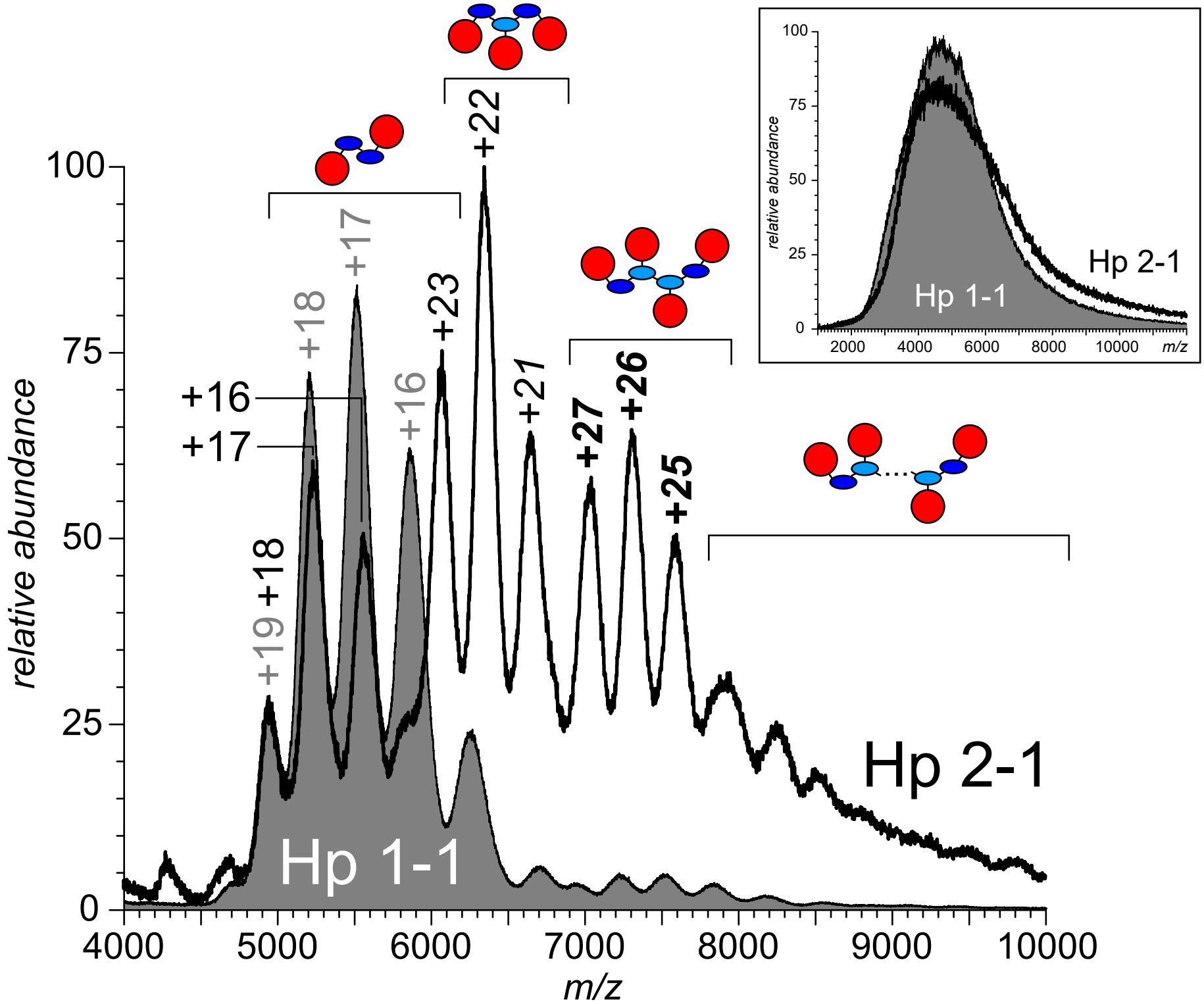
chorionic gonadotropin protein at the intact level by HILIC-MS and comparison with RPLC-MS. *Anal. Bioanal. Chem.* **2020**, *412*, 4423-4432.
(14) Wang, S.; Liu, A. P.; Yan, Y.; Daly, T. J.; Li, N. Characterization of product-related low molecular weight impurities in therapeutic monoclonal antibodies using hydrophilic interaction chromatography coupled with mass spectrometry. *J. Pharm. Biomed. Anal.* **2018**, *154*, 468-475.
(15) Abzalimov, R. R.; Kaltashov, I. A. Electrospray ionization mass spectrometry of highly heterogeneous protein systems: Protein ion charge state assignment via incomplete charge reduction. *Anal. Chem.* **2010**, *82*, 7523-7526.
(16) Pawlowski, J. W.; Carrick, I.; Kaltashov, I. A. Integration of On-Column Chemical Reactions in Protein Characterization by Liquid Chromatography/Mass Spectrometry: Cross-Path Reactive Chromatography. *Anal. Chem.* **2018**, *90*, 1348-1355.
(17) Andersen, C. B.; Torvund-Jensen, M.; Nielsen, M. J.; de Oliveira, C. L.; Hersleth, H. P.; Andersen, N. H.; Pedersen, J. S.; Andersen, G. R.; Moestrup, S. K. Structure of the haptoglobin-haemoglobin complex. *Nature* **2012**, *489*, 456-459.
(18) Graves, K. L.; Vigerust, D. J. Hp: an inflammatory indicator in cardiovascular disease. *Future Cardiol.* **2016**, *12*, 471-481.
(19) Arredouani, M. S.; Kasran, A.; Vanoirbeek, J. A.; Berger, F. G.; Baumann, H.; Ceuppens, J. L. Haptoglobin dampens endotoxin-induced inflammatory effects both in vitro and in vivo. *Immunology* **2005**, *114*, 263-271.
(20) Weatherhead, J. E.; Clark, E.; Vogel, T. P.; Atmar, R. L.; Kulkarni, P. A. Inflammatory syndromes associated with SARS-CoV-2 infection: dysregulation of the immune response across the age spectrum. *J. Clin. Invest.* **2020**, *130*, 6194-6197.
(21) Cavezzi, A.; Troiani, E.; Corrao, S. COVID-19: hemoglobin, iron, and hypoxia beyond inflammation. A narrative review. *Clin. Pract.* **2020**, *10*, 1271.
(22) Van Gucht, S.; Atanasova, K.; Barbé, F.; Cox, E.; Pensaert, M.; Van Reeth, K. Effect of porcine respiratory coronavirus infection on lipopolysaccharide recognition proteins and haptoglobin levels in the lungs. *Microbes Infect.* **2006**, *8*, 1492-1501.
(23) Wan, J.; Sun, W.; Li, X.; Ying, W.; Dai, J.; Kuai, X.; Wei, H.; Gao, X.; Zhu, Y.; Jiang, Y.; Qian, X.; He, F. Inflammation inhibitors were remarkably up-regulated in plasma of severe acute respiratory syndrome patients at progressive phase. *Proteomics* **2006**, *6*, 2886-2894.
(24) Janz, D. R.; Bastarache, J. A.; Sills, G.; Wickersham, N.; May, A. K.; Bernard, G. R.; Ware, L. B. Association between haptoglobin, hemopexin and mortality in adults with sepsis. *Crit. Care* **2013**, *17*, R272.
(25) Kerchberger, V. E.; Bastarache, J. A.; Shaver, C. M.; Nagata, H.; McNeil, J. B.; Landstreet, S. R.; Putz, N. D.; Yu, W. K.; Jesse, J.; Wickersham, N. E.; Sidorova, T. N.; Janz, D. R.; Parikh,

C. R.; Siew, E. D.; Ware, L. B. Haptoglobin-2 variant increases susceptibility to acute respiratory distress syndrome during sepsis. *JCI Insight* **2019**, *4*.
(26) Messner, C. B.; Demichev, V.; Wendisch, D.; Michalick, L.; White, M.; Freiwald, A.; Textoris-Taube, K.; Vernardis, S. I.; Egger, A. S.; Kreidl, M.; Ludwig, D.; Kilian, C.; Agostini, F.; Zelezniak, A.; Thibeault, C.; Pfeiffer, M.; Hippenstiel, S.; Hocke, A.; von Kalle, C.; Campbell, A., et al. Ultra-High-Throughput Clinical Proteomics Reveals Classifiers of COVID-19 Infection. *Cell Syst.* **2020**, *11*, 11-24.e14.
(27) Maresca, B.; Cigliano, L.; Spagnuolo, M. S.; Dal Piaz, F.; Corsaro, M. M.; Balato, N.; Nino, M.; Balato, A.; Ayala, F.; Abrescia, P. Differences between the glycosylation patterns of haptoglobin isolated from skin scales and plasma of psoriatic patients. *PloS one* **2012**, *7*, e52040.
(28) Kim, J. H.; Lee, S. H.; Choi, S.; Kim, U.; Yeo, I. S.; Kim, S. H.; Oh, M. J.; Moon, H.; Lee, J.; Jeong, S.; Choi, M. G.; Lee, J. H.; Sohn, T. S.; Bae, J. M.; Kim, S.; Min, Y. W.; Lee, H.; Lee, J. H.; Rhee, P. L.; Kim, J. J., et al. Direct analysis of aberrant glycosylation on haptoglobin in patients with gastric cancer. *Oncotarget* **2017**, *8*, 11094-11104.
(29) Kratz, E. M.; Waszkiewicz, N.; Kaluza, A.; Szajda, S. D.; Zalewska-Szajda, B.; Szulc, A.; Zwierz, K.; Ferens-Sieczkowska, M. Glycosylation changes in the salivary glycoproteins of alcohol-dependent patients: a pilot study. *Alcohol Alcohol.* **2014**, *49*, 23-30.
(30) Lee, S. H.; Jeong, S.; Lee, J.; Yeo, I. S.; Oh, M. J.; Kim, U.; Kim, S.; Kim, S. H.; Park, S. Y.; Kim, J. H.; Park, S. H.; Kim, J. H.; An, H. J. Glycomic profiling of targeted serum haptoglobin for gastric cancer using nano LC/MS and LC/MS/MS. *Mol. Biosyst.* **2016**, *12*, 3611-3621.
(31) Andersen, C. B. F.; Stødkilde, K.; Sæderup, K. L.; Kuhlee, A.; Raunser, S.; Graversen, J. H.; Moestrup, S. K. Haptoglobin. *Antioxid. Redox Signal.* **2017**, *26*, 814-831.
(32) Fries, E.; Wicher, K. B. Evolutionary aspects of hemoglobin scavengers. *Antioxid. Redox Signal.* **2010**, *12*, 249-259.
(33) Griffith, W. P.; Kaltashov, I. A. Mass spectrometry in the study of hemoglobin: from covalent structure to higher order assembly. *Curr. Org. Chem.* **2006**, *10*, 535-553.
(34) Griffith, W. P.; Mohimen, A.; Abzalimov, R. R.; Kaltashov, I. A. Characterization of Protein Higher Order Structure and Dynamics with ESI MS. In: *Comprehensive Analytical Chemistry, Vol. 52: Protein Mass Spectrometry*, Whitelegge, J. P., Ed.; Elsevier: San Diego, CA, **2008**, pp. 47-62.
(35) Yang, W.; Tu, Z.; Wang, H.; Zhang, L.; Kaltashov, I. A.; Zhao, Y.; Niu, C.; Yao, H.; Ye, W. The mechanism of reduced IgG/IgE-binding of beta-lactoglobulin by pulsed electric field pretreatment combined with glycation revealed by ECD/FTICR-MS. *Food & function* **2018**, *9*, 417-425.
(36) Ambrogelly, A.; Gozo, S.; Katiyar, A.; Dellatore, S.; Kune, Y.; Bhat, R.; Sun, J.; Li, N.;

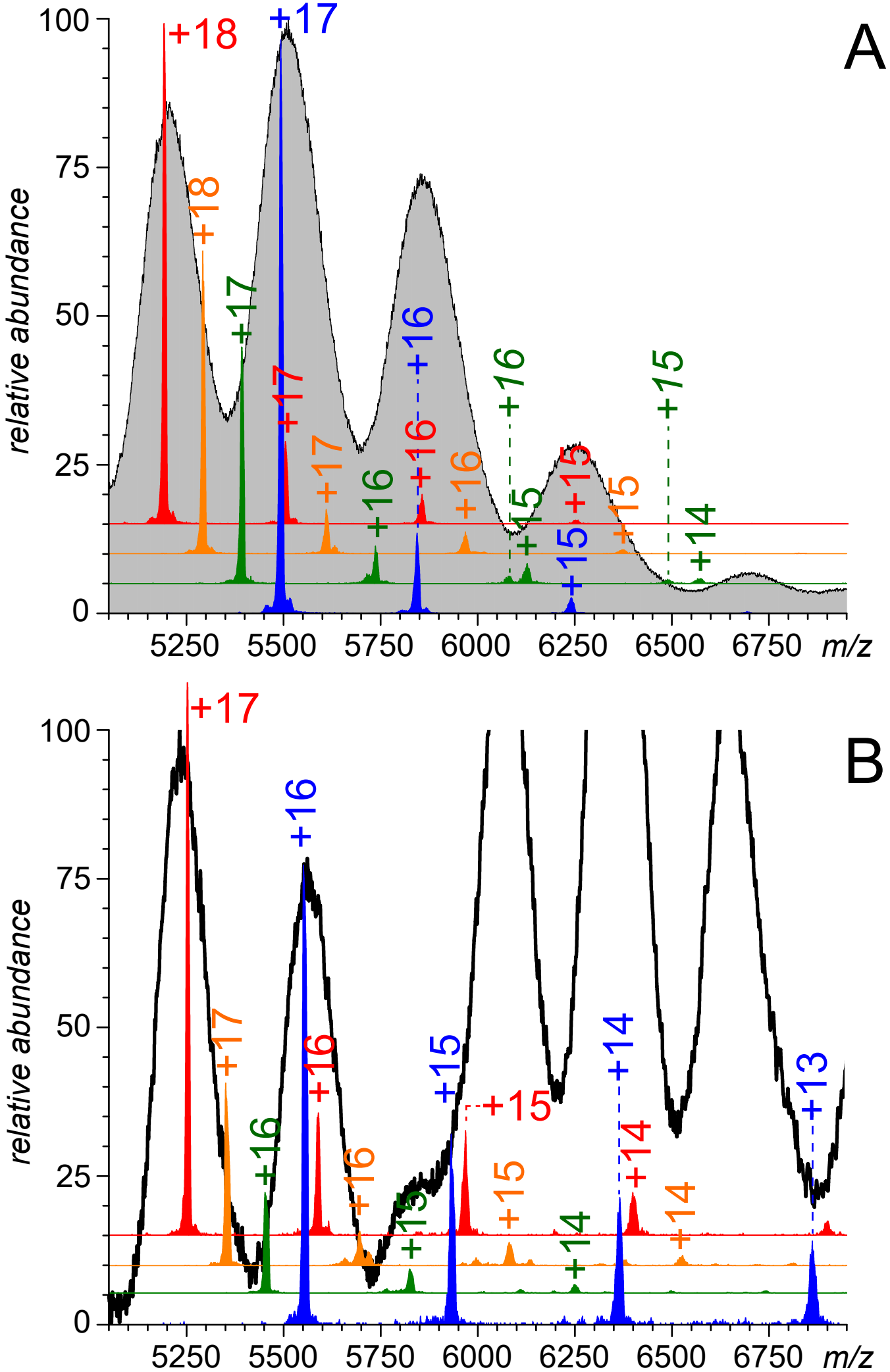
Wang, D.; Nowak, C.; Neill, A.; Ponniah, G.; King, C.; Mason, B.; Beck, A.; Liu, H. Analytical comparability study of recombinant monoclonal antibody therapeutics. *MAbs* **2018**, *10*, 513-538.

(37) Wang, S.; Xing, T.; Liu, A. P.; He, Z.; Yan, Y.; Daly, T. J.; Li, N. Simple Approach for Improved LC–MS Analysis of Protein Biopharmaceuticals via Modification of Desolvation Gas. *Anal. Chem.* **2019**, *91*, 3156-3162.

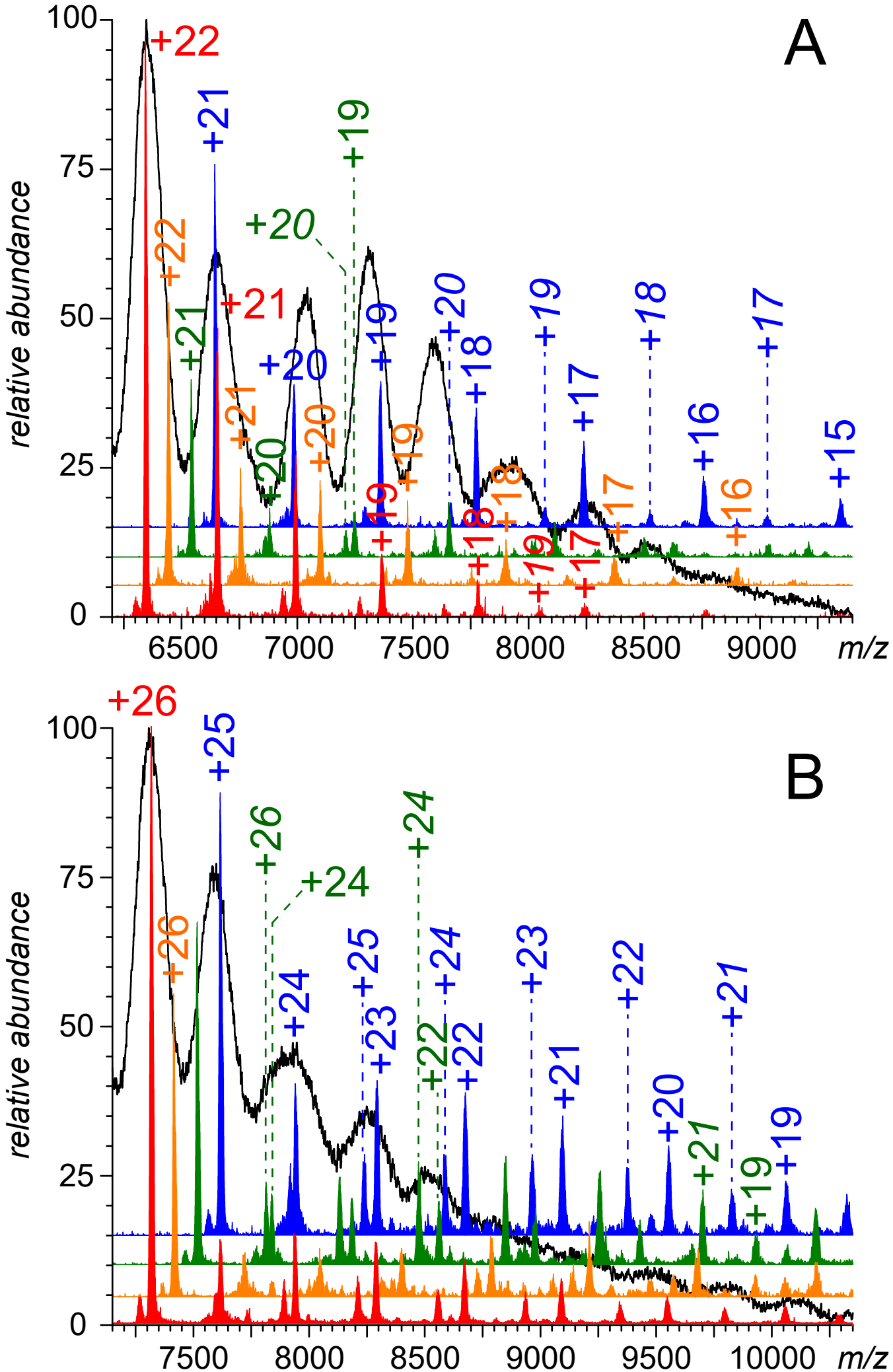
(38) Yang, Y.; Chendi, N.; Cedric, B.; Igor, K. Ionic Charge Manipulation Using Solution and Gas-Phase Chemistry to Facilitate Analysis of Highly Heterogeneous Protein Complexes in Native Mass Spectrometry. *ChemRxiv* **2020**, *doi: 10.26434/chemrxiv.13379975.v1*.



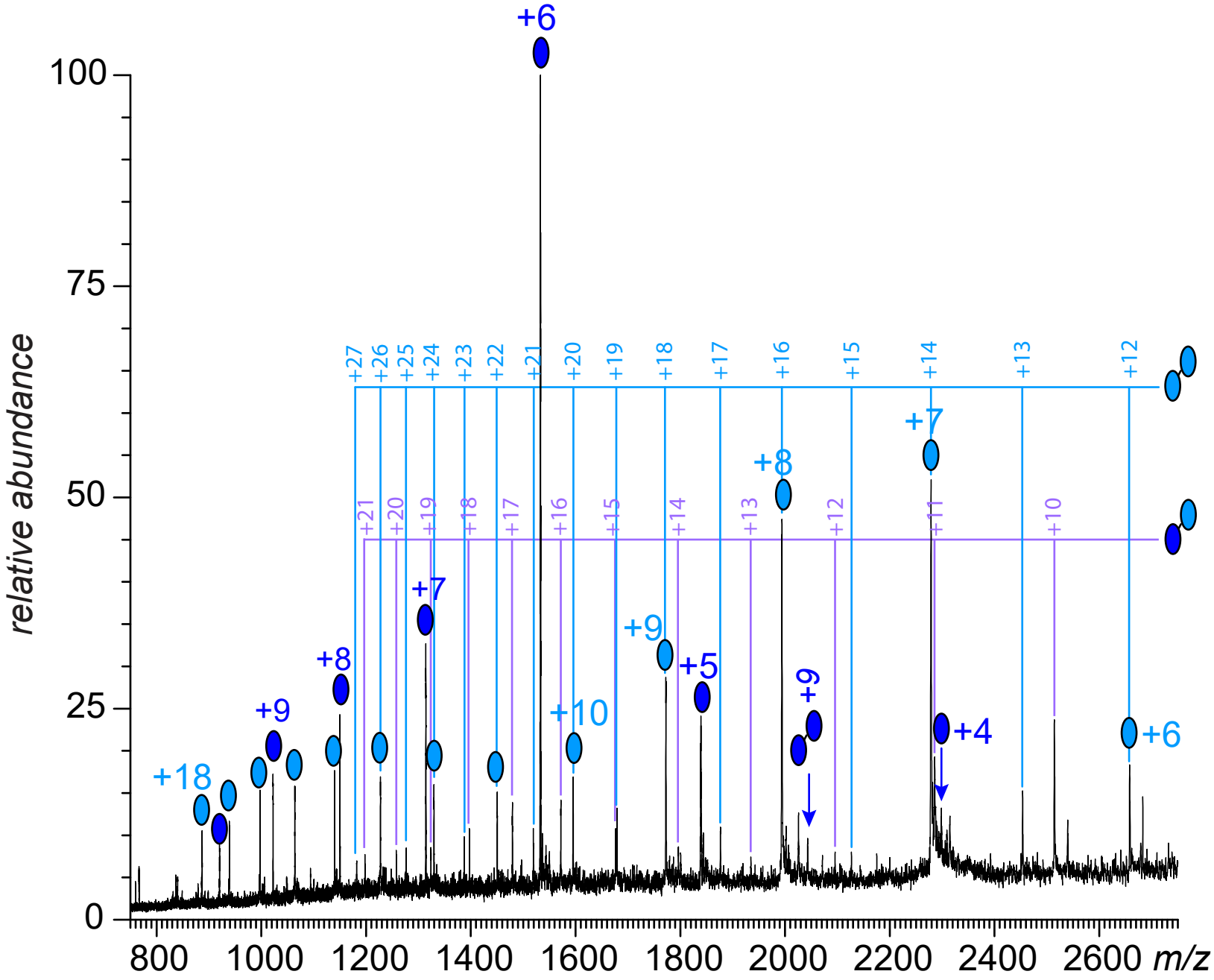
Yang, *et al*. **Figure 1**



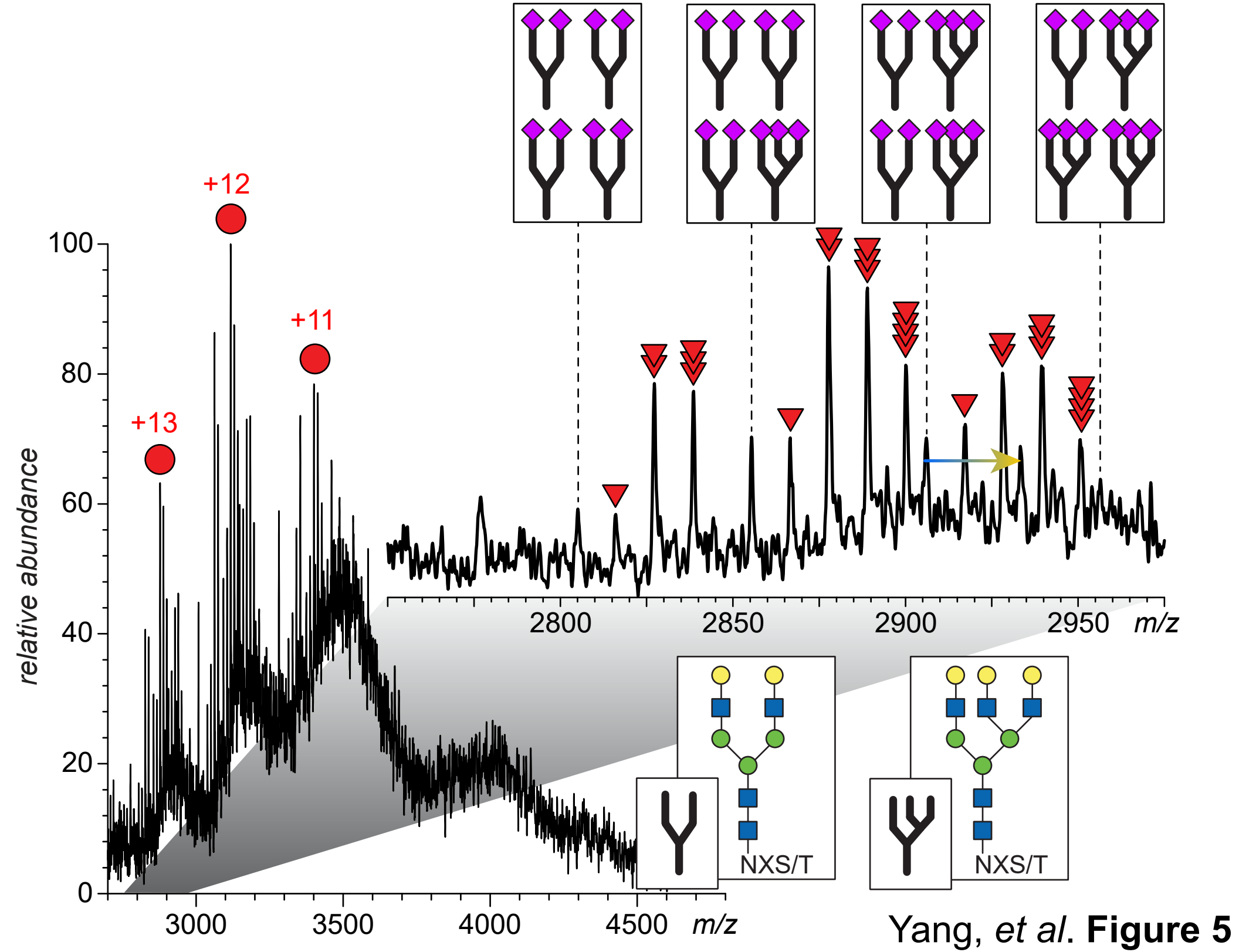
Yang, *et al.* **Figure 2**



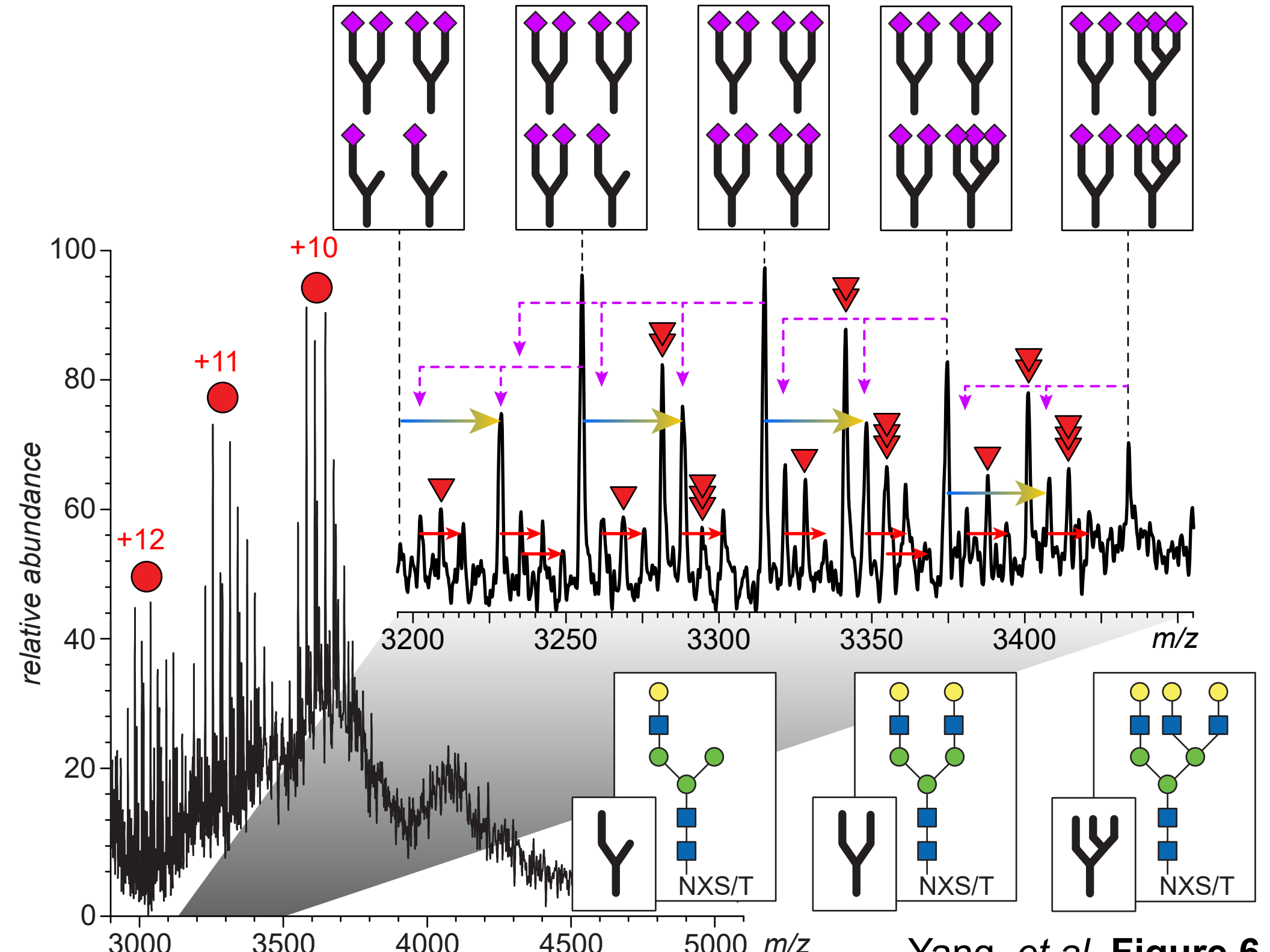
Yang, *et al.* **Figure 3**



Yang, *et al*. **Figure 4**



Yang, *et al*. **Figure 5**



Yang, *et al*. **Figure 6**

Supplementary Information for

Evaluation of the extent of haptoglobin glycosylation using two orthogonal intact-molecule MS approaches

Yang Yang, Jake W. Pawlowski, Ian Carrick, and Igor A. Kaltashov

Department of Chemistry, University of Massachusetts-Amherst, 240 Thatcher Road, Amherst, MA 01003

Contents

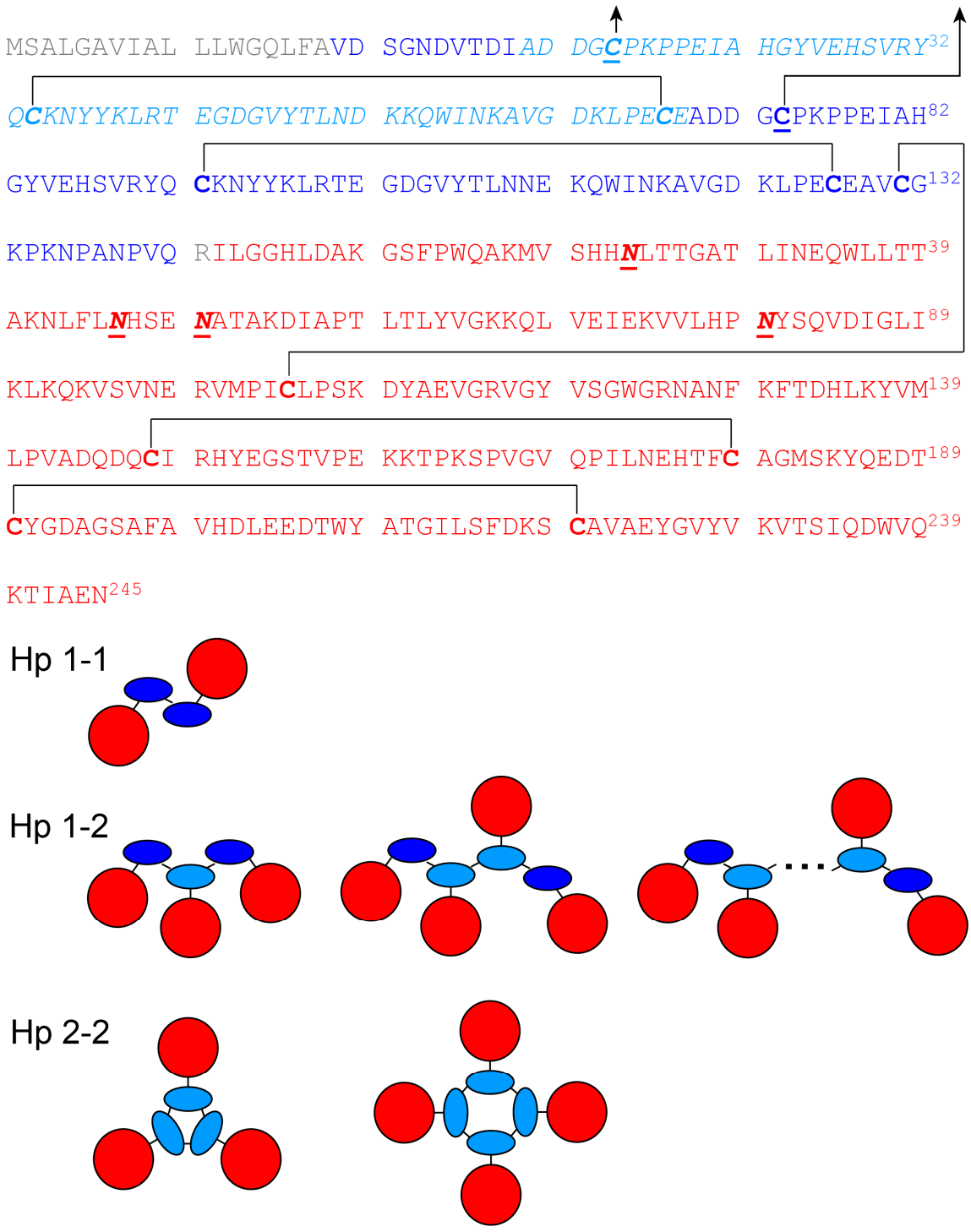


Figure S1. Top: the amino acid sequence of human haptoglobin pre-protein (P00738). The sequences of the light (L) and heavy (H) chains are highlighted in blue and red, respectively. The segment highlighted in purple corresponds to the additional insert present in the light chain L*. The black brackets show the disulfide bonds; arrows indicate the external disulfide bonds (L-L, L-L* and L*-L*). Bottom: schematic representations of the quaternary organization of haptoglobin oligomers.

Table S1. Relative intensities of the ionic species chosen as precursors for limited charge reduction measurements[1] and the abundance of corresponding charge-reduced species.[2]

	Precursor	Charge reduced group	Mass (Da)
Hp 1-1, H-L-L-H	m/z 5193, +18 (88.65%)	+17 (14.58%), +16 (5.23%), +15 (0.64%)	93, 620±130
	m/z 5293, +18 (51.41%)	+17 (7.52%), +16 (3.61%), +15 (0.70%)	95, 415±120
	m/z 5393, +17 (42.43%)	+17 (1.80%), +16 (1.36%), +15 (0.80%)	97, 193±110
		+16 (6.73%), +15 (3.54%), +14 (0.91%)	91, 809±110
	m/z 5493, +17 (100.00%)	+16 (14.01%), +15 (2.74%)	93, 494±100
Hp2-1, H-L-L-H	m/z 5253, +17 (100.00%)	+16 (22.14%), +15 (18.91%), +14 (7.7%)	89, 449±120
	m/z 5353, +17 (48.85%)	+17(2.09%), +16(2.04%), +15(1.46%)	96, 034±230
		+16 (9.28%), +15 (6.09%), +14 (2.78%)	91, 162±150
		+15(3.69%), +14(1.63%)	85, 758±130
	m/z 5453, +16 (41.54%)	+15 (9.89%), +14 (3.54%)	87, 359±100
	m/z 5553, +16 (85.24%)	+15 (35.22%), +14 (23.63%), +13 (15.44%)	89, 020±150
Hp2-1, HL-(HL*)-LH	m/z 6345, +22 (100.00%)	+23 (7.46%), +22 (4.93%), +21 (3.54%), +20 (2.30%), +19 (2.02%)	152, 573±140
		+21 (49.87%), +20 (27.91%), +19 (10.92%), +18 (6.36%), +17 (2.44%)	139, 784±150
	m/z 6445, +22 (53.93%)	+23 (5.95%), +22 (3.12%), +21 (2.95%), +20 (2.20%), +19 (1.97%), +18 (1.64%)	155, 128±70
		+21 (22.18%), +20 (19.81%), +19 (16.20%), +18 (8.68%), +17 (4.93%), +16 (3.37%)	141, 916±110

	m/z 6545, +21 (36.52%)	+21 (4.23%), +20 (4.70%), +19 (4.43%), +18 (3.61%), +17 (2.95%)	144, 225±50
		+20 (9.96%), +19 (9.47%), +18 (11.01%), +17 (7.25%), +16 (3.03%)	137, 621±160
	m/z 6645, +21 (67.74%)	+22 (3.74%), +21 (3.70%), +20 (3.56%), +19 (3.17%), +18 (2.65%), +17 (2.06%)	153, 287±140
		+20 (26.66%), +19 (27.14%), +18 (22.25%), +17 (15.98%), +16 (9.45%), +15 (5.29%)	139, 758±160
Hp2-1, HL-$(HL^*)_2$-LH	m/z 7316, +26 (100.00%)	+27 (4.90%), +26 (7.27%), +25 (7.92%), +24 (5.50%), +23 (5.21%), +22 (3.66%)	205, 311±110
		+25 (13.99%), +24 (14.51%), +23 (13.53%), +22 (10.64%), +21 (7.47%), +20 (4.49%)	190, 402±120
	m/z 7416, +26 (56.47%)	N/A, N/A, +24 (3.18%), +23 (3.32%), +22 (4.02%), +21 (3.03%)	208, 123±190
		N/A, N/A, +23 (3.28%), +22 (5.29%), +21 (6.13%), +20 (4.13%)	191, 910±240
		+25 (8.18%), +24 (8.14%), +23 (9.06%), +22 (10.62%), +21 (13.83%), +20 (8.06%)	193, 150±80
	m/z 7516, +25 (61.71%)	+26 (14.92%), +25 (15.05%), +24 (18.55%), +23 (19.58%), +22 (16.90%), +21 (13.44%)	203, 335±150
		+24 (12.81%), +23 (11.85%), +22 (11.27%), +21 (10.10%), +20 (7.99%), +19 (5.53%)	188, 322±160
	m/z 7616, +25 (77.77%)	+26 (12.64%), +25 (14.11%), +24 (13.96%), +23 (14.06%), +22 (11.94%), +21 (8.05%)	206, 080±110
		+24 (41.66%), +23 (27.20%), +22 (25.00%), +21 (20.81%), +20 (15.78%), +19 (9.55%)	190, 927±130

[1]The relative intensity of the most abundant precursor ion is set as 100% for each Hp oligomer.

[2] the listing order of different charge-reduced species in each group (top to bottom) follows the left-to-right order in the mass spectrum