Rapid evaluation of the extent of haptoglobin glycosylation using orthogonal intact-mass MS approaches and multivariate analysis

Daniil G. Ivanov,[a] Yang Yang,[a,i] Jake W. Pawlowski,[ii] Ian J. Carrick,[iii] and Igor A. Kaltashov*

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

Abstract

Intact-mass measurements are becoming 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 is 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 the entire complement of 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) despite a significant overlap of their ionic signals. More detailed characterization of their glycosylation patterns is enabled by the recently introduced technique of cross-path reactive chromatography (XP-RC) with on-line MS detection, which 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. N-glycosylation patterns of both haptoglobin phenotypes were found to be consistent with bi- and tri-antennary structures of complex type that exhibit significant level of fucosylation and sialylation. However, multivariate analysis of haptoglobin 1-1 reveals higher number of the tri-antennary structures, in comparison to haptoglobin 2-1, as well as a higher extent of fucosylation. The glycosylation patterns deduced from the XP-RC/MS measurements are in agreement with the conclusions of the intact-mass analysis supplemented by limited charge reduction, suggesting that the latter technique can be employed in situations when fast assessment of protein heterogeneity is needed (*e.g.*, process analytical technology applications).

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 undoubtedly improve the accuracy of the existing (non-MS based) diagnostic methods of biomarker measurements.[5] 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] While the de-novo characterization of protein glycoforms involves both localization of the glycosylation sites within the protein and determination of their structure,[7] 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.[8] 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),[9] although it can certainly benefit from incorporating an on-line separation step prior to MS analysis.[10] Hydrophilic interaction chromatography (HILIC) remains the most popular option for the glycoform separation prior to the MS step;[11,12] 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.[13]

In this work we use an alternative approach to evaluate structural heterogeneity of larger glycoproteins (90-190 kDa) using a combination of native MS supplemented with limited charge reduction[14] and a recently introduced cross-path reactive chromatography (XP-RC) platform with on-line MS detection.[15] The model used to test this approach is haptoglobin (Hp), an acute phase plasma glycoprotein, which captures free hemoglobin (Hb) in circulation during intravascular

hemolysis and transports it to macrophages.[16,17] 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 Hp 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 the acute respiratory distress syndrome during sepsis.[25] Hp expression levels have also been shown to correlate with the severity of COVID-19,[26] and notably lower levels of this protein were observed within the intubated and deceased groups of patients.[27] Lastly, Hp is likely to be an important factor modulating the severity of some rare adverse reactions to mRNA vaccines, where acute hemolysis is a prime manifestation.[28,29]

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[30] and cancer[31] to chronic alcoholism.[32] 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 that requires highly specialized MS equipment.[33] 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[34] (see ***Supplementary Material*** for more detail). In humans, Hp is expressed in two allelic forms, HP1 and HP2, which differ by the size of the light chain.[35] 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 the H-L-L-H connectivity.[16] 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%,[34] making this protein highly heterogeneous. In fact, even the 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;[36] the MS analysis of other phenotypes is more difficult still[37]). In this work we use limited charge reduction[14] 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,[15] a technique that allows ionic signal of multiple glycoforms to be resolved. 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 protein heterogeneity.

Experimental

Materials. Human Hp (phenotypes 1-1 and 2-1) were purchased from Athens Research & Technologies, Inc. (Athens, GA). More details on sample preparation are provided in the ***Supplementary Material*** section.

Methods. All native MS measurements were carried out with a Synapt G2 HDMS (Waters, Milford, MA) hybrid quadrupole/time-of-flight mass spectrometer equipped with a nanospray ion source. The following set of parameters was used in the ESI interface region: capillary voltage, 1.2 kV; sampling cone voltage, 40 V; extraction cone voltage 4 V. Isolation of ionic populations for limited charge reduction in the trap cell for limited charge reduction measurements was performed by setting the quadrupole LM resolution values 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 0.4-1.5 V; the discharge current was set at 20 μA. The MS data were analyzed by using MassLynx 4.2 (Waters, Milford, MA); deconvolution of the mass spectra was carried out

using a UniDEC algorithm.

In-line (on-column) chemical reactions were carried uisng a TSKgel G3000SWxl (Tosoh, Tokyo, Japan) SEC column and an HP 1100 (Agilent, Santa Clara, CA) HPLC system. A flow rate of 0.1 mL/min was typically used in XP-RC/MS measurements for the analysis of Hp glycosylation patterns. The reagent plug was composed of 50 mM TCEP and 4 M guanidinium chloride, and a 80:20 (v:v) mixture of 100 mM ammonium acetate (pH 6.8) and acetonitrile was used as a mobile phase. The reagent plug was introduced using a manual injector with a loop volume of 500 μL, which was placed between the sample injector and the SEC column. The protein samples were injected with a delay time of 30 sec with respect to the reagent plug injection. On-line MS measurements were performed using Synapt G2 HDMS with a LockSpray ESI source. The following parameters were use in the ESI interface: spray voltage 3 kV; sampling cone 100 V, extraction cone 10 V, source temperature 100 ^{O}C, desolvation temperature 300 ^{O}C, cone gas flowrate 100 L/h, desolvation gas flow rate 300 L/hr.

The data analysis was carried out using MassLynx followed by additional processing with the Anaconda distribution of Python 3. Mass convolutions were performed using an FFT based algorithm and SciPy library tools. Multivariate analysis was performed using the Scipy.stats module operating with data loaded into the data frames within the Pandas library.

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. The two mass spectra largely overlap, and neither contains significant discernable features that can be used to estimate the ionic charge states (inset in **Figure 1**). 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 clusters of peaks corresponding to specific oligomers (**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 almost overlap with the signal of Hp 1-1 (containing exclusively H-L-L-H proteoforms). Each distribution contains partially resolved peaks representing different charge states, suggesting that the average mass of the H-L-L-H proteoform can be determined in each case based on the native MS data once the charge states are assigned to all ionic peaks. However, direct identification of the charge states using the apexes of adjacent peaks does yields inconsistent results with the *z*

values showing (*i*) significant deviations from the nearest integers and (*ii*) inconsistent variations across the charge state distribution. This failure to assign charge states to the protein ion peaks is echoed by the ambiguity of the molecular mass distribution generated by robust deconvolution algorithms such as UniDEC[38] (the top inset in **Figure 1**), indicating the need for an orthogonal method of charge state assignment to be deployed.

Limited charge reduction allows the charge state of protein ions to be determined directly by generating well-defined charge ladders following selection of ionic populations within narrow m/z windows followed by their exposure to either electrons or anions.[14] Application of this technique to the ionic signal of Hp 1-1 and Hp 2-1 allows the most abundant peak in the lower *m/z* range (below 5,700) to assigned a charge state +19 (**Figure 2**). While this allows the mass of the most abundant protein ion mass to be calculated as 94.1 kDa for Hp 1-1 (based on the apexes of the ion peak, and the average mass of 93.9 kDa based on the peak centroid), the entire mass range populated by different proteoforms of Hp 1-1 can be readily determined as well by moving the ion isolation window across the protein peak. The resulting mass range is 90.8 - 96.0 kDa for Hp 1-1 (using the peak valleys positions as terminal points – see **Figure 2**). This is consistent with mass range obtained by fitting the experimentally determined peak shape of the +19 charge state with normal distribution, which yields the mass range of 90.3 - 97.3 (defined as points of 10% of the maximum intensity) with the 93.9 kDa average value **(*Supplementary Material*, Figure S6A)**. A similar analysis for the low-molecular weight species of Hp 2-1 (having the same molecular architecture as Hp 1-1, *i.e*. H-L-L-H) yields 92.9 kDa as the most abundant ion mass, and 91.9-95.6 kDa as the range populated by various proteoforms. Approximation of this peak shape with a normal distribution provided the mass range of 89.9 - 96.1 kDa (93.0 kDa average).

The noticeable difference between the masses of the protein species with identical molecular architecture present in the two Hp phenotypes may point 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,[39] 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 and using the amino acid sequences of the L-chains, the average masses of the H-chains can be estimated as 37.7 kDa (Hp 1-1) and 37.3 kDa (Hp 2-1). The polypeptide mass of the H-chain (calculated based on its amino acid sequence) is 27.3 kDa; therefore, the average carbohydrate component of this subunit is 10.5 kDa (or 2.6 kDa per a single glycosylation site) in Hp 1-1. This number is slightly lower for Hp 2-1 (10.0 kDa per subunit, or 2.5 kDa per a single glycosylation site).

The same analysis was performed to estimate the mass-distribution of the more complex Hp 2-1 species (see **Figure S2** in the ***Supplementary Material***). Limited charge reduction applied to the second distinct set of ion peaks in the mass spectrum of Hp 2-1 (*m/z* 5,700 - 6,500, presumably corresponding to the HL-(HL*)-LH species) yields 145.9 kDa as the mass of the most abundant species and 141.9-149.5 kDa as the mass range (10% of the maximum intensity). These experimentally measured masses provide 37.2 kDa as an estimate of the H-chain average mass, which corresponds to the total glycan mass of 9.9 kDa within each H-chain. The other higher-molecular weight population in the Hp 2-1 mass spectrum (m/z range 6.700-7.900, presumably corresponding to the larger species HL-$(HL^*)_2$-LH) yields 198.5 kDa as the most abundant species mass and 193.5-203.5 kDa as the mass range (see **Figure S2**, bottom in the ***Supplementary Material*** for more detail). These experimentally measured masses provide 37.0 Da as an estimate of the H-chain average mass, which corresponds to the total glycan mass of 9.8 kDa for each H-chain.

The ability to detect the differences in the extent of Hp glycosylation assisted by limited charge reduction measurements (despite the fact that the resolution afforded by MS1 measurements is clearly insufficient for detection of distinct glycoforms) could be exploited for comparative analysis between different proteoforms of the same protein even for highly-heterogenous biopolymers. 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 intact mass 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 become 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[15] 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 within protein molecules) with separation of the reaction products from the unconsumed reagents and/or reaction byproducts prior to MS detection, which allows even metastable species to be characterized without sacrificing the quality of MS data.[15] Previously we have demonstrated that this technique can be used to reduce Hp 1-1 molecules inside the size exclusion column, followed by separation and MS characterization of the L- and H-chains.[15] In this work we used XP-RC/MS for in-line reduction and subunit characterization of both Hp 1-1 and Hp 2-1.

Using XP-RC with online MS detection, no PTMs were observed within the L-chains of Hp 1-1 at the intact polypeptide level (**Figure S3**), which is consistent with the previous findings.[15] Furthermore, no detectable PTMs were present within the L- and L*-chains of Hp 2-1 (**Figure S4**). Even though the chromatographic resolution was insufficient to separate the two light chain types (L and L*), 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 S4**). 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 with 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*).

In-line reduction of both Hp phenotypes gives rise to abundant ionic signal of H-chains in XP-RC/MS spanning a range of charge states from +10 to +13 (**Figure 3**). A detailed analysis of these mass spectra reveals a presence of multiple peaks at each charge state which appear to represent a range of glycoforms based on characteristic spacing between adjacent peaks revealing the presence of deoxyhexoses, hexoses, acetyl hexose amines and acetylneuraminic acid residues. Furthermore, the ionic signal contains abundant components whose *m/z* values are consistent with the mass of the fully reduced H-chain (calculated based on its amino acid sequence) which incorporates four fully sialylated, fucose-free biantennary N-glycans of the complex type $Hex_5HexNAc_4NeuAc_2$ (calculated mass of the neutral species 36,084.4 Da vs. the measured mass of 36,086.3 Da). It might be tempting to assign the rest of the ionic peaks in each charge state cluster by relating their distances from the “base” peaks (labeled with arrows in **Figure 3**) to a specific set of saccharide residues added to (or removed from) the mature glycans $Hex_5HexNAc_4NeuAc_2$. For example, the only abundant peak on the left-hand side of the base peak in each charge state cluster corresponds to a mass loss of 291.2 Da and, therefore, can be readily assigned as a loss of a single sialic acid residue from one of the glycans (calculated mass loss 291.095 Da). However, applying this approach to the entire complement of ionic species detected in XP-RC/MS experiments is problematic due to ambiguities caused by multiple

combinations of monosaccharides giving rise to mass increments that are difficult or indeed impossible to distinguish from one another at the mass measurement accuracy levels that can be routinely achieved for intact proteins (5 ppm in the present work). This is illustrated in **Figure 4**, which shows calculated masses for all possible variations of the glycan structure within the four N-glycans present in the H-chain assuming they are all based on the complex-type template. The multiple overlaps are evident in this diagram, indicating that the naïve approach to the glycoform identification in the XP-RC/MS spectra of haptoglobin H-chain is both impractical and fraught with potential errors.

A straightforward analysis and comparison of the Hp glycosylation patterns on the basis of mass measurements alone is challenging due to the possible presence of multiple near-isobaric glycoforms. In order to circumvent this difficulty, we applied a multivariate statistical analysis approach, which allows the qualitative differences between the two patterns glycosylation patterns to be identified while tolerating possible errors in assigning the specific glycans.

A visual comparison of the ionic signals of the H-chains derived from two different isoforms in **Figure 4** suggest that the average mass is higher for the Hp 1-1 isoform. This could reflect one or more of the following: (a) a difference in the fractions of tri-antennary structures, (b) the difference in the extent of fucosylation, and (c) the difference in the sialylation levels. To evaluate contributions of each of these factors, all of them were examined independently.

The extent of fucosylation was evaluated by grouping all protein ion peaks whose masses fall within a 10 Da windows of masses of glycoforms from a set ($HexNAc_{16+t}Hex_{20+t}NeuAc_{8+t-n}$) representing various combinations of bi-antennary and tri-antennary structures (t = 0 - 4) with varying extent of sialylation (n = 0 - 6). The mass-distribution was obtained by converting the *m/z* values of H-chain ions at +12 charge state to neutral protein masses for both Hp 2-1 and Hp 1-1. A singly fucosylated satellite was identified for each of these glycoforms, followed by calculating the ratio of peak intensities for each pair. This array of numbers forms a smooth distribution for each isoform, allowing them to be compared (**Figure 5A**). This approach is more reliable than the direct comparison of the glycoform distributions, as any possible misidentification of peaks does not shift the entire distribution, but only introduces an outlier in the distribution. **Figure 5A** shows that the percentage of the fucose-free species is significantly lower for the Hp 1-1 isoform compared to Hp 2-1. The percentage of glycoforms carrying multiple fucose residues follows the same trend, which certainly contributes to the higher relative mass of intact Hp 1-1 species compared to the Hp 2-1 species with the same quaternary structure (H-L-L-H).

Evaluation of the differences between the two Hp isoforms in terms of the relative numbers of incorporated tri-antennary structures was focused on examining the peaks whose relative abundance are significantly different between Hp 2-1 and Hp 1-1 across all detected charge states from a statistical point of view. The mass spectra of the H-chains derived from both isoforms were deconvoluted spectra using ionic signals of charge states ranging from +13 to +10. Since all these charge states have no overlapped peaks, the H-chain mass distribution was obtained by converting the ionic *m/z* values to neutral masses using the corresponding ionic charges. The glycoform assignment in the H-chain mass distributions for both isoforms was followed by generation of a volcano plot to identify specific isoforms that are overrepresented in one haptoglobin isoform versus the other (**Figure 5B**). The normalized peak intensities averaged across all charge states were used to calculate the magnitude of change and the Student's *t*-statistics *p*-values was used as the "significance metrics" for "volcano-plot". All peaks were colored according to the number of tri-antennary chains presented in each specific glycoform. This analysis clearly shows that isoform 1-1 contains a larger fraction of tri-antennary structures compared to Hp 2-1.

Lastly, the two haptoglobin isoforms were compared vis-à-vis their sialylation levels. This was accomplished by plotting the intensity of each fucose-free isoform as a function of the number of incorporated sialic acid residues (**Figure S5**). No significant differences between Hp 2-1 and 1-1 were found; in fact, the "sialic acid saturation" profiles for glycoforms with zero, one and two tri-antennary chains were found to be almost identical for the H-chains derived from the Hp 2-1 and Hp 1-1 isoforms. This indicates that both isoforms exhibit identical sialylation levels, while the differences in the extent of fucosylation and the tri-/bi-antennary structures ratios are the chief contributors to the differences in the mass profiles of the H-chains derived from the two isoforms. The similar trend in differences in glycoform distribution was observed by Tamara et. al. who compared the glycosylation patterns of de-sialylated isoforms of Hp 1-1 and Hp 2-2 phenotypes.[40]

In order to evaluate the consistency of the results of XP-RC/MS measurements with the protein mass profiles extracted from the intact MS measurements, the mass distributions of the intact H-L-L-H molecules were calculated for each Hp phenotype by applying a conventional convolution algorithm[41] to the XP-RC/MS data. **Figure S6** in the ***Supplementary Material*** section shows the juxtaposition of the mass-distributions of the H-L-L-H species in both phenotypes deduced from the XP-RC/MS and native MS data. In both cases, the mass distributions derived from the XP-RC/MS appear to be slightly shifted towards lower masses, most likely due to less effective desolvation of ions during intact mass-measurements. Nevertheless, the magnitude of this shift

is not sufficient to mask the incongruence between the mass profiles of the H-L-L-H species from the two protein phenotypes. The mass distribution of Hp 1-1 extends to a notably higher mass range compared to the tetrameric Hp 2-1 species, reflecting the differences in their glycan compositions. A good agreement is also observed between XP-RC/MS and native MS for larger protein oligomers in Hp 2-1 (**Figure 5C** and **D)**. Therefore, native MS supplemented with limited charge reduction provides reliable estimates for the extent of protein glycosylation in two Hp phenotypes.

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 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 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 became 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, PAT applications and clinical analyses). In those settings the baseline information on protein glycosylation profiles is frequently available from prior bottom-up work, and the ability to observe the mass distribution shifts at the intact-molecule level would allow the relevant glycosylation changes to be readily interpreted without the need to carry additional lengthy and labor-intensive measurements. Further improvements of this methodology are possible, *e.g.* by integrating this approach with front-end separations (as has been recently demonstrated in the analysis of PEGylated protein therapeutics[42]), application of ionic suppressor to increase the desolvation efficiency,[43] as well as by complementing limited charge reduction in the gas phase with solution-phase charge manipulation techniques.[44]

Supporting Information

Supporting information contains (*i*) 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,

(*ii*) native MS/limited charge reduction analysis of higher-mass proteoforms of Hp 2-1, and (*iii*) mass information on all analyzed protein species in Hp 1-1 and Hp 2-1.

Author Information

[a] These two authors contributed equally to this work

[i] Current address: Analytical Development, Biogen, Cambridge, MA 02142

[ii] Current address: Analytical Development, Amgen, Cambridge, MA 02142

[iii] Current address: Department of Chemistry, Purdue University, West Lafayette, IN 47907

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

Authors' contributions: IAK designed the study, YY and DGI designed the experimental work; DGI, YY, JWP and IC carried out the experimental work; DGI designed the statistical algorithms for data analysis; DGI, YY and IAK interpreted the data, IAK and DGI wrote the manuscript. All authors read the final draft of the manuscript and gave their consent to submitting it for publication.

Acknowledgements

This work was supported by grants R01 GM132673 from the National Institutes of Health and CHE-1709552 from the National Science Foundation. All measurements were carried out at the Mass Spectrometry 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 100 mM ammonium acetate, pH 6.8 shown with a gray-filled curve and a black trace, respectively. Charge state assignments of resolved peaks were made based on the results of limited charge reduction measurements (refer to **Figures 2** and **S2**). The insets on the right show the results of UniDEC deconvolution[38] of the ionic signals in m/z range 4200-5700 (top) and ESI mass spectra acquired under denaturing conditions (bottom).

Figure 2. Limited charge reduction of ESI-generated ions of Hp 1-1 (**top panel**) and the lowest-molecular weight species of Hp 2-1 (**bottom panel**). The precursor ions were selected in each case at the apex of one of the most abundant peaks and at the lowest signal intensity points between two partially resolved peaks (isolation windows < 50 *m/z* units). The colored charge ladders represent limited charge reduction measurements for different precursor ions (with the signal intensity plotted on a log-scale).

Figure 3. Mass spectra of H-chains obtained by in-line reduction of Hp 1-1 (**top**) and Hp 2-1 (**bottom**) using XP-RC with on-line MS detection. Arrows indicate m/z values of ions representing the fully reduced H-chain carrying four fully sialylated, fucose-free biantennary N-glycans of the complex type (calculated mass of a neutral species 36,084.4 Da).

Figure 4. The comparison of mass-distribution of glycoforms between H-chains of Hp 1-1 and 2-1 species. On the top of diagram there is an identification of glycoforms that was done using combinatorial approach. Only non-fucosylated and single-fucosylated species are shown to avoid the overlapping of labels with species with different sialylation level.

Figure 5. The intensity ratio distributions of ions representing singly-fucosylated species and their non- fucosylated counterparts for all identified glycoforms, calculated for Hp 1-1 and Hp 2-1 heavy chains (**A**). The volcano-plot showing statistical significance (the *p*-value from Student's *t*-test) versus the magnitude of change (Hp 1-1 vs. Hp 2-1) for each peak. The peaks were colored according to the number of tri-antennary chains that occupy N-glycosylation sites. The *x*-axis values refer to the specific peak intensity ratios (H-chains of Hp 2-1 vs. Hp 1-1); *y*-axis is a negative logarithm of the *p*-value between. The glycoforms with fewer tri-antennary structures than bi-antennary ones are marked with red dots; others are marker with blue dots (**B**).

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

(8) Yang, W.; Tu, Z.; McClements, D. J.; Kaltashov, I. A. A systematic assessment of structural heterogeneity and IgG/IgE-binding of ovalbumin. *Food & function* **2021**, *12*, 8130-8140.

(9) 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.

(10) 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.

(11) 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.

(12) 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.

(13) 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.
(14) 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.
(15) 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.
(16) 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.
(17) Naryzny, S. N.; Legina, O. K. Haptoglobin as a Biomarker. *Biochem. Mosc. Suppl. B Biomed. Chem.* **2021**, *15*, 184-198.
(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) Yağcı, S.; Serin, E.; Acicbe, Ö.; Zeren, M.; Odabaşı, M. S. The relationship between serum erythropoietin, hepcidin, and haptoglobin levels with disease severity and other biochemical values in patients with COVID-19. *Int. J. Lab. Hematol.* **2021**, *43 Suppl 1*, 142-151.

(28) Brito, S.; Ferreira, N.; Mateus, S.; Bernardo, M.; Pinto, B.; Lourenço, A.; Grenho, F. A Case of Autoimmune Hemolytic Anemia Following COVID-19 Messenger Ribonucleic Acid Vaccination. *Cureus* **2021**, *13*, e15035.

(29) Murdych, T. M. A case of severe autoimmune hemolytic anemia after a receipt of a first dose of SARS-CoV-2 vaccine. *Int. J. Lab. Hematol.* **2021**, *in press (doi: 10.1111/ijlh.13653)*.

(30) 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.

(31) 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.

(32) 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.

(33) 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.

(34) 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.

(35) Fries, E.; Wicher, K. B. Evolutionary aspects of hemoglobin scavengers. *Antioxid. Redox Signal.* **2010**, *12*, 249-259.

(36) 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.
(37) 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.
(38) Marty, M. T.; Baldwin, A. J.; Marklund, E. G.; Hochberg, G. K.; Benesch, J. L.; Robinson, C. V. Bayesian deconvolution of mass and ion mobility spectra: from binary interactions to polydisperse ensembles. *Anal. Chem.* **2015**, *87*, 4370-4376.
(39) 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.
(40) Tamara, S.; Franc, V.; Heck, A. J. R. A wealth of genotype-specific proteoforms fine-tunes hemoglobin scavenging by haptoglobin. *Proc. Natl. Acad. Sci. U. S. A.* **2020**, *117*, 15554-15564.
(41) Yang, Y.; Ivanov, D. G.; Kaltashov, I. A. The challenge of structural heterogeneity in the native mass spectrometry studies of the SARS-CoV-2 spike protein interactions with its host cell-surface receptor. *Anal. Bioanal. Chem.* **2021**, *413*, 7205-7214.
(42) 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.
(43) Zaia, J.; Khatri, K.; Klein, J.; Shao, C.; Sheng, Y.; Viner, R. Complete Molecular Weight Profiling of Low-Molecular Weight Heparins Using Size Exclusion Chromatography-Ion Suppressor-High-Resolution Mass Spectrometry. *Anal. Chem.* **2016**, *88*, 10654-10660.
(44) Yang, Y.; Niu, C.; Bobst, C. E.; Kaltashov, I. A. Charge Manipulation Using Solution and Gas-Phase Chemistry to Facilitate Analysis of Highly Heterogeneous Protein Complexes in Native Mass Spectrometry. *Anal. Chem.* **2021**, *93*, 3337–3342.

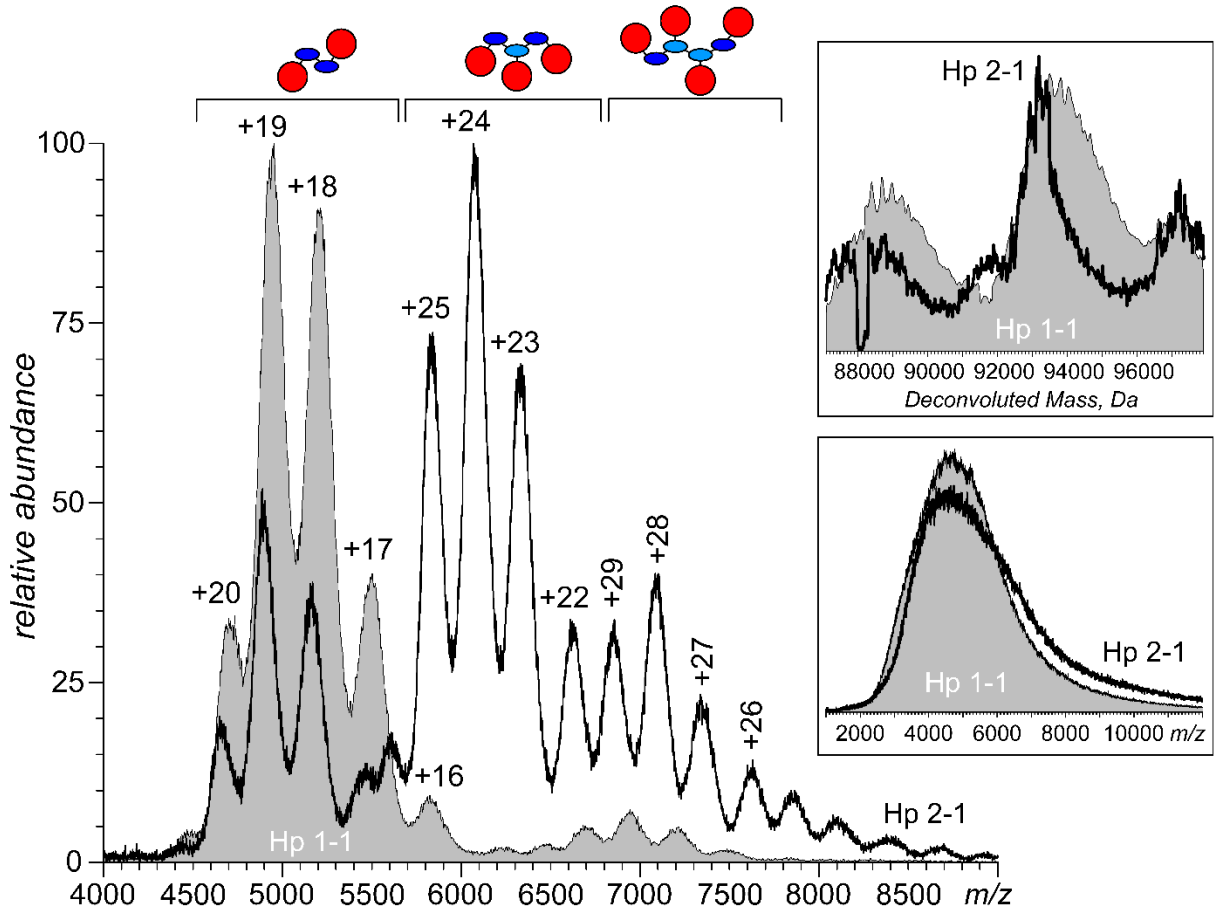

Figure 1. ESI mass spectra of aqueous solutions of Hp 1-1 (0.43 mg/mL) and Hp 2-1 (1.0 mg/mL) in 100 mM ammonium acetate, pH 6.8 shown with a gray-filled curve and a black trace, respectively. Charge state assignments of resolved peaks were made based on the results of limited charge reduction measurements (refer to **Figures 2** and **S2**). The insets on the right show the results of UniDEC deconvolution of the ionic signals in m/z range 4200-5700 (top) and ESI mass spectra acquired under denaturing conditions (bottom).

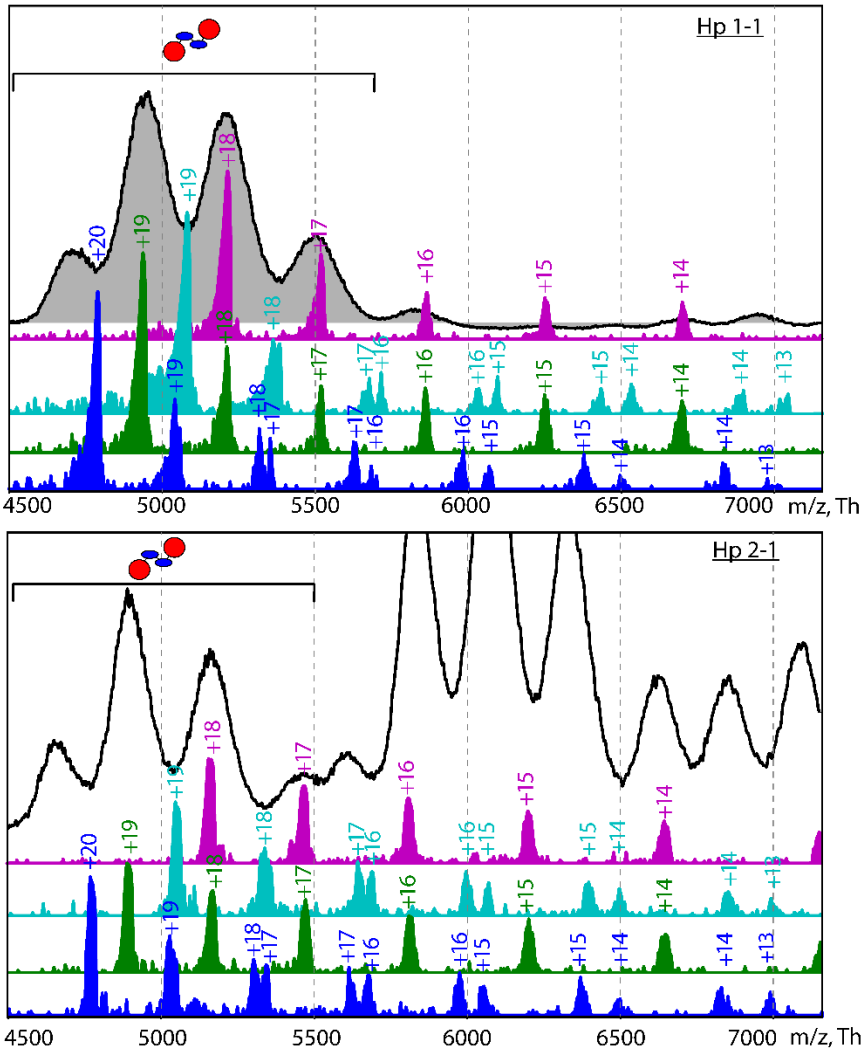

Figure 2. Limited charge reduction of ESI-generated ions of Hp 1-1 (**top panel**) and the lowest-molecular weight species of Hp 2-1 (**bottom panel**). The precursor ions were selected in each case at the apex of one of the most abundant peaks and at the lowest signal intensity points between two partially resolved peaks (isolation windows < 50 *m/z* units). The colored charge ladders represent limited charge reduction measurements for different precursor ions (with the signal intensity plotted on a log-scale).

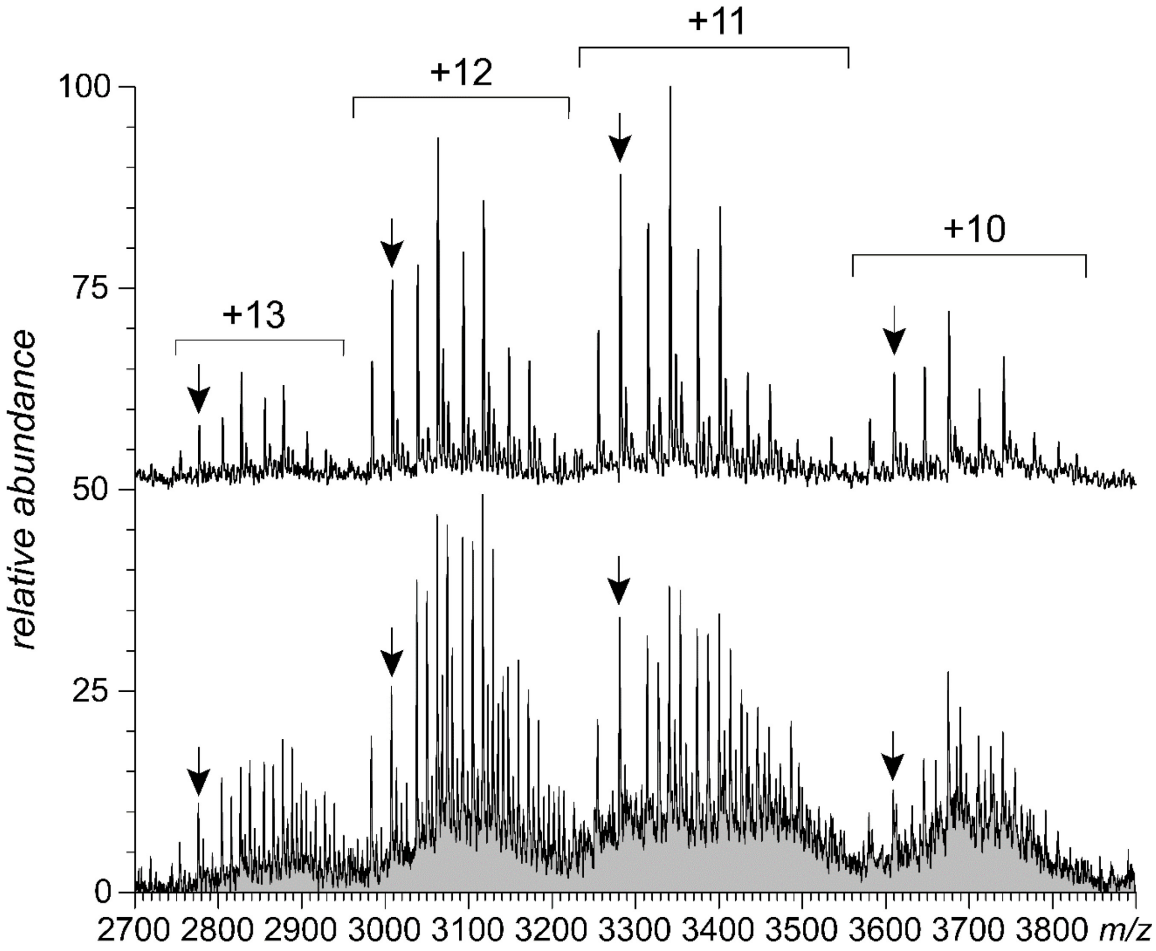

Figure 3. Mass spectra of H-chains obtained by in-line reduction of Hp 1-1 (**top**) and Hp 2-1 (**bottom**) using XP-RC with on-line MS detection. Arrows indicate the m/z values of ions representing a fully reduced H-chain carrying four fully sialylated, fucose-free biantennary N-glycans of the complex type (calculated mass of a neutral species 36,084.4 Da).

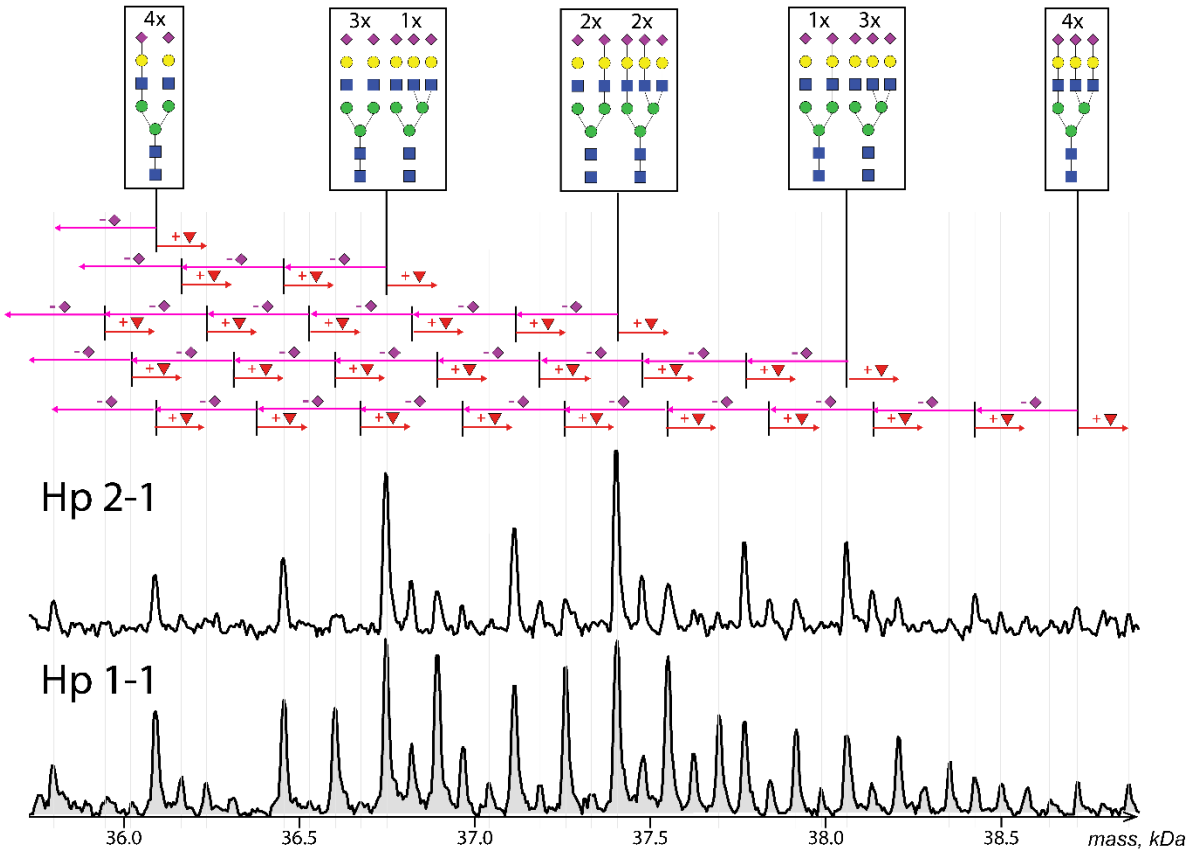

Figure 4. The comparison of mass-distribution of glycoforms between H-chains of Hp 1-1 and 2-1 species. On the top of diagram there is an identification of glycoforms that was done using combinatorial approach. Only non-fucosylated and single-fucosylated species are shown to avoid the overlapping of labels with species with different sialylation level.

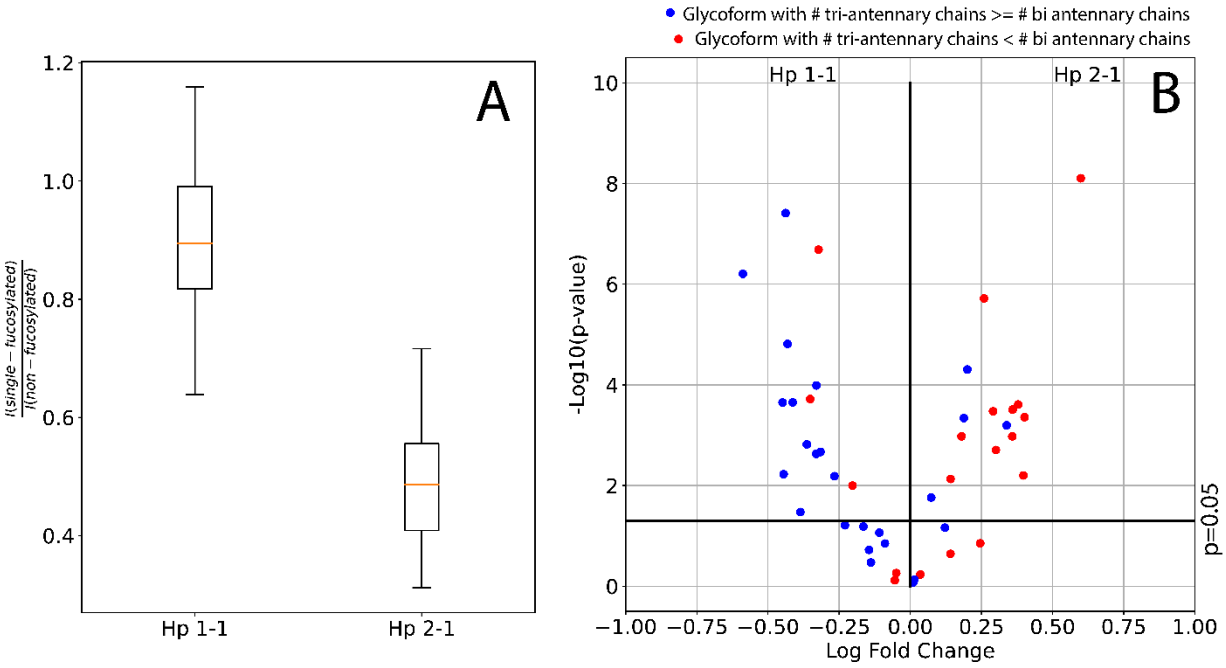

Figure 5. The intensity ratio distributions of ions representing singly-fucosylated species and their non-fucosylated counterparts for all identified glycoforms, calculated for Hp 1-1 and Hp 2-1 heavy chains (**A**). The volcano-plot showing statistical significance (the *p*-value from Student's *t*-test) versus the magnitude of change (Hp 1-1 vs. Hp 2-1) for each peak. The peaks were colored according to the number of tri-antennary chains that occupy N-glycosylation sites. The *x*-axis values refer to the specific peak intensity ratios (H-chains of Hp 2-1 vs. Hp 1-1); *y*-axis is a negative logarithm of the *p*-value between. The glycoforms with fewer tri-antennary structures than bi-antennary ones are marked with red dots; others are marker with blue dots (**B**).