

Exploring the Chemical Space of Protein Glycosylation in Non-covalent Protein Complexes: An Expedition Along Different Structural Levels of Human Chorionic Gonadotropin Employing Mass Spectrometry

Maximilian Lebede,^{||1,2} Fiammetta Di Marco,^{||1,2} Wolfgang Esser-Skala,^{1,2,3} René Hennig,^{4,5} Therese Wohlschlager,^{1,2} Christian G. Huber*^{1,2}

[1,2] M. Lebede, F. Di Marco, Dr. W. Esser-Skala, Dr. T. Wohlschlager, Prof. C.G. Huber
Department of Biosciences, Bioanalytical Research Labs
University of Salzburg
Hellbrunner Straße 34, 5020 Salzburg, Austria
and
Christian Doppler Laboratory for Innovative Tools for Biosimilar Characterization
University of Salzburg
Hellbrunner Straße 34, 5020 Salzburg, Austria
E-mail: c.huber@sbg.ac.at

[3] Dr. W. Esser-Skala
Department of Biosciences, Computational Systems Biology Group
University of Salzburg
Hellbrunner Straße 34, 5020 Salzburg, Austria

[4,5] Dr. R. Hennig
glyXera GmbH,
Brenneckestraße 20 – ZENIT, 39106 Magdeburg, Germany
and
Max Planck Institute for Dynamics of Complex Technical Systems
Sandtorstraße 1, 39106 Magdeburg, Germany

[||] These authors contributed equally, Co-first authors

Abstract: The description of the structure and function of protein glycoforms remains one of the most fundamental challenges of biological chemistry. We report on the application of high-resolution mass spectrometry (MS) and systematic computational data evaluation approaches to unravel the multitude of glycoforms present in recombinant human chorionic gonadotropin (hCG, Ovitrelle®). Upon combining established techniques, such as released glycan and glycopeptide analysis, with novel approaches employing HPLC-MS to characterize protein subunits, and native MS we demonstrate the in-depth analysis of the heavily glycosylated non-covalent hCG heterodimer. Starting from the deconvoluted mass spectrum of dimeric hCG comprising about 50 signals, we explore the chemical space of hCG glycosylation and elucidate concealed complexity. Systematic, stepwise integration of data obtained at the levels of released glycans, glycopeptides, and subunits using a computational annotation tool allows us to reveal 1031 underlying glycoforms. Additionally, two batches of Ovitrelle® are compared to assess potential product variability.

Introduction

The existence of proteoforms – protein variants due to sequence variation and/or post-translational modifications (PTMs) – significantly augments the chemical space of protein structures and has been recognized as a major contribution to the functional diversity and regulation of proteins.^[1] Thus, the structural and functional characterization of proteoforms is a prerequisite for understanding how chemical structures are involved in the complex network of biochemical reactions responsible for the functioning of living organisms. Among the

multitude of PTMs, glycosylation is the one contributing most to protein heterogeneity.

Glycosylation, *i.e.* the conjugation of carbohydrates with a protein backbone, is crucial to the structure and function of certain proteins. A plethora of coexisting glycoforms arises from macroheterogeneity, the presence or absence of a carbohydrate structure at a specific glycosylation site within the protein, as well as microheterogeneity, the occurrence of different glycan structures at a specific site.^[2] Due to the complexity of glycan structures and the subtle structural differences between glycoforms, characterization of glycoproteins poses a substantial analytical challenge.

To address this challenge, analytical techniques such as high-performance liquid chromatography (HPLC) and mass spectrometry (MS) have been employed at different levels of structural complexity, ranging from released glycans, glycopeptides, to intact proteins, and protein complexes.^[3,4] Released glycan analysis upon removal of the glycans from the protein backbone provides information on glycan structure but not on the localization within the protein.^[5] In contrast, glycopeptide analysis allows acquisition of site-specific glycosylation data.^[6] Molecular context can be preserved by employing MS at the intact protein level to acquire information on the whole glycoprotein and to reveal co-occurrence of different glycoforms.^[7] Finally, to unravel biomacromolecular higher-order structure, non-covalent protein complexes can be analyzed by MS under non-denaturing conditions, referred to as native MS.^[8–13] In the context of glycoproteins, native MS permits elucidation of coexisting multi-glycoform complexes arising from combinations of different glycoforms.^[14,15]

Nevertheless, isobaric carbohydrate building units as well as the high complexity of glycan structures lead to the presence of many glycoforms with identical mass, thus indistinguishable by MS. Furthermore, native measurement of protein complexes is often limited in resolving glycoforms of similar mass. Specifically, the natural width of isotopologue clusters hampers the resolution

of small mass differences in mass spectra of proteins or protein complexes.^[16,17] At this level, one cannot discriminate between glycoforms comprising two fucose residues (292 Da) or one sialic acid residue (291 Da). Hence, enzymatic dissection may be employed to reduce spectral complexity and to enable unambiguous identification of different glycoforms.^[18]

As no individual analytical technique suffices to comprehensively characterize protein glycosylation, current developments tend towards a combination of approaches at different structural levels to assess glycosylation. This so-called hybrid MS strategy overcomes the shortcoming of the individual methods but generates a large amount of interdependent data, necessitating bioinformatic tools for evaluation.^[19–23]

In this study, we assess different levels of glycoprotein characterization, ranging from released glycans to intact non-covalent glycoprotein complexes. As a model for proteoform heterogeneity, we utilize human chorionic gonadotropin (hCG), a highly glycosylated protein hormone involved in early embryonal communication and promotion of pregnancy.^[24–26] The protein exists as a hetero-dimer comprising two non-covalently interacting subunits, alpha (hCG α) and beta (hCG β). *N*-glycosylation occurs at two sites in each hCG α (N52, N78) and hCG β (N13, N30).^[27] *O*-glycosylation sites are unique to hCG β and have been described in its unfolded serine-rich C-terminal region (S121, S127, S132, S138).^[28] A fifth *O*-glycosylation site (S130 or T140) has been proposed by Bai et al. based on an approach combining enzymatic dissection and tandem MS (MS/MS).^[29]

As a therapeutic, hCG is applied in the treatment of infertility in men and women.^[30] One of the commercially available drug products, Ovitrelle®, contains recombinant hCG, expressed in Chinese hamster ovary (CHO) cells. From a functional perspective, *O*-glycosylation and sialylation in gonadotropins increase serum half-life and biological activity, while removal of *N*-glycans reduces receptor binding affinity.^[31] Glycosylation therefore constitutes a critical quality attribute (CQA) of gonadotropin drug products.

Characterization of hCG has been addressed by HPLC or capillary electrophoresis (CE), mostly hyphenated to MS, at the level of released glycans,^[32] glycopeptides,^[28,29,33,34] and, rarely, at the intact protein level.^[35–40] Camperi et al.^[36–38,40,41] employed different techniques for separation of hCG α glycoforms; annotation of mass spectra was attempted by a non-combinatorial approach resulting in only approximately forty identified glycoforms.^[41] With regard to hCG β , neither separation of glycoforms nor acquisition of mass spectra of this subunit was achieved. Toll et al.^[35] previously accomplished analysis of both intact hCG α and hCG β by HPLC-MS. Applying a combinatorial approach, they were able to assign specific glycoforms for hCG α . However, high spectral complexity and isobaricity of variants prevented unambiguous peak assignment for hCG β .

Here, we introduce a global approach to characterize glycosylation patterns of recombinant hCG in the drug product Ovitrelle® at multiple structural levels applying multiplexed capillary gel electrophoresis with laser-induced fluorescence detection (xCGE-LIF), HPLC-MS, native MS and the computational annotation tool MoFi.^[18,42] In order to integrate the data, we build libraries at the level of released glycans, glycopeptides and subunits for stepwise knowledge transfer across all structural levels up to the non-covalent protein complex. Furthermore, we scrutinize underlying glycoforms by enzymatic dissection of intact protein subunits as well as of the non-covalent complex to reduce spectral complexity. For the first time we acquire isotopically resolved spectra of intact hCG β subunit as well as native mass spectra of the hCG dimer, which proves useful for highly informative batch-to-batch comparisons of biopharmaceuticals.

Results and Discussion

Glycoform profiles represent “fingerprints” of hCG drug products.

Native MS is a powerful tool to obtain global information about composition and PTMs of non-covalent protein complexes avoiding lengthy sample preparation or potential introduction of artificial modifications. We analyzed two batches of the hCG-based drug product Ovitrelle® to generate mass fingerprints for batch-to-batch comparison (for raw spectra, see Figure S1). To maintain the dimeric hCG complex in a quasi-native conformation, samples were buffer-exchanged into 150 mM ammonium acetate and analyzed by direct infusion using static nano-ESI-MS. The resulting deconvoluted mass spectra of two Ovitrelle® batches were highly complex with two distinct sets of peaks, ranging from 34.5 to 39.6 kDa (Figure 1). We concluded that the detected masses correspond to the hCG heterodimer existing in a multitude of glycoforms. Considering the difference between theoretical mass of the protein-backbone of the subunits (25.7 kDa) and observed masses, we attributed $\approx 30\%$ of molecular mass (12 kDa) to attached carbohydrate structures. The spectra revealed reoccurring mass differences arising from either increasing content of sialic acids (291 Da) or multiple fucoses ($2 \times 146=292$ Da, Figure 1). Moreover, mass differences of 365 Da and 146 Da were observed, corresponding to HexNAc₁Hex₁ and one fucose residue, respectively. Both batches of Ovitrelle® resulted in signals with equal masses, which varied slightly in abundance, reflecting overall comparable glycoform patterns.

Mass spectra of the hCG complex obtained by native MS thus provided a holistic picture of heterogeneity arising from glycosylation. Furthermore, information on total glycan mass could be obtained. However, no information on individual glycan structures or site-occupancy was gained, preventing the assignment of specific glycoforms to observed signals. Thus, we devised a systematic workflow for in-depth analysis of hCG glycoforms at different structural levels to gain further insight into the chemical space of hCG glycosylation, applicable in the comparison of two Ovitrelle® batches (Figure S2).

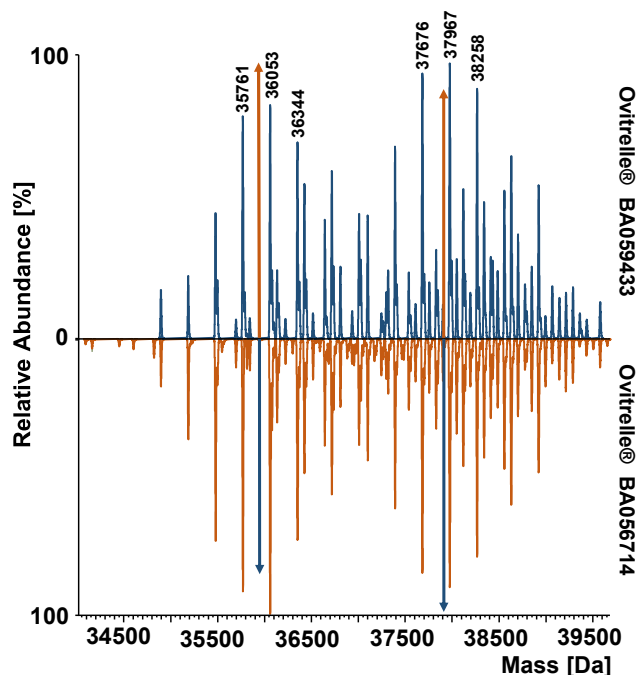


Figure 1. Mirror plot of deconvoluted mass spectra of dimeric hCG of Ovitrelle® batches BA059433 and BA056714. Spectra were obtained with an instrument resolution setting of $R=17,500$ at m/z 200. Colored arrows represent peak height of the highest peak in the signal series of the mirrored batch.

Released glycan analysis provides insight into the origin of complexity.

Initially, *N*-glycans were released by treatment with PNGase F, labelled with aminopyrene-1,3,6-trisulfonic acid, and analyzed using xCGE-LIF (Figure S3). Forty-three peaks were detected and identified via migration time matching with database entries, and additionally confirmed with repeated measurements after exoglycosidase digests. Mainly, complex type *N*-glycans carrying sialic acids (Neu5Ac) were observed. Specifically, biantennary *N*-glycans, such as A2S1G1 and A2S2 predominated (a summary of all possible glycan structures with name and composition can be found in the Sup. File 1, Released glycan data). Additionally, mono- and triantennary structures were detected at lower abundances. Some structures also exhibited core fucosylation.

Furthermore, we analyzed chemically released *O*-glycans by porous graphitized carbon HPLC-MS/MS (Figure S4). *O*-glycan structures were assigned based on elution times and fragmentation patterns. Identified *O*-glycans were of the core 1-type (HexNAc₁Hex₁) with either one or two attached sialic acids. The varying degree of sialylation and fucosylation of the identified *N*- and *O*-glycan structures are well in accordance with the delta mass series of 146 and 291 Da previously observed at the level of the hCG dimer (Figure 1).

Glycopeptide analysis reveals the site-specific context of glycan structures.

All possible *N*- and *O*-glycan structures were combined in a glycan library (Sup. File 1, Released glycan data) for use in identification of glycopeptides. Glycopeptides were analyzed after tryptic digestion by nano HPLC-MS/MS and quantified in a label free approach by means of extracted ion current chromatograms (Figure S5). Since Neu5Gc is isobaric to Neu5Ac plus oxygen, quantitative assessment of the glycopeptides containing Neu5Gc based on MS1 was impossible due to the presence of simultaneous oxidation of Met in the corresponding peptide. Nevertheless, due to their low abundance of less than 3% at released glycan level we decided to remove these glycans from the library. In accordance with released glycan data, predominantly biantennary glycans were identified for the two *N*-glycosylation sites of both subunits at the peptide level. However, we observed site-specific differences with regard to *N*-glycan structures, e.g. type, fucosylation or sialylation (Figure S6).

Unambiguous site-specific assignment of *O*-glycans was hampered by the co-existence of eight serine residues and one threonine residue in close sequential proximity, which led to the generation of hCG β peptides with multiple possible *O*-glycosylation sites. Detected tryptic glycopeptides (115-122 β , 123-133 β , 134-145 β) carried up to two *O*-glycans each, indicating a maximum of six *O*-glycans present at intact level (Figure S4). Therefore, we were unable to assign *O*-glycosylation to specific sites within those peptides (Figure S7).

Analysis of subunits provides information about molecular context of glycosylation.

Data on *N*- and *O*-glycan structures and glycosylation site occupancy together with information about relative glycan abundances (site-specific glycan library) constitutes the basis for the assignment of possible glycoforms at the level of hCG subunits.

Interpretation of the glycoform patterns at a mass tolerance of 20 ppm was realized by a two-step algorithm (modification finder, MoFi). MoFi combines monosaccharide composition determined via intact molecular mass measurements and the site-specific glycan library, as described in detail in ref.^[42] MoFi is able to calculate the contribution to peak height of individual isobaric glycoforms (hit scores) based on site-specific relative abundance of glycans retrieved from glycopeptide analysis.

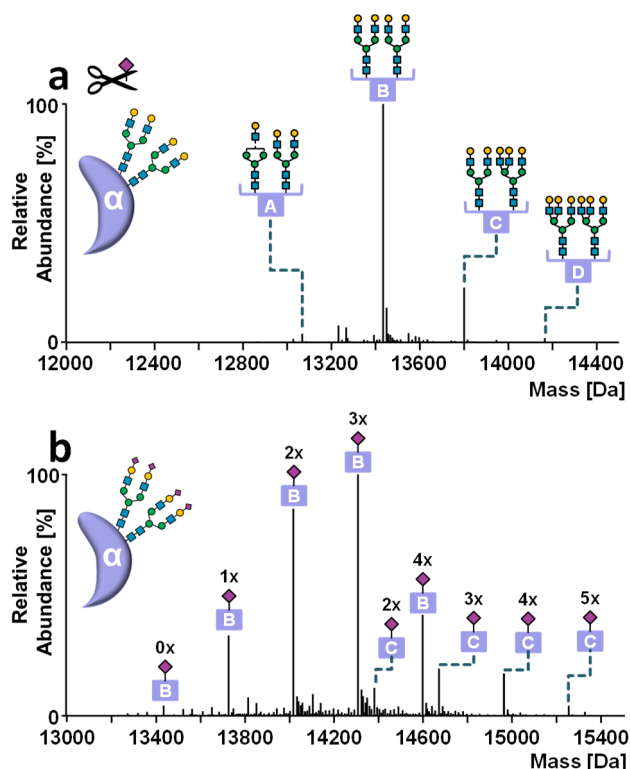


Figure 2. Annotation of hCG α glycoforms. Deconvoluted mass spectra of hCG α obtained by HPLC-MS analysis a after desialylation or b without enzymatic treatment. Glycoforms are indicated by letters A–D. Numbers above the peaks refer to Neu5Ac residues. Peak lists with all possible glycoform assignments for untreated and desialylated hCG α are available in Supplementary Data, Annotations hCG α BA059433. Instrument resolution setting was 140,000 at *m/z* 200.

In order to obtain subunit spectra, RP-HPLC-MS analysis of Ovitrelle® was carried out under denaturing conditions and revealed two sets of chromatographic peaks. The first can be attributed to hCG α and the second arises from hCG β , according to molecular masses (see Figure S8 for the chromatogram and S9d for raw spectra). Employing a mass spectrometric resolution setting of 140,000, we were able to record isotopically resolved spectra of both subunits (Figure S9d and S10).

Due to the high spectral complexity of subunits arising from many underlying glycoforms, reduction of molecular diversity upon glycosidase digestion was used to support the annotation of glycoform profiles, as outlined in detail in ref.^[18]

For desialylated hCG α , fucosylated and non-fucosylated bi- and triantennary *N*-glycan structures were readily assignable in the deconvoluted spectra of hCG α treated with sialidase (Figure 2a) using a site-specific glycan library, where sialic acids were removed in silico and abundances were recalculated. The 194 signals ranging from 13.0-15.5 kDa of untreated hCG α were consecutively annotated using the site-specific glycan library (Figure 2b). The most prominent glycoform series comprised two biantennary *N*-glycans differing in the number of terminal sialic acids, ranging from zero (A2G2/A2G2) to four (A2S2/A2S2) (letter B, Figure 2b).

Due to the existence of isobaric variants, multiple glycoforms were assigned for each mass, taking into account individual glycan contributions. For example, the most abundant hCG α signal at 14307.1 Da was assigned to the glycoform A2S1G1/A2S2, contributing to 98.9% of peak abundance. The remaining 1.1% is explained by minor *N*-glycoforms.

For hCG β , analysis of de-*N*-glycosylated and desialylated hCG revealed the number of *O*-glycan cores mainly ranging from four to six (Figure 3a). In a next step, we analyzed PNGase F-treated hCG to obtain information on sialic acid variants of *O*-glycoforms (Figure 3b). Intriguingly, an hCG β glycoform

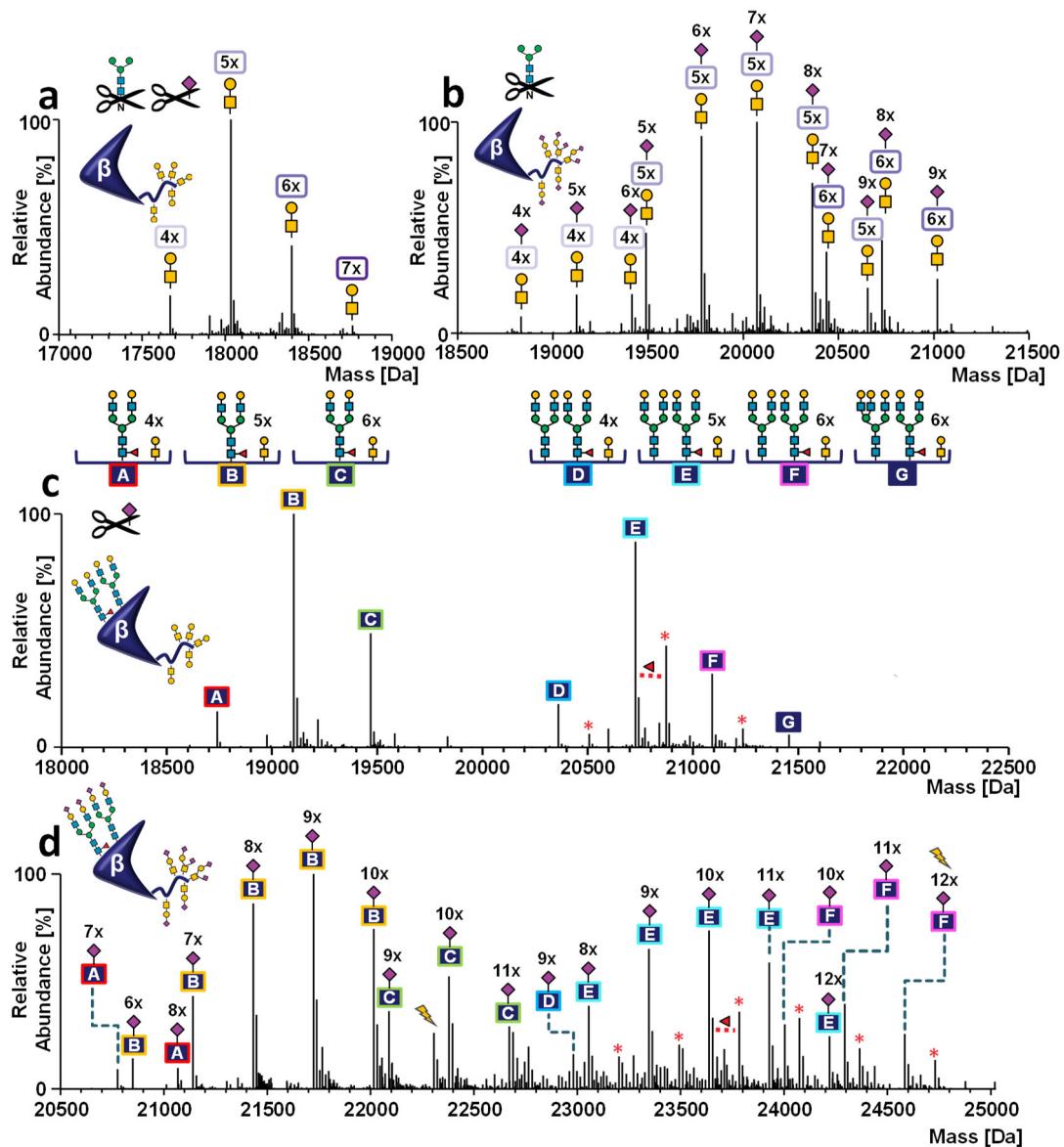


Figure 3. Annotation of hCG β glycoforms. Deconvoluted mass spectra of carbamidomethylated hCG β obtained by HPLC-MS analysis **a** after sialidase and PNGase F treatment, **b** after PNGase F treatment, **c** after sialidase treatment, and **d** in untreated form. Glycoforms are indicated by letters A–F. Red asterisks indicate fucose variants. The number of Neu5Ac residues is indicated by the numbers above the peaks. Lightning bolts indicate two peaks that are further discussed in the main text. Peak lists with all possible glycoform assignments for untreated and desialylated hCG β are available in Supplementary Data, Annotations hCG β BA059433. Instrument resolution setting was 140,000 at m/z 200.

comprising five O-glycans and nine sialic acids was detected, which was not supported by glycopeptide data. This variant can only be explained when at least one O-glycopeptide carries two O-glycan cores with four sialic acid, but no such glycopeptide was detected (Figure S7). Subsequent analysis of sialidase-treated hCG β revealed two abundant series of signals (Figure 3c). Using the approach described for the alpha subunit (in silico desialylated glycan library), we were able to assign desialylated N-glycoforms revealing the degree of galactosylation (Figure 3c). The two series of signals corresponded to glycoforms carrying one (letters A–C) or two N-glycans (letters D–G) and four to six O-glycan cores. Specifically, we detected biantennary N-glycan A2G2F (letters A–C, Figure 3c), the two biantennary N-glycans with one fucose A2G2/A2G2F (letters D–F, Figure 3c), or the two fucosylated biantennary N-glycans A2G2F/A2G2F (Figure 3c, glycoforms indicated by red asterisks). Glycopeptide data indicates site-occupancy of 62.6% at N13 β and fucosylation one third of N-glycans at this position. Moreover, site-specific differences were observed, as N30 β glycans were almost completely fucosylated. These findings are also reflected in the intact spectrum of hCG β .

Finally, we analyzed sialylated N- and O-glycoforms of the intact hCG β subunit. The deconvoluted mass spectrum of intact hCG β (408 signals ranging from 20.5–25.0 kDa) exhibited a signal pattern that is explained by a combination of the identified N- and O-glycoforms including terminal sialic acids as revealed upon integration of site-specific glycan library including sialic acids (letters A–F, Figure 3d). Complete annotation lists comprising 755 (hCG α) and 16894 (hCG β) entries can be found in Sup. File 2, Annotations hCG α BA059433 and Annotations hCG β BA059433.

Annotation of the deconvoluted hCG β mass spectrum was successful and plausible for all but two peaks (lightning bolts in Figure 3d). MoFi annotation of the peak at 22306.2 Da was A3S2G1F + LacN1/M5-A1G1 and five core-1 with one sialic acid with a hit score of 100. From the PNGase F digest (Figure 3b) we know that no O-glycoform comprising five core-1 and only a single Neu5Ac exists. Considering that this peak shows a delta mass of +291 Da compared to the last peak of the B-series, it is likely that the correct annotation is unmodified/A2S2F and 5 core-1 + 9 S, which is supported by PNGase F digest data. The wrong

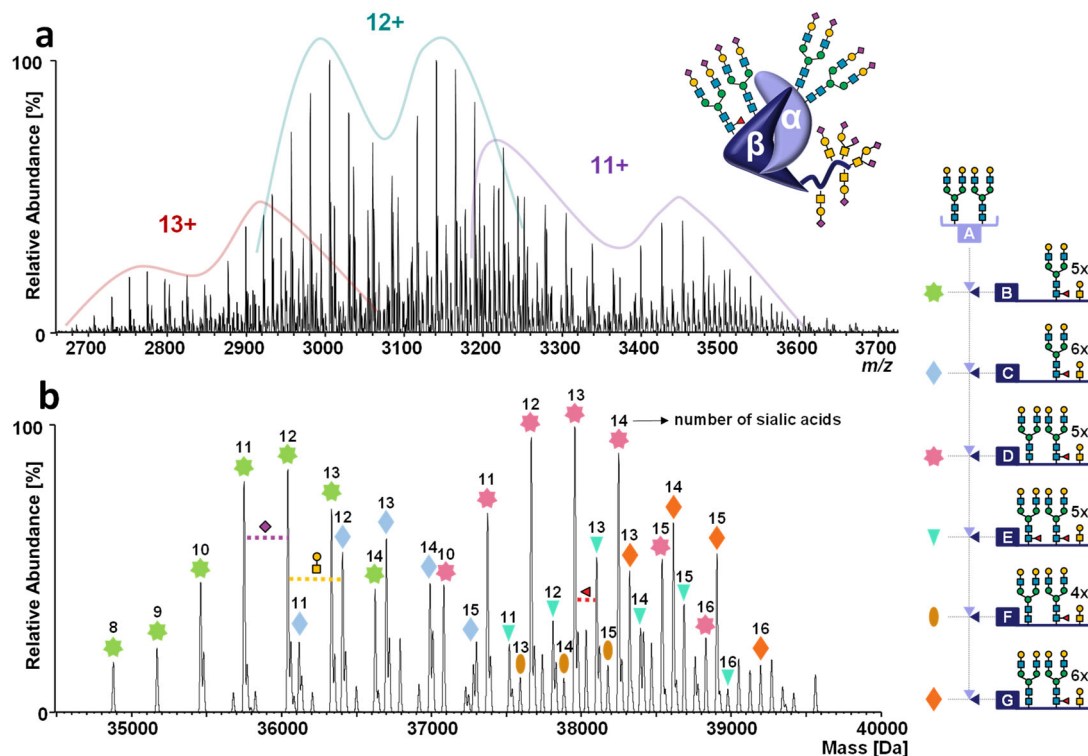


Figure 4. Annotation of dimeric hCG glycoforms. **a** Raw mass spectrum of dimeric hCG obtained by native MS. Charge states are indicated. **b** Deconvoluted mass spectrum of dimeric hCG. The glycoforms contributing most to peak abundance are indicated by colored symbols and represent a combination of glycans present on hCG α (A) and hCG β (B–G). Numbers above these symbols correspond to the number of sialic acids present. The spectrum was obtained with an instrument resolution setting of $R=17,500$ at m/z 200.

annotation stems from the absence of tryptic O-glycopeptides carrying two core-1 and four sialic acids in the site-specific glycan library, as this variant was not detected by nano HPLC-MS/MS.

Additionally, no glycoform could initially be attributed to the mass of 24583.7 Da using a mass tolerance of 20 ppm. Annotation was only possible when a mass tolerance of more than 44 ppm was set. The resulting glycoform A2S2/A2S2F and 6 core-1 + 8 S had a theoretical mass of 24584.8 Da. This unusually high mass error was most likely caused due to an artifact of deconvolution resulting in a mass shift of -1 Da, a phenomenon often observed for proteins with a sulfur content above average.^[43]

Stepwise data integration of structural levels facilitates deciphering the chemical space of dimeric hCG.

By integrating data of lower structural levels (released glycan, glycopeptide), we were able to elucidate and assign glycosylation patterns of intact hCG α and hCG β subunits. With this information in hand, we tackled the initially obtained native spectrum of the complex and investigated how the subunits combine within the heterodimer.

The raw spectrum (Figure 4a) showed three peak series in the m/z range between 2500–4000, corresponding to charge states from 11+ to 13+. The deconvoluted spectrum (Figure 4b) featured 48 signals in a mass range spanning from 34.5 to 39.6 kDa. Some peaks can be observed as doublets due to sodium adducts formed (mass difference +22 Da) which may be attributed to incomplete buffer exchange. Analogous to the glycopeptide library, we built a library of hCG α and hCG β glycoforms to be used in MoFi to annotate the deconvoluted mass spectrum of the hCG complex. Using subunit instead of glycopeptide data, we were able to reduce the number of theoretical combinations by more than seventeen-fold (3×10^6 instead of 51×10^6 possibilities considering glycoforms and excluding oxidation). In case of hCG data, MoFi was limited to approximately 120 entries in the glycan

list (since the combinatorial search space would explode with more entries). Thus, we decided to limit the number of glycoforms used for the library by employing a cut-off of 0.3% fractional abundance. Hence, the considered glycoforms (42 for hCG α and 58 for hCG β) described 89.4% and 59.9% of total glycoform abundance, respectively. For hCG β , we assumed that individual contributions of low-abundant subunit variants to peak intensity were negligible because 98.7% of hCG β glycoforms (10984 out of 11126) are each below the 0.1% fractional abundance threshold and make up only $\approx 35.7\%$ of the total intensity. Thus, we assume that peak annotation was not significantly impacted by eradication of these glycoforms. When calculating the abundance of subunit glycoforms, we did not consider oxidation variants observed at intact level ($\Delta m = +16$ Da) as distinct proteoforms and summed the abundances of corresponding non-oxidized and oxidized species. Using the subunit library, 1031 glycoforms were identified for 48 detected masses (a complete list of annotations can be found in [Sup. File 2, Annotations Dimer BA059433](#)).

In the deconvoluted spectrum of dimeric hCG shown in Figure 4b, the variant explaining the highest portion of the abundance of each peak (as calculated by MoFi) is reported. Most N-glycoforms were composed of sialylated biantennary complex type glycans, while triantennary N-glycans were observed to a lesser extent. As expected from subunit data, the deconvoluted spectrum shows five main signal series of glycovariants with delta masses of 291 Da, corresponding to a varying number of sialic acid residues. The two main patterns (green and pink stars in Figure 4b) are due to sialic acid variants of N-glycans with biantennary structures. A delta mass of 365 Da is also observable and may be attributed to variants with an additional antenna (GlcNAc₁Gal₁ for N-glycans or GalNAc₁Gal₁ for O-glycans; light blue and orange rhomboids in Figure 4b).

Moreover, a difference in mass of 146 Da was observed arising from variants comprising two fucosylated N-glycans on hCG β (turquoise triangles and brown ovals in Figure 4b). At an

instrument resolution setting of 17.500 it is not possible to distinguish between a glycoform comprising a sialic acid (291 Da) or two fucose (292 Da) residues based on the delta mass. Since all abundant variants carried sialic acid residues, enzymatic desialylation was performed in order to clarify these mass differences. After desialylation, the raw spectrum (Figure S11a) showed significantly lower complexity compared to the intact dimer. Charge state distribution was practically unaffected by removal of sialic acids, as the three main charge states remained 11+ to 13+. The deconvoluted mass spectrum (Figure S11b) showed 13 peaks in a mass range between 31 and 36 kDa. In Figure S11b, the most abundant hCG glycoforms are reported as a combination of hCG α and β glycovariants (the complete annotation is available in Sup. File 2, Annotations Dimer BA059433). No mass differences of 291 or 292 Da were observed in this spectrum, confirming mass shifts to arise from sialic acids and not multiple fucoses in the spectrum of the untreated complex (e.g. pink stars in Figure 4b).

Glycoform annotations are relevant for biopharmaceutical quality control.

In-depth hCG characterization of the second Ovitrelle® batch (Sup. File 2, Annotations Dimer, hCG α or hCG β , BA056714) allowed the detailed assessment of glycosylation for batch-to-batch comparison at multiple structural levels. First, we decided to derive rather global parameters such as the overall degree of sialylation and core-fucosylation from the glycoform annotation results. To obtain the degree of sialylation we divided the amount of sialic acids observed in a given glycopeptide or glycoform by the calculated maximum number of sialic acids that could be present on this glycopeptide/glycoform (one per *N*-glycan antenna, two per *O*-glycan). All ratios were weighted for abundance to calculate the mean degree of sialylation. Core-fucosylation was addressed in a similar fashion considering a maximum of one core-fucose per *N*-glycan. Indeed, core-fucosylation and sialylation degrees were consistent between the two Ovitrelle® batches across all structural levels (Table 1).

Table 1. Comparison of the degrees of fucosylation and sialylation in both Ovitrelle® batches at different structural levels.

Structural level	Target	Degree of fucosylation		Degree of sialylation	
		BA059433	BA056714	BA059433	BA056714
Glyco-peptides	hCG α	0.028	0.026	0.61	0.59
	hCG β	0.66	0.68	0.63	0.63
Subunits	hCG α	0.10	0.071	0.65	0.64
	hCG β	0.71	0.73	0.67	0.67
Dimer	hCG dimer	0.33	0.34	0.71	0.71

Looking at individual glycoform signals, slight apparent differences in abundance of glycoforms were observable in the deconvoluted native spectra of Ovitrelle® batches (Figure 1) and verified by comparing the charge state signals in the raw spectra (Figure S1). Batch BA059433 showed higher abundance of glycoforms lacking *N*-glycosylation at N13 β (signal series ranging from 34.8 to 37.1 kDa in Figure 1). In contrast, fully *N*-glycosylated glycoforms were more prominent in BA056714 (signal series ranging from 37.1 to 39.6 kDa in Figure 1). However, after calculating the fractional abundance of glycoforms according to MoFi score and relative peak abundance, we observed that the most abundant peak (37967.0 Da for BA059433 and 36052.6 Da for BA056714) did not contain the most abundant glycoform. Interestingly, the variant A2S1G1/A2S2/unmodified/A2S2F/1 x Core + 2 x S/2 x Core + 3 x S/2 x Core + 3 x S with a mass of 36343.5 Da (Figure 1) was identified as the most abundant for

both batches, despite the relative abundance of the corresponding mass peak in the deconvoluted spectrum of the hCG dimer was 71.1% for BA059433 and 73.0% for BA056714.

Conclusion

The integration of comprehensive data on protein glycosylation from different structural levels facilitates the exploration of the chemical space of highly heterogeneous glycoforms in protein complexes and provides valuable insights into the structural intricacies of glycoproteins. Analyzing intact hCG subunits and hCG dimer highlights the benefit of intact protein/protein complex over glycopeptide analysis, because the structural context of microheterogeneity as well as macroheterogeneity remains preserved and allows for direct assessment of specific glycoforms. Moreover, in analyzing the non-covalent protein complex of hCG consisting of highly glycosylated subunits, we took advantage of two assets of native MS – the increased spatial resolution and the preservation of non-covalent interactions. The resulting mass spectrum discloses heterogeneity of the hCG heterodimer and provides detailed information on glycoforms of the intact complex.

Although the number of distinguishable signals in the mass spectrum of the intact complex is in the range of only 50, computer-aided integration of semiquantitative data of released glycans, glycopeptides, and protein subunits facilitates the assignment of more than 1000 underlying glycoforms that can explain the experimentally observed pattern of proteoforms. The discrepancy between the number of observed signals and the number of underlying proteoforms rests within (1) the fundamental inability of MS to resolve isobaric proteoforms as well as (2) overlapping isotope clusters of variants having small mass differences. This prevents the distinction of proteoforms having mass differences in the order of 25 Da or less for a 38 kDa protein, irrespective of achievable mass spectrometric resolution (see ref. [16]).

Nevertheless, our probability-based calculation of the contribution of distinct glycoforms to overall observed signal intensity (hit scores in Sup. File 2, Annotations Dimer BA059433) reveals that in most cases, a limited number of 5 different glycoforms contribute to the major portion of observed intensity in a distinct signal. Hence, about 385 different glycoforms of more than 0.04% fractional abundance explain 90% of all signal intensities in the glycoform pattern shown in Figure 4b.

Finally, we successfully demonstrate the capability of our approach to investigate drug product quality attributes. Obtained glycoprotein profiles may serve as fingerprints, e.g. in the assessment of batch-to-batch variability. We believe that advancement in instrument and software performances will increase the robustness of the approaches involving native MS and enable quick characterization of glycosylated biopharmaceuticals at intact level in the future.

Acknowledgements

This work was funded by the Austrian Federal Ministry for Digital and Economic Affairs, the National Foundation of Research, Technology, and Development, a Start-up Grant of the State of Salzburg, as well as the Austrian Science Fund (W1213).

We thank U. Lohrig (Novartis) and K. Scheffler (Thermo Fisher Scientific) for critically proofreading the manuscript and Erdmann Rapp as well as Xenia K. Hoffmann (glyXera GmbH) for released glycan analysis.

Conflict of interest

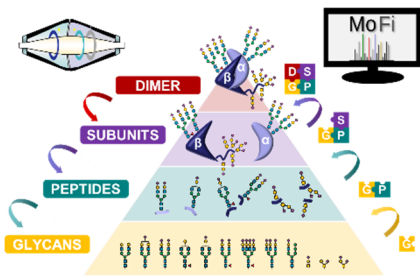
Novartis AG / Sandoz GmbH as well as Thermo Fisher Scientific provide financial support for the Christian Doppler Laboratory for Innovative Tools for Biosimilar Characterization. René Hennig is employee of glyXera GmbH, Magdeburg, Germany. The salaries

of Wolfgang Esser-Skala and Therese Wohlschlager are fully funded; Christian G. Huber's salary is partly funded by the Christian Doppler Laboratory for Biosimilar Characterization. The authors declare no other competing financial interest.

Keywords: bioinformatics • biopharmaceutical • glycoprotein characterization • human chorionic gonadotropin • mass spectrometry

- [1] L. M. Smith, N. L. Kelleher, *Nat. Methods* **2013**, *10*, 186–187.
- [2] L. F. Zacchi, B. L. Schulz, *Glycoconj. J.* **2016**, *33*, 359–376.
- [3] K. Sandra, I. Vandenheede, P. Sandra, *J. Chromatogr. A* **2014**, DOI 10.1016/j.chroma.2013.11.057.
- [4] A. Beck, E. Wagner-Rousset, D. Ayoub, A. Van Dorsselaer, S. Sanglier, C. Cianférani, **2012**, DOI 10.1021/ac3032355.
- [5] M. Wuhler, A. R. De Boer, A. M. Deelder, *Mass Spectrom. Rev.* **2009**, *28*, 192–206.
- [6] L. R. Ruhaak, G. Xu, Q. Li, E. Goonatilleke, C. B. Lebrilla, *Chem. Rev.* **2018**, *118*, 7886–7930.
- [7] J. D. Tipton, J. C. Tran, A. D. Catherman, D. R. Ahlf, K. R. Durbin, N. L. Kelleher, **2011**, DOI 10.1074/jbc.R111.239442.
- [8] A. J. R. Heck, *Nat. Methods* **2008**, *5*, 927–933.
- [9] R. J. Rose, E. Damoc, E. Denisov, A. Makarov, A. J. R. Heck, *Nat. Methods* **2012**, *9*, 1084–1086.
- [10] S. Rosati, R. J. Rose, N. J. Thompson, E. Van Duijn, E. Damoc, E. Denisov, A. Makarov, A. J. R. Heck, *Angew. Chemie - Int. Ed.* **2012**, *51*, 12992–12996.
- [11] A. J. R. Heck, R. H. H. Van Den Heuvel, *Mass Spectrom. Rev.* **2004**, *23*, 368–389.
- [12] N. J. Thompson, S. Rosati, A. J. R. Heck, *Methods* **2014**, *65*, 11–17.
- [13] L. F. Schachner, A. N. Ives, J. P. McGee, R. D. Melani, J. O. Kafader, P. D. Compton, S. M. Patrie, N. L. Kelleher, *J. Am. Soc. Mass Spectrom.* **2019**, 1–9.
- [14] D. Wu, W. B. Struwe, D. J. Harvey, M. A. J. Ferguson, C. V. Robinson, *Proc. Natl. Acad. Sci. U. S. A.* **2018**, *115*, 8763–8768.
- [15] S. Tamara, V. Franc, A. J. R. Heck, *Proc. Natl. Acad. Sci.* **2020**, 202002483.
- [16] C. Regl, T. Wohlschlager, W. Esser-Skala, I. Wagner, M. Samonig, J. Holzmann, C. G. Huber, *MAbs* **2019**, *11*, 569–582.
- [17] P. Lössl, J. Snijder, A. J. R. Heck, *J. Am. Soc. Mass Spectrom.* **2014**, *25*, 906–917.
- [18] T. Wohlschlager, K. Scheffler, I. C. Forstenlehner, W. Skala, S. Senn, E. Damoc, J. Holzmann, C. G. Huber, *Nat. Commun.* **2018**, *9*, DOI 10.1038/s41467-018-04061-7.
- [19] Y. Yang, F. Liu, V. Franc, L. A. Halim, H. Schellekens, A. J. R. Heck, *Nat. Commun.* **2016**, *7*, 1–10.
- [20] Y. Yang, G. Wang, T. Song, C. B. Lebrilla, A. J. R. Heck, *MAbs* **2017**, *9*, 638–645.
- [21] Y. Yang, V. Franc, A. J. R. Heck, *Trends Biotechnol.* **2017**, *35*, 598–609.
- [22] R. Upton, L. G. Migas, K. J. Pacholarz, R. G. Beniston, S. Estdale, D. Firth, P. E. Barran, *Chem. Sci.* **2019**, *10*, 2811–2820.
- [23] Y. H. Lin, J. Zhu, S. Meijer, V. Franc, A. J. R. Heck, *Mol. Cell. Proteomics* **2019**, *18*, 1479–1490.
- [24] C. Ticconi, A. Zicari, A. Belmonte, M. Realacci, C. V. Rao, E. Piccione, *Placenta* **2007**, *28*, S137–S143.
- [25] M. Tsampalás, V. Grیدهlet, S. Berndt, J. M. Foidart, V. Geenen, S. P. d'Hauterive, *J. Reprod. Immunol.* **2010**, *85*, 93–98.
- [26] L. A. Cole, *Reprod. Biol. Endocrinol.* **2010**, *8*, 1–14.
- [27] A. J. Laphorn, D. C. Harris, A. Littlejohn, J. W. Lustbader, R. E. Canfield, K. J. Machin, F. J. Morgan, N. W. Isaacs, *Nature* **1994**, *369*, 455–461.
- [28] L. Valmu, H. Alfthan, K. Hotakainen, S. Birken, U. H. Stenman, *Glycobiology* **2006**, *16*, 1207–1218.
- [29] X. Bai, D. Li, J. Zhu, Y. Guan, Q. Zhang, L. Chi, *Anal. Bioanal. Chem.* **2015**, *407*, 1857–1869.
- [30] E. G. Papanikolaou, H. Fatemi, M. Camus, D. Kyrou, N. P. Polyzos, P. Humaidan, B. Tarlatzis, P. Devroey, H. Tournaye, *Fertil. Steril.* **2010**, *94*, 2902–2904.
- [31] R. C. Anderson, C. L. Newton, R. A. Anderson, R. P. Millar, *Endocr. Rev.* **2018**, *39*, 911–937.
- [32] A. Gervais, Y. A. Hammel, S. Pelloux, P. Lepage, G. Baer, N. Carte, O. Sorokine, J. M. Strub, R. Koerner, E. Leize, A. Van Dorsselaer, *Glycobiology* **2003**, *13*, 179–189.
- [33] H. Zhu, C. Qiu, A. C. Ruth, D. A. Keire, H. Ye, *AAPS J.* **2017**, *19*, 846–855.
- [34] S. Perchepied, N. Eskenazi, C. Giangrande, J. Camperi, T. Fournier, J. Vinh, N. Delaunay, V. Pichon, *Talanta* **2020**, *206*, 120171.
- [35] H. Toll, P. Berger, A. Hofmann, A. Hildebrandt, H. Oberacher, H. P. Lenhof, C. G. Huber, *Electrophoresis* **2006**, *27*, 2734–2746.
- [36] J. Camperi, A. Combes, J. Guibourdenche, D. Guillarme, V. Pichon, T. Fournier, N. Delaunay, *J. Pharm. Biomed. Anal.* **2018**, *161*, 35–44.
- [37] J. Camperi, V. Pichon, T. Fournier, N. Delaunay, *J. Pharm. Biomed. Anal.* **2019**, *174*, 495–499.
- [38] J. Camperi, A. Combès, T. Fournier, V. Pichon, N. Delaunay, *Anal. Bioanal. Chem.* **2020**, DOI 10.1007/s00216-020-02684-8.
- [39] D. Thakur, T. Rejtar, B. L. Karger, N. J. Washburn, C. J. Bosques, N. S. Gunay, Z. Shriver, G. Venkataraman, *Anal. Chem.* **2009**, DOI 10.1021/ac901506p.
- [40] J. Camperi, B. De Cock, V. Pichon, A. Combes, J. Guibourdenche, T. Fournier, Y. Vander Heyden, D. Mangelings, N. Delaunay, *Talanta* **2019**, *193*, 77–86.
- [41] A. Al Matarí, A. Combès, J. Camperi, T. Fournier, V. Pichon, N. Delaunay, *Anal. Bioanal. Chem.* **2020**, 1–13.
- [42] W. Skala, T. Wohlschlager, S. Senn, G. E. Huber, C. G. Huber, *Anal. Chem.* **2018**, *90*, 5728–5736.
- [43] M. W. Senko, S. C. Beu, F. W. McLafferty, *J. Am. Soc. Mass Spectrom.* **1995**, *6*, 229–233.

Entry for the Table of Contents



Exploration of a highly glycosylated non-covalent protein complex by mass spectrometry is a challenging task due to isobaricity of the majority of glycoforms. Integration of data from multiple structural levels permits unraveling the hidden diversity of the human chorionic gonadotropin heterodimer, constituting the basis for the study of complex glycosylated protein assemblies.

Keywords: bioinformatics • biopharmaceutical • glycoprotein characterization • human chorionic gonadotropin • mass spectrometry