CrossMark

Research Article

A LC-MS All-in-One Workflow for Site-Specific Location, Identification and Quantification of N-/O- Glycosylation in Human Chorionic Gonadotropin Drug Products

Hongbin Zhu,[1] Chen Qiu,[1] Ashley C. Ruth,[1,2] David A. Keire,[1] and Hongping Ye[1,3]

Abstract. Site-specific characterization of the N- and O-linked glycosylation on a set of different human chorionic gonadotropin (hCG) drug products was performed by a LC-MS method combining high resolution (120K at *m/z* 200) mass spectrometry, multiple dissociation methods, tandem mass tag (TMT 10plex) labeling, and partial least squares-discriminant analysis (PLS-DA). In total, the data provided identification, relative quantification, and comparison of site-specific glycosylation of protein therapeutics with a single experiment. Ten different lots and/or brands of commercial therapeutic hCG were labeled with TMT 10plex reagents after tryptic digestion. The labeled intact glycopeptides were then analyzed by high resolution LC-MS with online alternating HCD/ETD/CID dissociation methods. For digested hCG drugs, 1000 intact N- and O-linked glycopeptides were identified. The relative amount of each glycopeptide from hCG products was determined based on the reporter signal intensities of the TMT labeling reagents. Moreover, with the help of TMT 10plex, through just one LC-MS run, PLS-DA was performed to ascertain the differences in glycosylation among different sources of hCG drug products. The results of PLS-DA showed that 167 glycopeptides were found to be significantly different between the naturally derived and recombinant hCG products. The results demonstrate the suitability of this method for similarity assessments and counterfeit identification of hCG as well as other glycoproteins.

KEY WORDS: glycopeptide; hCG; LC-MS; PLS-DA; protein therapeutic.

INTRODUCTION

Glycosylation is one of the most common and the most structurally diverse protein post-translational modifications (PTM) (1,2). Full characterization of glycoprotein drugs is essential, as even small changes in the glycosylation sites and/ or glycan types may dramatically affect the efficacy and safety of a glycosylated drug product (3–7). Of the available tools, mass spectrometry (MS) is the most widely used to characterize the glycosylation sites and glycoforms for glycoproteins.

One approach is bottom-up deglycosylated proteomics where the intact glycoprotein is digested to smaller glycopeptide parts and then analyzed by LC-MS/MS with the help of

hydrazide chemistry (8). The bottom-up proteomics workflow allows for direct analysis of the glycans, including their attachment sites in the protein(s) (9). In most cases, high resolution MS data and appropriate MS/MS data are necessary for confident glycopeptide identification.

For MS/MS fragmentation, the most commonly used dissociation techniques include collision-induced dissociation (CID), higher energy collisional dissociation (HCD), and electron transfer dissociation (ETD). These dissociation techniques are usually applied in separate experiments then the data is analyzed and the results combined to provide the majority of the information necessary for the complete identification of intact glycopeptides (10). For intact glycopeptides, CID can generate fragment ions from cleavage of glycosidic bonds that can be used to elucidate oligosaccharide sequence. In contrast to CID, ETD dissociation occurs at the peptide backbone rather than at glycosidic bonds, which preserves labile glycan-peptide bonds and allows the glycopeptide identification as well as the assignment of glycosylation sites (11–13). MS technology continues to develop and the advanced performance of newer mass spectrometers, including the quadrupole/linear ion trap/Orbitrap tribrid mass spectrometer (Q-OT-qIT, Orbitrap Fusion) used in this study,

---

[1] Division of Pharmaceutical Analysis, Center for Drug Evaluation and Research, U.S. Food and Drug Administration, 645 S. Newstead Ave, St. Louis, Missouri 63110, USA.
[2] *Present Address:* Biotechlogic Inc., Glenview, Illinois 60025, USA.
[3] To whom correspondence should be addressed. (e-mail: Hongping.Ye@fda.hhs.gov)

aaps

plays an important role for intact glycopeptide analysis (1,14–18).

An equally important and developing aspect of mass spectral data analysis is the use of multivariate statistical analysis of the complex data streams generated by glycoprotein MS data acquisition (19,20). For example, partial least squares-discriminant analysis (PLS-DA) multivariate analysis has been widely used in mass spectrometry-based -omics studies. The primary goal of PLS-DA is to identify class differences from a multivariate dataset (21,22). Usually, the intensities of each *m/z* (and, in some cases peak areas) from LC-MS data are used as variables. For proteomics studies, this kind of multivariate analysis is rarely used because the data file sizes increase by several orders of magnitude with LC-MS/MS analysis of peptide fractionations. The run-to-run variation of the LC-MS performance during data acquisition for multiple samples is also a concern.

As a therapeutic glycoprotein, human chorionic gonadotrophin (hCG, ~26 kDa) is used for multiple purposes including maturing human ova, triggering ovulation, treating cryptorchidism, and hypogonadotropism. Naturally derived urinary hCG (u-hCG), extracted from urine of pregnant women, has been used for more than 30 years for these purposes. To overcome some limitations of humans-sourced products, recombinant gonadotrophin (r-hCG), derived from genetically engineered Chinese hamster ovary cells through recombinant DNA technology, has been commercially available to patients for more than 10 years (23). The safety and efficacy of r-hCG compared to u-hCG have been studied clinically (23–27). However, there are some reported differences on clinical efficacy and applications (28,29). For example, the r-hCG has been reported to be more effective on inducing follicular maturation and early luteinization (23,28,30), whereas u-hCG has been recommended as a better choice for ovulation induction (23). hCG is a complex and highly glycosylated protein (29,31). The secretion, biological activity, and half-life of hCG are dependent on its glycosylation level and pattern. The different manufacturing processes of r-hCG and u-hCG lead to the differences of their glycan composition. The specific glycoforms of r-hCG and u-hCG can potentially affect their efficacy, stability, and safety. Therefore, a comprehensive study of the glycosylation difference between r-hCG and u-hCG is fundamental for understanding the differences in their clinical efficacy.

In this work, comprehensive glycosylation comparisons among different sources of hCG products with site-specific glycoforms identification and relative quantification are presented. Only a limited number of studies have examined the glycosylation of hCG. To our knowledge, there is no previous report for the comparison of the site-specific glycosylation of r-hCG and u-hCG. Furthermore, for the first time, PLS-DA and TMT 10plex labeling techniques are combined to map the differences of site-specific glycosylation that can be used for protein therapeutics assessment. By using tandem mass tag (TMT) reagents, multivariate analysis was performed in a unique and fast manner. The samples labeled with different tags were combined into one sample mixture. All variables (reporter ions) from the ten samples were acquired simultaneously in a single LC-MS run, so the LC-MS run time was reduced 10-fold. In addition, the peptide identification, quantification, and multivariate analysis were performed using data from a single LC-MS run (three LC-MS runs when considering technical replicates). Because the TMT labeling and sample combining steps are before fractionation of the peptides mixture, the sample size will be reduced ten times and the drift of LC-MS instrument conditions over time is less likely to affect the comparison between protein samples. Therefore, in this study, a comparative study of N- and O-linked glycosylation of glycoprotein therapeutics using isobaric labeling, multivariate data analysis, and LC-MS on a quadrupole/linear ion trap/orbitrap tribrid mass spectrometer is reported.

EXPERIMENTAL SECTION

Reagents and Materials

The following drugs were purchased from a pharmacy: three lots of Novarel® (u-hCG, Lot # 5H033A, 5C256A, and 5B134A, Ferring Pharmaceuticals Inc. Parsippany, NJ); three lots of Ovidrel® (r-hCG, Lot # BA023013, BA030650 and BA029711, EMD Serono, Inc. Rockland, MA); one lot of chorionic gonadotropin (u-hCG, Lot# 6108010, APP Pharmaceuticals, LLC. Schaumburg, IL); one lot of Pregnyl® (u-hCG, Lot# L034962, Merck & Co., Inc. Whitehouse Station, NJ). USP Human Chorionic Gonadotropin standard (u-hCG, Lot# 29700H) was purchased from USP Reference Standards (Rockville, MD). The Tandem Mass Tag 10-plex reagent kit, Optima formic acid, and Optima LC-MS solvents were purchased from Fisher Scientific (Pittsburgh, PA). Mass spectrometry-grade trypsin was purchased from Promega (Madison, WI).

Sample Preparation

Desalting. The drug products were dissolved in LC-MS grade water, desalted using Amicon Ultra-0.5 mL Centrifugal Filters (3K, EMD Millipore Corporation, Billerica, MA), quantified using bicinchoninic acid assay (BCA assay kit, Pierce Biotechnology, Rockford, IL), and aliquoted to 80 μg of protein per tube. All sample tubes were stored at −20°C after freeze dried until digestion and labeling.

Digestion. Samples were digested and labeled in triplicates using our optimized protocol as previously described (32) with some minor modification. Briefly, 20 μL of 0.5 M triethylammonium bicarbonate (TEAB) and 1 μL of denaturant (2% sodium dodecyl sulfate, SDS) were added to each sample tube containing 80 μg of protein; 2 μL of tris(2-carboxyethyl)phosphine hydrochloride (TCEP, 50 mM) was added as reducing regent at 37°C (1 h); after reduction, 0.5 μL of 200 mM iodoacetamide (IAA) was added at room temperature in the dark for 30 min to block the cysteine residues. After the samples were diluted to 2 μg/μL with 0.5 M TEAB, Promega Gold® trypsin was added at the ratio of 1:40 overnight (37°C, 12 to 16 h) to digest the samples. Immediately before use, TMT reagents were equilibrated to room temperature and dissolved in 70 μL of absolute ethanol. The labeling reactions were quenched after 55 min at room temperature by adding 8 μL of an aqueous solution of 5%

hydroxylamine and incubating for 15 min. The labeled samples were combined, vacuum concentrated for 1 h, and then freeze dried. Before LC-MS/MS analysis, the samples were dissolved with 150 μL of an aqueous solution containing 5% ACN/5% formic acid.

Ten samples, including nine different lots of drug products and a different package from the same Lot# BA029711 (r-hCG), were digested and labeled. Lot # BA023013 and BA030650 (r-hCG) were labeled with TMT 128N and 128C separately. Two different packages from Lot# BA029711 (r-hCG) were labeled with TMT 129N and 129C separately. Lot # 5H033A, 5C256A, 5B134A, 6108010, L034962, and 29700H (u-hCG) were labeled with TMT 126, 127C, 127N, 130N, 130C, and 131 separately.

Liquid Chromatography-Mass Spectrometry

An Orbitrap Fusion™ Tribrid™ mass spectrometer (Thermo Scientific, Bremen, Germany) coupled with a Thermo Scientific™ Dionex™ Ultimate™ 3000 UHPLC instrument was used for LC-MS/MS analysis. Isobarically labeled digests were separated with a Waters SunFire C18 column (2.1 × 150 mm, 3.5 μm). Gradient elution began with 5% acetonitrile containing 0.1% formic acid for 7 min at 0.2 mL/min (to waste), then increased to 30% acetonitrile over 163 min with a flow rate of 0.1 mL/min (to MS), then immediately increased to 95% acetonitrile and held for 10 min to wash the column before column re-equilibration at 5% acetonitrile for the next injection.

The relevant parameters used for MS and MS/MS data acquisition were as follows: top speed mode with 3 s cycle time; FTMS: scan range (m/z) = 380–2000; resolution = 120 K; AGC target = 2.0e5; maximum injection time = 50 ms; Filter: precursor selection range = 380–1800; include charge state = 2–8; dynamic exclude after n times = 2; Decision: precursor priority = highest charge state then most intense; $FTMS^2$ (HCD): isolation mode = quadrupole; isolation window = 2; collision energy (%) = 40; resolution = 30 K; AGC target = 2.0e5; maximum injection time = 250 ms; microscan = 1; $ITMS^2$ (CID): isolation mode = quadrupole; isolation window = 2; collision energy (%) = 30; AGC target = 1.0e4; maximum injection time = 150 ms; microscan = 1; $ITMS^2$ (ETD): isolation mode = quadrupole; isolation window = 2; use calibrated charge dependent ETD parameters = false; ETD reaction time = 60 ms; AGC target = 1.0e4; maximum injection time = 150 ms; microscan = 1.

Data Processing

Samples were prepared in triplicate and each replicate was injected three times for LC-MS analysis. Three raw files from the same replicate were combined and used for database searches using Byonic™ (version 2.3.1) and Proteome Discoverer (version 2.1). N- and O-glycosylation were searched separately using Byonic nodes for Proteome Discoverer with the following search parameters: precursor mass tolerance = 10 ppm, fragmentation type = both: HCD&ETD, fragment mass tolerance (HCD) = 40 ppm, fragment mass tolerance (ETD) = 0.7 Da; modifications: TMT6plex/ +229.162932 @ K and N-term (fixed), Carbamidomethyl/ +57.021464 @ C (fixed), Oxidation/+15.994915 @ M (common 1), Dehydrated/-18.010565 @S, T (Rare 2), (De) Carbamidomethyl/-57.021464 @ C (rare 1), (De)TMT6plex/- 229.162932 @ N-term (rare 1), N-glycan (common 1) or O- glycan (common 3) (the glycan database provided by Byonic and filtered from the literature were both used for glycopeptides search). Spectrum-level false discovery rate (FDR) was set as 1% cut and peptide spectra matches (PSMs) with manually score cut at 300 were used.

The N-linked glycopeptides identified from Byonic were verified manually. Thermo Scientific's Xtract software was used to convert the raw data into singly charged monoisotopic spectra for interpretation of protonated molecular ions. For quantitative analysis, ten different hCG samples labeled with different TMT tags were prepared in triplicates and combined to three samples, one for each set of TMT 10plex, for LC-MS/ MS analysis. The signal to noise (S/N) ratio of each TMT Reagent reporter ion mass was extracted using Proteome Discoverer with a mass tolerance of 10.0 ppm.

Multivariate analysis was performed using the PLS Toolbox (v.7.3.1, Eigenvector Research, Wenatchee, WA) run in MATLAB (R2013a, MathWorks Inc., Natick, MA). The S/N of all TMT 10plex reporter ions for glycopeptides with Byonic Score>300 were extracted by Proteome Discoverer. The S/N data were imported into MATLAB software for PCA and PLS-DA. The loadings plots from PLS-DA were used to target the important variables (glycopeptides) which could be used as markers for different sources of hCG products. Data preprocessing includes standard normal variate (SNV) and mean centering. For PLS-DA, the pre-class of the ten samples was based on the PCA classification results (Information about PCA can be found in Fig. S1). S/N of TMT reporter channels (126, 127N, 127C) were set as class 1 which were different lots of the same brand (u-hCG). S/N of TMT reporter channels (130N, 130C, and 131) were set as class 2 which were from other three different brands of u- hCG. S/N of TMT reporter channels (128N, 128C, 129N, and 129C) were set as class 3 which were all from r-hCG. Venetian blinds method was used for cross-validation (max- imum number of LVs, 10; number of data splits, 5; samples per blind, 1). Two latent variables (LVs) were selected (92.22% coverage) to build the score plot and loading plot.

RESULTS AND DISCUSSION

Identification of Intact N-Linked Glycopeptides by ETD, HCD, and CID

Most of the glycopeptides were present at very low abundances relative to the peptides without glycosylation in the tryptic peptide mixture. In order to obtain the maximum signal for glycopeptide identification, the LC, full MS, and MS/MS settings, including LC gradient, AGC target, maxi- mum injection time, collision energy (HCD MS/MS), and ETD reagent reaction times, were optimized. Collision energy (HCD MS/MS) and ETD reagent reaction times were found to have the most influence on the number of identified TMT labeled glycopeptides. For HCD MS/MS fragmentation, higher collision energy (40%) was required for TMT labeled glycopeptides compared to non-glycopeptides/non-labeled

peptides. For ETD MS/MS fragmentation of TMT labeled glycopeptides, the optimized ETD reaction time was 60 ms, which provided significant improvement relative to default manufacturer calibration settings. Generally, long gradient and/or off-line fragmentation is used in the analysis of peptides mixtures from multiple digested proteins (proteomics). In this study, a single highly purified protein was being analyzed (hCG α chain and β chain) and the long gradient was used to separate the peptides from the glycopeptides online. The optimized HPLC gradient was 180 min at a flow rate of 0.1 mL/min. The protein digestion and TMT labeling protocols and instrumentation methods were optimized for analysis of intact glycopeptides. Protein sequence coverage was determined to be 100% for both α-chain and β-chain.

The intact N-linked glycopeptides were identified by combining the match of accurate mass acquired from high resolution full MS scan (120K) and HCD MS/MS or ETD MS/MS Peptide

Spectra Match (PSMs). The ETD settings were optimized for intact glycopeptides analysis. The ETD reaction times were shown to be the most important parameter that could affect the quality of ETD spectra. Instead of using reaction time from "charge-state-specific calibration settings," a time of "50 ms" was found to be optimal for intact glycopeptides which could achieve the best identification and ETD spectra on the spectrometer used in this work. In addition, because precursors with higher charge states could have better ETD fragmentation, the precursors with the highest charge states were selected as the priority for MS2 analysis. Furthermore, TMT labeling increased the average precursor charge for glycopeptides and made them more amenable to ETD and resulted in better spectral signal to noise. Figure 1 shows the MS/MS spectra of a N-linked glycopeptide (K.[+TMT]N(HexNAc$_3$Hex$_4$NeuAc$_1$)VTSESTCCVAK[+TMT].S) acquired with multiple dissociation techniques. The peptide sequence was identified by ETD and HCD MS/MS spectra. As
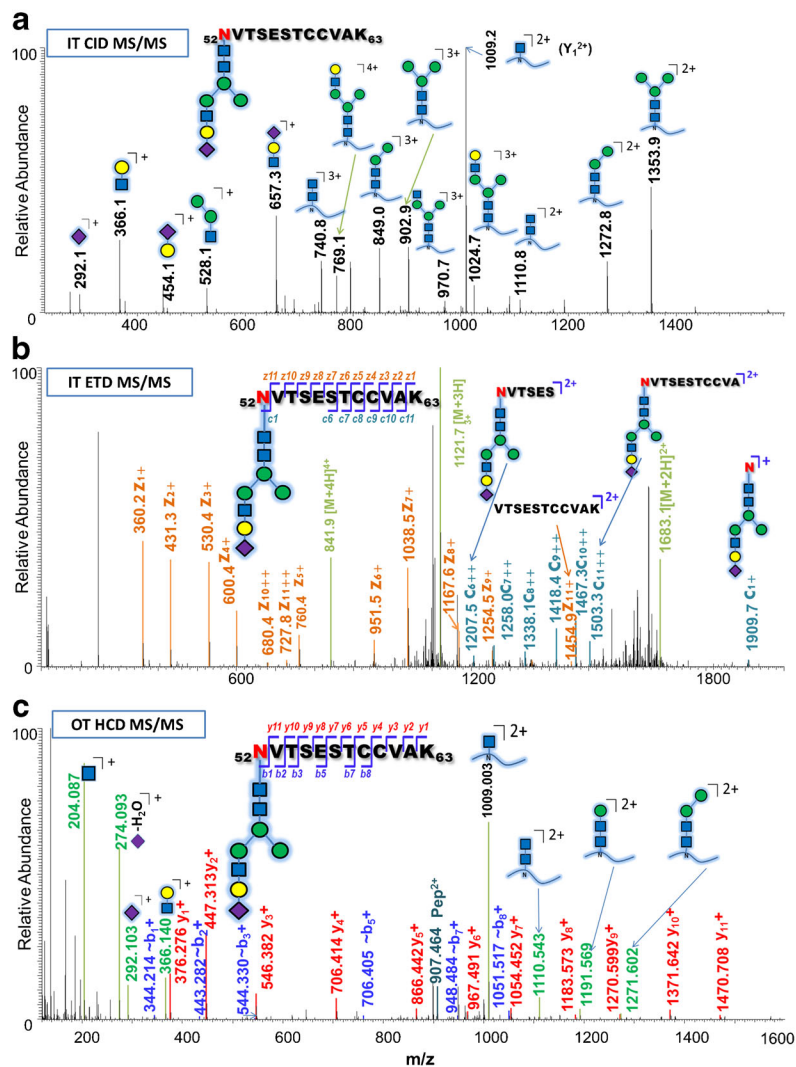

Fig. 1. The MS/MS spectra from multiple dissociation methods (CID, ETD, and HCD) of a N-linked glycopeptide (4+, *m/z* 841.3714 calculated and *m/z* 841.3722 observed, 0.9 ppm). These data confirmed a sialic acid containing N-linked glycan structure, HexNAc$_3$Hex$_4$NeuAc$_1$. *Yellow circle* = galactose (Hex); *blue square* = N-acetylglucosamine (HexNAc); *blue circle* = mannose (Hex); *purple diamond* = N-acetylneuraminic acid (NeuAc). *IT* = ion trap (low resolution), *OT* = Orbitrap (high resolution)

shown in Fig. 1b, c, there were a large number of c/z and b/y ions for peptide backbone identification. With ETD technology, the labile glycan was retained without signs of carbohydrate cleavage as shown in Fig. 1b, while the peptide backbone was cleaved. For this peptide, with the help of c ions and z ions, especially c_1^+ (observed m/z 1909.8, calculated m/z 1909.7, $\Delta m/z = 0.1$ Da) and z_{11}^+ (observed m/z 1454.9, calculated m/z 1454.7, $\Delta m/z = 0.2$ Da), the glycosylation site was clearly determined at position Asn52 and peptide sequence was also identified with series of c/z ions.

Figure 1c is the HCD spectrum of this glycopeptide. HCD experiments generated ions from both amide bond breaking and glycosidic bond cleavage. Therefore, N-linked glycopeptide was also identified by HCD fragmentation with high confidence with the help of accurate mass. Some free glycan fragments and glycopeptide fragments were very useful for N-linked glycopeptide confirmation. Low-energy CID spectra (Fig. 1a) give little or no peptide backbone fragmentation, but the abundant glycosidic bond cleavage information is helpful for glycan structure confirmation. As shown in Fig. 1a, the doubly charged Y_1^{2+} ion (observed m/z 1009.2, calculated m/z 1009.0, $\Delta = 0.2$ Da) was the base peak in the CID spectrum, which indicated that the glycan was fragmented efficiently. There were very few fragments (b/y ions at very low abundance) from peptide backbone cleavage with CID. Most of the high abundant peaks in the CID spectrum were assigned to the glycan fragments or glycopeptides fragments with glycosidic bond cleavage only. With the help of these numerous matched glycan fragments, this glycan structure ($HexNAc_3Hex_4NeuAc_1$) could be confirmed unambiguously.

In this study, online alternating HCD/ETD/CID scans were performed for all peptides. Both HCD and ETD PSMs were used for N-linked glycopeptides identification, while CID spectra were used for glycan structure confirmation. Combination of peptide fragmentation information from the ETD and HCD spectra provided the maximum number of identified intact N-linked glycopeptides. As shown in Fig. 2a, when using a peptide score of $\geq$300 for cutting (Byonic), a total of 755 unique N-linked glycopeptides were identified from ETD and HCD spectra. Overall, for this protein, 34% more glycopeptides were identified by HCD/ETD compared to using HCD alone.

The charge state distribution and $[M+H]^+$ distribution of the N-linked glycopeptides identified by HCD and ETD are shown in Fig. 2b, c. The results indicated that for intact N-linked glycopeptide identification, HCD and ETD were complementary. Most of the N-linked glycopeptides had mass of 4000–6000 and charge state of 4+ or 5+. More glycopeptides were identified with HCD than ETD for lower charge peptides (3+, 4+), while for glycopeptides having higher charge state (5+, 6+, 7+, and 8+), ETD identified more glycopeptides than HCD. This agreed with the fact that the charge state of precursor ion can affect the dissociation efficiency of ETD. Because almost all of the intact tryptic glycopeptides are multiple charged ($\geq$3+) when analyzed under normal LC-ESI-MS (m/z 300–2000) conditions, ETD can identify more N-linked glycopeptides and is more suitable for intact N-linked glycopeptides identification. However, because of the greater diversity of N-linked glycan structures and the size of N-linked glycans, their mass could be hundreds to thousands and the glycopeptides have very broad charge state distribution. Therefore, HCD can be used to complement ETD dissociation to improve glycopeptide identification as well as peptide sequence coverage.
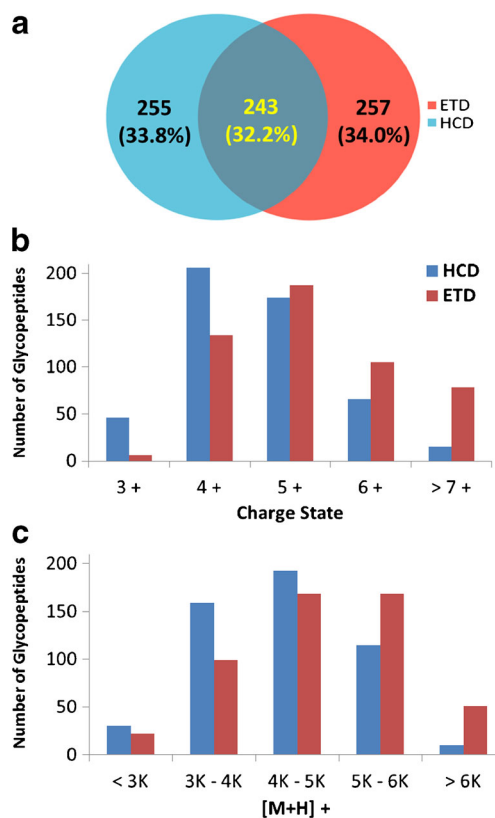
Fig. 2. a Number of N-linked glycopeptides identified by HCD and ETD; **b** the charge state distributions of the N-linked glycopeptides identified by HCD and ETD, **c** the $[M+H]^+$ distribution of N-linked glycopeptides identified by HCD and ETD

As shown in Table I (b), most of the identified N-linked glycosylation sites are close to the N-termini of the tryptic peptides. Due to the steric hindrance from the large N-linked glycans, the TMT labeling efficiency for N-termini (75%) was relatively lower than lysine residue (92%). To map the potential glycans, mislabeled N ((De) TMT) as well as other common peptide modifications, such as deamidation and oxidation, were included in database searches using Byonic. With these peptide modifications and N-linked glycosylation considered, a total of 755 unique N-linked glycopeptides were identified with a good Byonic score ($\geq$300). When only glycosylation was considered (i.e., no other modifications), 332 unique site-specific N-linked glycopeptides were identified (Supporting Information_2.xlsx). While ignoring the site of modification information, there were 165 unique N-glycans identified in these hCG products. These results indicated that the glycosylation of hCG was highly heterogeneous.

As shown in Fig. 3, the types of glycans identified at the four N-glycosylation sites were found to be very different. More N-linked glycans were identified on the α-chain than on the β-chain and only half of these glycans were conserved between the two chains of hCG (Fig. 3a, d). Site Asn78 on α-chain of hCG had the most N-linked glycans identified (133 and 119 for r-hCG and u-hCG, respectively), about twice more than any of the other three sites (Fig. 3b, e). The heterogeneity of the glycosylation observed indicated the importance of developing methods for the site-specific characterization of glycoprotein therapeutics so the variability in these products can be measured. Glycans at different sites may play different biological roles and may affect the quality of glycoprotein products.

Table I. Number of N-Linked Glycopeptides and Glycans Identified

(a)

	Total	In u-hCG	In r-hCG
Unique N-linked glycopeptides with other modifications	755	649	729
Unique site-specific glycopeptides	332	290	323
Unique glycans	165	149	163

(b)

Site in protein	Peptide sequence	Number of site-specific glycopeptides		
		Total	In u-hCG	In r-hCG
αN52	N_{52}VTSESTCCVAK	72	60	71
αN78	VEN$_{78}$HTACHCSTCYYHK	134	119	133
βN13	PIN$_{13}$ATLAVEK	72	65	69
βN30	EGCPVCITVN$_{30}$TTICAGYCPTMTR	54	46	50

Investigation of glycoproteins as intact glycopeptides is essential to obtaining the required site-specific information.

Identification of O-Linked Glycopeptides by ETD and CID

Because there is not an O-linked amino acid consensus sequence, O-glycosylation was much more complicated than N-glycosylation and HCD could not locate the exact position of O-glycan, only ETD and CID were used for O-linked glycopeptide identification. The reported O-glycosylation sites of the hCG forms are Ser-121, Ser-127, Ser-132, and Ser-138, all of which are located on the β-chain. The amino acid sequence of the β-chain from 113 to 145 is "PRF QDSSS S$_{121}$KAPP PS$_{127}$LPS PS$_{132}$RLP GPS$_{138}$DT PILPQ." This β-chain sequence yields three tryptic peptides when fully cleaved and 9 of the 33 amino acids can be potential O-linked glycosylation sites.

In this case, HCD did not provide enough information for determination of O-glycosylation sites. By contrast, ETD provided glycopeptide sequence and unambiguous assignment of the site of glycosylation. O-linked glycopeptides were identified by ETD when aided with the glycan structure confirmation provided by abundant glycosidic bond cleavage signals obtained by CID spectra. HCD MS/MS spectra were also acquired for O-linked glycopeptides for the purpose of quantification.

Figure 4 shows the ETD and CID MS/MS spectra of an O-linked glycopeptide (R.[+TMT]FQDSSSS(HexNAc$_1$Hex$_1$NeuAc$_1$)K[+TMT].A). The peptide backbone was identified by ETD MS/MS spectra, as shown in Fig. 4a. The O-glycosylation site was assigned at position Ser121 by rich c/z ions from the ETD spectrum, especially z_1^+(observed m/z 360.3, calculated m/z 360.3, $\Delta = 0$ Da), z_2^+(observed m/z 1103.3, calculated m/z 1103.5, $\Delta = 0.2$ Da), c_6^+ (observed m/z 898.3, calculated m/z 898.4, $\Delta = 0.1$ Da), and c_7^+ (observed
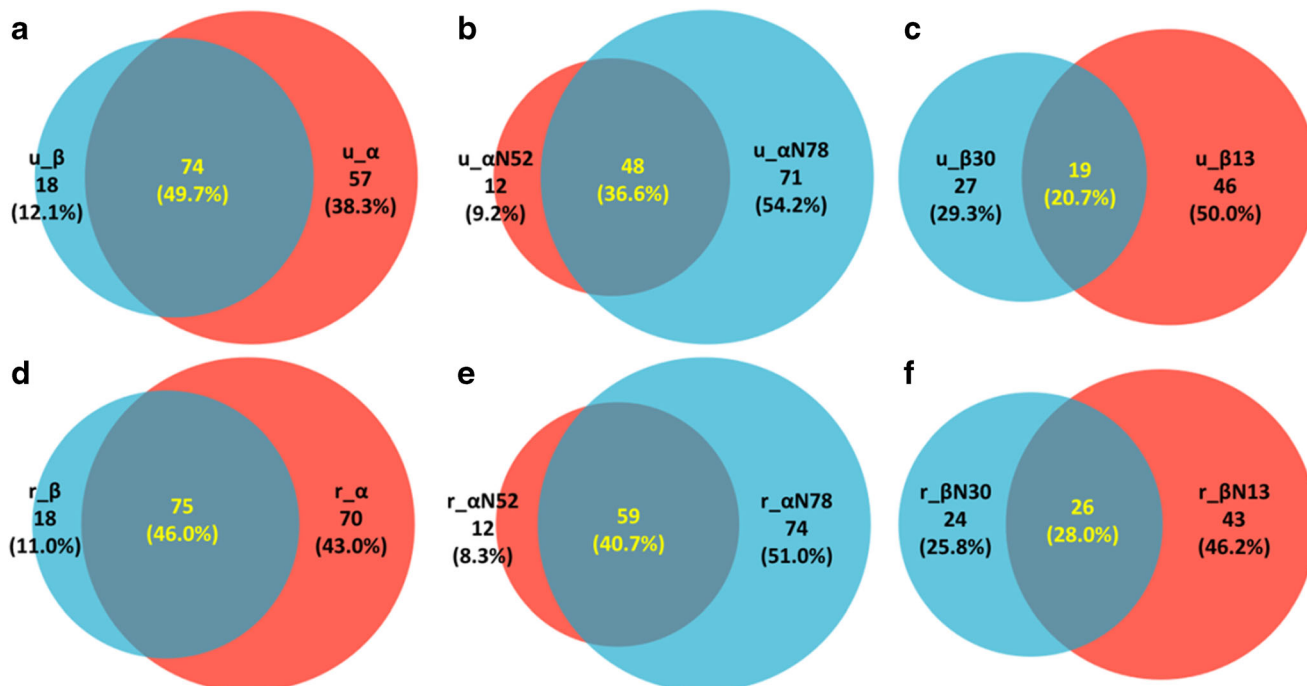

Fig. 3. The unique site-specific glycans comparison between different N-linked glycosylation sites. 'r' means r-hCG, 'u' means u-hCG
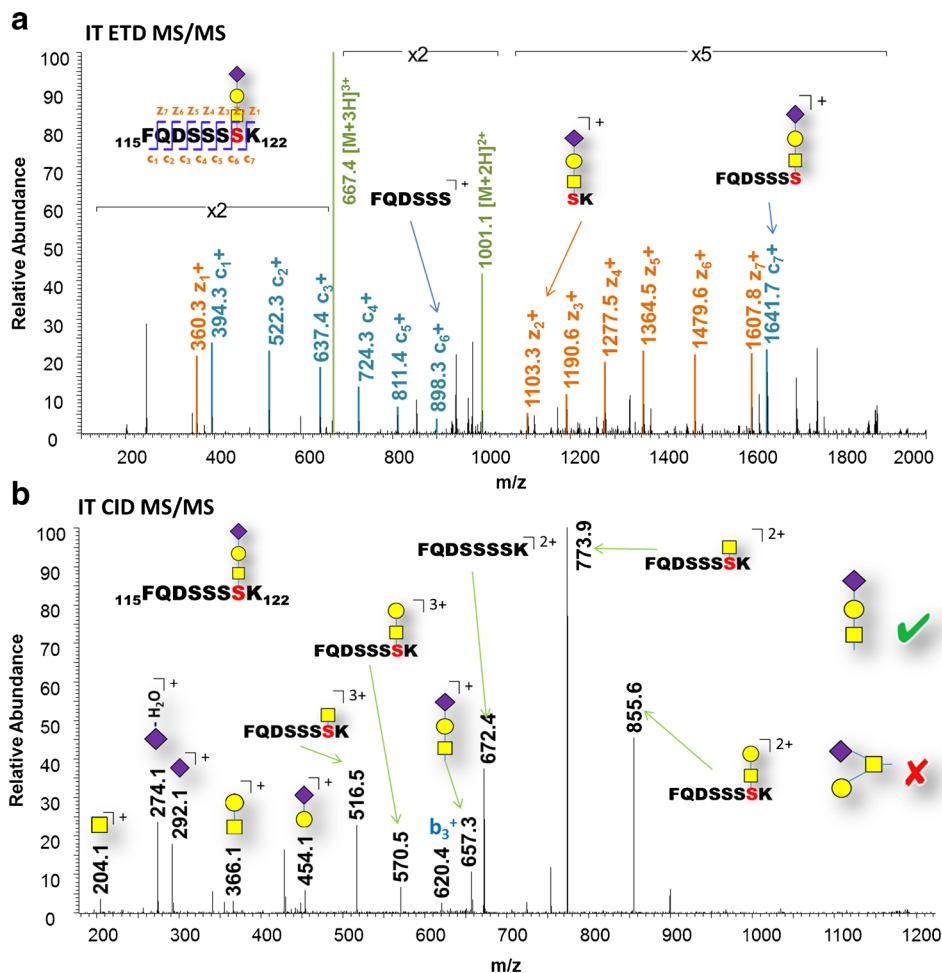
Fig. 4. The MS/MS spectra from multiple dissociation techniques (CID, ETD) of an O-linked glycopeptide (3+, *m/z* 667.3210 calculated and *m/z* 667.3214 observed, 0.5 ppm). These fragments confirmed a sialic acid containing O-linked glycan structure, $HexNAc_1Hex_1NeuAc_1$, at position Ser121. *Yellow square* = N-acetylgalactosamine (HexNAc), *yellow circle* = galactose (Hex); *purple diamond* = N-acetylneuraminic acid (NeuAc). *IT* = ion trap (low resolution)

m/z 1641.7, calculated *m/z* 1641.7, $\Delta = 0$ Da). The composition of this O-glycan identified by high resolution MS (by intact glycopeptide exact mass match, observed *m/z* 667.3214, calculated *m/z* 667.3210, $\Delta = 0.5$ ppm) and ETD MS/MS was $HexNAc_1Hex_1NeuAc_1$, with two possibilities of the glycan structure shown on the right side of Fig. 4b. The CID spectrum (Fig. 4b) gave minimal peptide sequence information, but abundant glycosidic bond cleavage information was helpful for glycan structure confirmation. From the CID MS/MS spectrum, the composition of the glycan and the mass of the peptide moiety could be deduced. With the help of the glycopeptide fragments at *m/z* 855.6 (calculated *m/z* 855.4, $\Delta = -0.2$ Da, $[Pep+HexNAc+Hex+2H]^{2+}$), *m/z* 570.5 (calculated *m/z* 570.6, $\Delta = 0.1$ Da, $[Pep+HexNAc+Hex+3H]^{3+}$) and glycan fragment at *m/z* 454.1 (calculated *m/z* 454.2, $\Delta = 0.1$ Da, $[Hex+NeuAc+H]^+$), the right glycan structure (shown on the top-right of the Fig. 4b) was confirmed unambiguously. Through the combination of ETD and CID, a total of 245 O-linked glycopeptides were identified. When only considering glycosylation and ignoring other modifications, 237 unique site-specific O-linked glycopeptides were found (Supporting Information_2.xlsx). While only

considering the types of glycans by ignoring the site information, there were 98 unique O-glycans identified in these hCG products.

Mapping the Glycosylation Differences Between u-hCG and r-hCG by PLS-DA and TMT 10-plex

A major aim of this work was to determine whether the present method is able to map the differences in glycosylation between u-hCG and r-hCG, which might be helpful for understanding the different biological activities of these two kinds of hCG. The isobaric labeling technique was useful for this purpose. In addition, multiplexing with 10plex allowed the comparative quantification of up to ten products (different brands or lots) simultaneously, saving time on sample preparation and the time demand on the high-end mass spectrometers. However, due to the large number of N- and O-linked glycopeptides (total of 1000) identified in this study, comparing the 10 TMT reporter ions for each peptide would be prohibitively time consuming. In addition, when the quantification data was analyzed manually, the glycosylation of the six urine products varied considerably from brand to

brand, which made the overall comparative quantification between u-hCG and r-hCG groups even more challenging.

PLS-DA is usually performed to target the variables which carry the class separating information. Here, PLS-DA was successfully applied to pinpoint the glycopeptides which were obviously different among different hCG products. Figure 5 shows the PLS-DA results obtained through ten TMT reporter ions from HCD spectra of 1000 glycopeptides. The score plot (Fig. 5a) shows a clear clustering of samples correlated with the sources of hCG products. In the first latent variables (LV1) direction, the hCGs are clustered into two groups, which represent r-hCG and u-hCG. In the second latent variables (LV2) direction, the u-hCG products are further clustered into two groups correlated with different brands. In r-hCG group, 128N, 128C, and 129N were from three different lots of the same brand, while 129N and 129C were from different packages of the same lot. In u-hCG group, 126, 127C, and 127N were from three different lots of the same brand, 130C and 130N were from two additional brands, and 131 was USP standard. Since the LV1 contributes to most of the separations (70.15%), the results indicate that urine-derived products from different manufacturers have a similar glycosylation patterns which differs from the recombinant products.

Variables with loadings at a given position in a loading plot contribute heavily to observations whose scores are found in a similar position in a score plot. The loading plot (Fig. 5b) shows these differences with the position of the dots (peptides). Variable Importance in Projection (VIP) scores estimate the importance of each variable in the projection used in PLS-DA and are often used for variable selection. In this study, dots (glycopeptides) in red were selected with a VIP score greater than 1.0. In LV1 direction, the red dots on the right side are the glycopeptides that have higher abundance in r-hCG than in u-hCG, while those on left side represent the glycopeptides that have higher abundance in u-hCG than in r-hCG. In LV2 direction, the red dots over the center of the black dots are the glycopeptides that have higher abundance in brand Novarel® than in other three u-hCG brands, while those dots below the black dots represent the glycopeptides that have lower abundance in brand Novarel® than in other three u-hCG brands. The composition of these glycopeptides and their relative TMT reporter ion intensities (S/N) are listed in Table S1. The abundance of TMT reporter ions of three selected examples are shown in Fig. 5c. Pep1 (Fig. 5c, Table S1) is an N-linked glycopeptide which has the biggest difference in abundance between r- and u-hCG products (higher abundance in r-hCG), with the
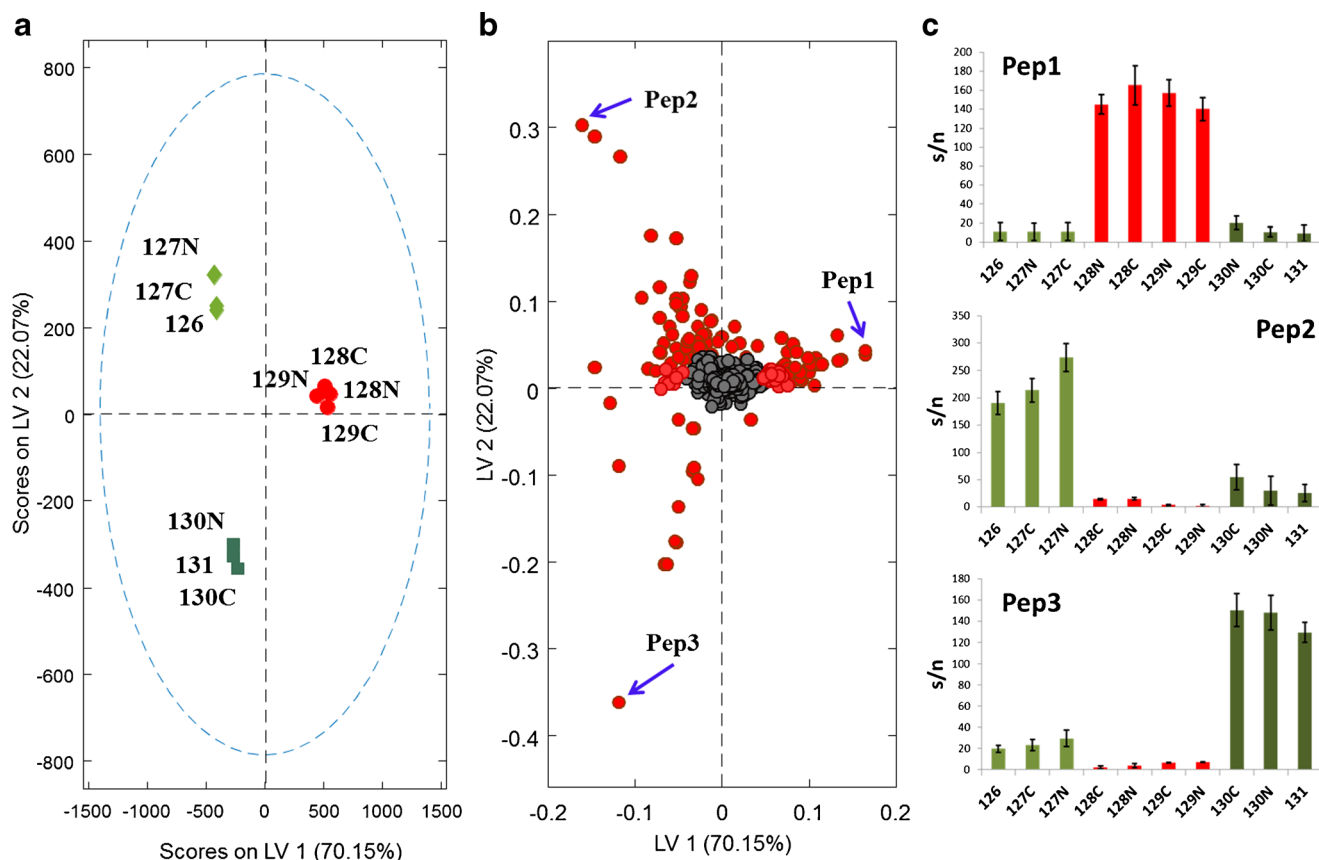

Fig. 5. PLS-DA results obtained from the S/N of TMT 10plex reporter ions. **a** Score plot shows the classification of hCG from different sources. **b** Loading plot shows the contribution of each glycopeptide to classification of hCG. The *dots* (glycopeptides) in *red* are selected with a VIP score greater than 1.0, which represent the glycopeptides that have obviously different abundance either between r- and u-hCG or between different brands of u-hCG. *Black dots* (VIP score lower than 1.0) in the middle of the plot are the glycopeptides that do not have much difference in abundance in these hCG products. **c** The S/N and quantitative variations of the three extreme glycopeptides (*Pep1*, N[+HexNAc$_3$Hex$_4$]VTSESTCCVAK; *Pep2*, N[+HexNAc]VTSESTCCVAK; *Pep3*, CRPIN[+HexNAc$_2$Hex$_2$]ATLAVEK)

average ratio of about 13:1 (r-hCG/u-hCG). Pep2 and Pep3 (Fig. 5c, Table S1) are two glycopeptides that have the most differences in abundance (higher abundance in u-hCG) with the average u-hCG to r-hCG ratios of 12:1 and 16:1, respectively. Pep2 and Pep3 are also the glycopeptides that have the biggest differences between the two u-hCG groups. The error bars in Fig. 5c represent standard deviations of three replicated measurements. For these three glycopeptides, the average RSD% is blow 10% for products with high abundance. For example, Pep1 has high abundance in r-hCG products with RSD% of 7.2, 12.4, 8.7, and 8.6% (average 9.2%) for 128N, 128C, 129N, and 129C, respectively. The lot-to-lot and brand-to-brand variations can be seen from Fig. 5c. For a particular glycopeptide, the S/N ratios of different lots from the same brand are very close (126, 127N, and 127C in Novarel®; 128N, 128C, 129N, and 129C in Ovidrel®). The lot-to-lot variations from the same brand can also be seen in Fig. 5a. The four red dots represent four Ovidrel® samples (three different lots). These four stacked dots in the score plot indicate that the glycosylation type and level among these four samples are very similar. The three diamond dots (126, 127N, 127N) represent three different lots of Novarel® which are very close and only slightly different in the LV2 direction. The results from PLS-DA and TMT quantification are consistent and indicate that the glycosylation of the products from different lots of same brand is relatively stable.

Using TMT labeling, the site-specific quantitative comparison of the glycopeptides among these hCG products was investigated based on the total intensity of the isotopic mass tags observed from HCD spectra (PSMs). The site-specific predominant glycopeptides were investigated using total reporter ion intensities and results can be found in Table S2. HexNAc$_4$Hex$_5$NeuAc$_1$ at Asn52 and HexNAc$_4$Hex$_5$NeuAc$_2$ at Asn78 (α-chain) were found to be the predominant N-linked glycopeptides in r-hCG and u-hCG products, respectively. The relative quantification of site-specific O-linked glycan analysis (Table S3) showed that HexNAc$_1$Hex$_1$NeuAc$_1$ at Ser121 was the predominant O-linked glycopeptide in both r-hCG and u-hCG products.

CONCLUSIONS

Here, a new method for identification, relative quantification, and comparison of the site-specific glycosylation of protein therapeutics within a single experiment is reported. To our knowledge, this is the first time a combination of TMT 10plex and PLS-DA have been used to make site-specific glycosylation comparisons that can be used as a framework for similarity assessment. Conducting PLS-DA analysis using the intensity information from TMT 10plex reporter ions saved data acquisition time and minimized run-to-run variability. The results obtained demonstrate the benefit of multiple dissociation methods for the identification and quantification of intact glycopeptides. The combination of ETD and HCD provided the maximum number of identified intact N-linked glycopeptides. CID was shown to be necessary and useful to elucidate glycan structures. This work supports the view that ETD has many advantages for defining glycan structures and glycosylation sites.

In this study, ten hCG drug samples were labeled with ten different TMT tags and analyzed in just one LC-MS run

(with three replicates). A total of 755 unique N-linked glycopeptides and 245 O-linked glycopeptides were identified. Without considering other modifications, a total of 332 unique site-specific N-linked glycopeptides and 237 unique site-specific O-linked glycopeptides were identified. Most of the high abundant glycans were observed to contain sialic acid, which was in agreement with previously reported results (29,33–35). Results showed that r-hCG and u-hCG products had distinct and conserved glycans. Through PLS-DA analysis, 167 (about 17%, see Table S1 (a) and Table S1 (b)) of the identified glycopeptides were found to have obvious differences in abundance between u-hCG and r-hCG. These glycosylation differences may contribute to their different clinical activities (23,28,30). This method can also be applied to any larger number of samples by setting one TMT channel as a reference standard. For example, up to 19 samples can be analyzed in two sets and with two LC-MS runs or 28 samples in three LC-MS runs. These observations demonstrate the suitability of this method for comparability determination or counterfeit identification of hCG and other glycoproteins.

COMPLIANCE WITH ETHICAL STANDARDS

Disclaimer This publication reflects the views of the authors and should not be construed to represent FDA's views or policies.

REFERENCES

1. Woo CM, Iavarone AT, Spiciarich DR, Palaniappan KK, Bertozzi CR. Isotope-targeted glycoproteomics (IsoTaG): a mass-independent platform for intact N- and O-glycopeptide discovery and analysis. Nat Methods. 2015;12(6):561–7.
2. Pan S, Chen R, Aebersold R, Brentnall TA. Mass spectrometry based glycoproteomics—from a proteomics perspective. Mol Cell Proteomics. 2011;10(1):R110.003251.
3. Jefferis R. Glycosylation as a strategy to improve antibody-based therapeutics. Nat Rev Drug Discov. 2009;8(3):226–34.
4. Carter PJ. Potent antibody therapeutics by design. Nat Rev Immunol. 2006;6(5):343–57.
5. Jefferis R. Glycosylation of recombinant antibody therapeutics. Biotechnol Prog. 2005;21(1):11–6.
6. Walsh G, Jefferis R. Post-translational modifications in the context of therapeutic proteins. Nat Biotechnol. 2006;24(10):1241–52.
7. Sinclair AM, Elliott S. Glycoengineering: the effect of glycosylation on the properties of therapeutic proteins. J Pharm Sci. 2005;94(8):1626–35.
8. Zhang H, Li X-J, Martin DB, Aebersold R. Identification and quantification of N-linked glycoproteins using hydrazide chemistry, stable isotope labeling and mass spectrometry. Nat Biotechnol. 2003;21(6):660–6.
9. Zhu Z, Su X, Go EP, Desaire H. New glycoproteomics software, GlycoPep evaluator, generates decoy glycopeptides de novo and enables accurate false discovery rate analysis for small data sets. Anal Chem. 2014;86(18):9212–9.
10. Scott NE, Parker BL, Connolly AM, *et al.* Simultaneous glycan-peptide characterization using hydrophilic interaction chromatography and parallel fragmentation by CID, higher energy collisional dissociation, and electron transfer dissociation MS applied to the N-linked glycoproteome of *Campylobacter jejuni.* Mol Cell Proteomics. 2011;10(2):M000031–MCP201.

11. Windwarder M, Altmann F. Site-specific analysis of the O-glycosylation of bovine fetuin by electron-transfer dissociation mass spectrometry. J Proteome. 2014;108:258–68.

12. Ye H, Boyne MT, Buhse LF, Hill J. Direct approach for qualitative and quantitative characterization of glycoproteins using tandem mass tags and an LTQ Orbitrap XL electron transfer dissociation hybrid mass spectrometer. Anal Chem. 2013;85(3):1531–9.

13. Håkansson K, Cooper HJ, Emmett MR, Costello CE, Marshall AG, Nilsson CL. Electron capture dissociation and infrared multiphoton dissociation MS/MS of an N-glycosylated tryptic peptide to yield complementary sequence information. Anal Chem. 2001;73(18):4530–6.

14. Richards AL, Hebert AS, Ulbrich A, et al. One-hour proteome analysis in yeast. Nat Protoc. 2015;10(5):701–14.

15. Senko MW, Remes PM, Canterbury JD, et al. Novel parallelized quadrupole/linear ion trap/Orbitrap tribrid mass spectrometer improving proteome coverage and peptide identification rates. Anal Chem. 2013;85(24):11710–4.

16. Wu S-W, Pu T-H, Viner R, Khoo K-H. Novel LC-MS2 product dependent parallel data acquisition function and data analysis workflow for sequencing and identification of intact glycopeptides. Anal Chem. 2014;86(11):5478–86.

17. Saba J, Dutta S, Hemenway E, Viner R. Increasing the productivity of glycopeptides analysis by using higher-energy collision dissociation-accurate mass-product-dependent electron transfer dissociation. Int J Proteome. 2012;2012

18. Frese CK, Altelaar AM, Hennrich ML, et al. Improved peptide identification by targeted fragmentation using CID, HCD and ETD on an LTQ-Orbitrap Velos. J Proteome Res. 2011;10(5):2377–88.

19. Gaspari M, Verhoeckx KC, Verheij ER, van der Greef J. Integration of two-dimensional LC-MS with multivariate statistics for comparative analysis of proteomic samples. Anal Chem. 2006;78(7):2286–96.

20. America AH, Cordewener JH, van Geffen MH, et al. Alignment and statistical difference analysis of complex peptide data sets generated by multidimensional LC-MS. Proteomics. 2006;6(2):641–53.

21. Pérez-Enciso M, Tenenhaus M. Prediction of clinical outcome with microarray data: a partial least squares discriminant analysis (PLS-DA) approach. Hum Genet. 2003;112(5–6):581–92.

22. Chevallier S, Bertrand D, Kohler A, Courcoux P. Application of PLS-DA in multivariate image analysis. J Chemom. 2006;20(5):221–9.

23. Youssef MA, Abou-Setta AM, Lam WS. Recombinant versus urinary human chorionic gonadotrophin for final oocyte maturation triggering in IVF and ICSI cycles. Cochrane Libr. 2011.

24. Chang P, Kenley S, Burns T, et al. Recombinant human chorionic gonadotropin (rhCG) in assisted reproductive technology: results of a clinical trial comparing two doses of rhCG (Ovidrel R) to urinary hCG (Profasi R) for induction of final follicular maturation in in vitro fertilization–embryo transfer. Fertil Steril. 2001;76(1):67–74.

25. Trinchard-Lugan I, Khan A, Porchet H, Munafo A. Pharmacokinetics and pharmacodynamics of recombinant human chorionic gonadotrophin in healthy male and female volunteers. Reprod Biomed Online. 2002;4(2):106–15.

26. Sakhel K, Khedr M, Schwark S, Ashraf M, Fakih MH, Abuzeid M. Comparison of urinary and recombinant human chorionic gonadotropin during ovulation induction in intrauterine insemination cycles: a prospective randomized clinical trial. Fertil Steril. 2007;87(6):1357–62.

27. Madani T, Yeganeh LM, Ezabadi Z, Hasani F, Chehrazi M. Comparing the efficacy of urinary and recombinant hCG on oocyte/follicle ratio to trigger ovulation in women undergoing intracytoplasmic sperm injection cycles: a randomized controlled trial. J Assist Reprod Genet. 2013;30(2):239–45.

28. Papanikolaou EG, Fatemi H, Camus M, et al. Higher birth rate after recombinant hCG triggering compared with urinary-derived hCG in single-blastocyst IVF antagonist cycles: a randomized controlled trial. Fertil Steril. 2010;94(7):2902–4.

29. Fournier T, Guibourdenche J, Evain-Brion D. Review: hCGs: different sources of production, different glycoforms and functions. Placenta. 2015;36:S60–5.

30. Group ERHCGS. Induction of final follicular maturation and early luteinization in women undergoing ovulation induction for assisted reproduction treatment—recombinant HCG versus urinary HCG. Hum Reprod. 2000;15(7):1446–51.

31. Fernández-Tejada A, Vadola PA, Danishefsky SJ. Chemical synthesis of the β-subunit of human luteinizing (hLH) and chorionic gonadotropin (hCG) glycoprotein hormones. J Am Chem Soc. 2014;136(23):8450–8.

32. Ye H, Hill J, Gucinski AC, Boyne II MT, Buhse LF. Direct site-specific glycoform identification and quantitative comparison of glycoprotein therapeutics: imiglucerase and velaglucerase alfa. AAPS J. 2015;17(2):405–15.

33. Kelly LS, Birken S, Puett D. Determination of hyperglycosylated human chorionic gonadotropin produced by malignant gestational trophoblastic neoplasias and male germ cell tumors using a lectin-based immunoassay and surface plasmon resonance. Mol Cell Endocrinol. 2007;260:33–9.

34. Toll H, Berger P, Hofmann A, et al. Glycosylation patterns of human chorionic gonadotropin revealed by liquid chromatography-mass spectrometry and bioinformatics. Electrophoresis. 2006;27(13):2734–46.

35. Valmu L, Alfthan H, Hotakainen K, Birken S, Stenman U-H. Site-specific glycan analysis of human chorionic gonadotropin β-subunit from malignancies and pregnancy by liquid chromatography—electrospray mass spectrometry. Glycobiology. 2006;16(12):1207–18.