



**HAL**  
open science

## Identification and semi-relative quantification of intact glycoforms by nano-LC–(Orbitrap)MS: application to the $\alpha$ -subunit of human chorionic gonadotropin and follicle-stimulating hormone

Amira Al Matari, Audrey Combès, Julien Camperi, Thierry Fournier, Valérie Pichon, Nathalie N. Delaunay

### ► To cite this version:

Amira Al Matari, Audrey Combès, Julien Camperi, Thierry Fournier, Valérie Pichon, et al.. Identification and semi-relative quantification of intact glycoforms by nano-LC–(Orbitrap)MS: application to the  $\alpha$ -subunit of human chorionic gonadotropin and follicle-stimulating hormone. *Analytical and Bioanalytical Chemistry*, 2020, 412 (23), pp.5729-5741. 10.1007/s00216-020-02794-3 . hal-03013700

**HAL Id: hal-03013700**

**<https://hal.sorbonne-universite.fr/hal-03013700v1>**

Submitted on 19 Nov 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# **Identification and semi-relative quantification of intact glycoforms by nanoLC-(Orbitrap)MS: application to the $\alpha$ -subunit of human chorionic gonadotropin and follicle-stimulating hormone**

Amira Al Matari<sup>1</sup>, Audrey Combès<sup>1</sup>, Julien Camperi<sup>1</sup>, Thierry Fournier<sup>2</sup>, Valérie Pichon<sup>1,3</sup>, Nathalie Delaunay<sup>1</sup>

<sup>1</sup>Laboratory of Analytical, Bioanalytical Sciences and Miniaturization, Chemistry, Biology and Innovation (CBI) UMR 8231, ESPCI Paris PSL, CNRS, PSL Research University, Paris, France

<sup>2</sup>Université de Paris, INSERM, «Pathophysiology & Pharmacotoxicology of the Human Placenta, pre & postnatal Microbiota», 3PHM, Paris, France.

<sup>3</sup>Sorbonne Université, France

Corresponding author: Nathalie Delaunay, UMR CBI 8231, ESPCI Paris PSL, 10 rue Vauquelin, 75231 Paris cedex 05, France. Tel: (+33)1 4079 4651. nathalie.delaunay@espci.fr

## **Compliance with Ethical Standards**

- a) The authors declare that there are no conflicts of interest.
- b) Research involving Human Participants and/or Animals : no.

**Keywords:** follicle-stimulating hormone, glycoprotein, human chorionic gonadotropin, intact protein, mass spectrometry, nano-liquid chromatography.

## **Abbreviations:**

ACN, acetonitrile; CHO, Chinese hamster ovary; CZE, capillary zone electrophoresis; FA, formic acid; FSH, follicle-stimulating hormone; FSH $\alpha$ , follicle-stimulating hormone alpha subunit; hCG, human chorionic gonadotropin; hCG $\alpha$ , human chorionic gonadotropin alpha subunit; Hex, hexose; HexNAc, N-acetylglucosamine; HILIC, hydrophilic interaction liquid chromatography; HRMS, high-resolution mass spectrometry; LC, high performance liquid chromatography; m/z, mass-to-charge; MS, mass spectrometry; MW, molecular weight; NeuAc, N-acetylneuraminic; PTM, post-translational modification; Q-TOF, quadrupole-time-of-flight; rFSH, a recombinant FSH-based drug (Puregon); rhCG, a recombinant hCG-based drug (Ovitrelle); RPLC, reversed-phase liquid chromatography; RSD, relative standard deviation; TIC, total ion current; XIC, extracted ion chromatogram.



## Abstract

Human chorionic gonadotropin (hCG) and follicle-stimulating hormone (FSH) belong to the family of glycoprotein polypeptide hormones called gonadotropins. They are heterodimers sharing the  $\alpha$ -subunit structure that has 2 N-glycosylation sites. A method based on nano-reversed phase liquid chromatography coupled to high resolution mass spectrometry with an Orbitrap analyzer was developed for the first time to characterize the glycosylation state of the  $\alpha$ -subunit at the intact level. A recombinant hCG-based drug, Ovitrelle®, was analyzed. This method combined to an appropriate data treatment allowed the detection of not only the major isoforms but also the minority ones with a high mass accuracy. More than 30 hCG $\alpha$  glycoforms were detected without overlapping of the isotopic patterns. The figures of merit of the method were assessed. The relative standard deviations (RSDs) of the retention time ranged between 0.1 and 6.08% (n=3), with an average of 0.4%. The RSDs of the peak area measured on the extracted ion chromatogram of each glycoform are below 38% (n=3), with an average of 16%, thus allowing semi-relative quantification. The ability to accurately profile glycosylated variants of hCG $\alpha$  was next demonstrated by comparing qualitatively and semi-quantitatively 3 batches of Ovitrelle®. The method was also used to analyze 3 batches of a recombinant FSH-based drug, Puregon®, and 30 FSH $\alpha$  glycoforms were detected and semi-quantified. This demonstrates the high potential of this method for fast quality control or comparison of the glycosylation of glycoprotein-based pharmaceutical preparations.

## 1. Introduction

Human chorionic gonadotropin (hCG) and follicle-stimulating hormone (FSH) belong to a family of related hormones called gonadotropins [1,2]. They are heterodimers sharing the  $\alpha$ -subunit structure non-covalently bound to a specific  $\beta$ -subunit resulting in differences in their physiological roles and signaling. The  $\alpha$ -subunit has 2 N-glycosylation sites, whereas the  $\beta$ -subunit of hCG has 2 N- and 4 O-glycosylation and of FSH contains 2 N-glycosylation sites. Similar to other hormones such as luteinizing hormone, hCG and FSH are used as therapeutic agents in reproductive medicine. However, their glycosylation state has an impact on their stability, solubility, biological activity, potential toxicity, and clearance [3]. This is why their characterization is required throughout their lifecycle (discovery, preclinical, and clinical steps), production, storage, and delivery.

However, protein glycosylation remains one of the most challenging problems in the analysis of post-translational modifications (PTMs), due to structural differences of glycans, number and types of glycosylation sites, and different degrees of sites occupancy leading to a large number of isoforms that are called glycoforms. Studies generally conducted for the analysis of glycoproteins can be done at three levels that complement each other: glycopeptides [4-6], released glycans [6,7], and intact glycoproteins [6,8,9]. The most commonly used method, called bottom-up, is the analysis of the glycopeptides resulting from the enzymatic digestion of the protein that allows the localization of the glycans along the protein chain in addition to the determination of the molecular weight (MW) of each glycan. The main disadvantages of this method are that it results in the loss of information regarding the different combinations of the glycopeptides together and the lack of certainty of the existence of glycoforms of the intact proteins. Other disadvantages are related to the numerous tedious sample preparation steps such as alkylation, reduction, and enzymatic digestion. These additional steps are costly, time-consuming, and could lead to the loss or modifications of certain glycopeptides. The same disadvantages are faced upon the use of the released glycans in addition to the fact that the obtained glycan profile does not give information on the micro-heterogeneity in presence of various glycosylation sites, and one glycan cannot be attributed to a given glycoform. To avoid the mentioned disadvantages, there is an ongoing effort to develop analytical methods at the intact level to evaluate their macro-heterogeneity by combining all glycans associated to each glycoform, and to obtain a global comprehensive picture of the glycosylation pattern [9]. Nevertheless, separation and identification of intact glycoforms can be a difficult task due to their high MW, number, and structural similarity.

By focusing on hCG and FSH, high-performance capillary zone electrophoresis (CZE) was used to separate

their isoforms [10-12]. For instance, our group recently developed a CZE separation method hyphenated with mass spectrometry (MS) allowing a fast glycoform fingerprinting of 2 hCG-based drugs with a turnaround of only  $\approx 6$  min [11]. Thakur *et al.* involved a CZE separation hyphenated with a linear trap quadrupole–Fourier Transform MS analyzer for assignment of intact glycoforms of 2 recombinant  $\alpha$ -subunits of hCG [12]. This method allowed in 20 min the identification of 9 and 20 major glycoforms of the  $\alpha$ -subunit of recombinant hCG obtained from Chinese hamster ovary (CHO) and murine cell line, respectively.

With regard to high performance liquid chromatography (LC), we recently optimized a hydrophilic interaction liquid chromatography (HILIC) separation hyphenated with UV that gave significantly different chromatographic profiles for 2 hCG-based drugs [13]. Its hyphenation with high-resolution MS (HRMS) using a quadrupole-time-of-flight (Q-TOF) analyzer was next studied, but the use of TFA that had been shown to be necessary for the separation of glycoforms caused a dramatic ion suppression effect in MS [14]. Moreover, the biggest challenge in HILIC for the analysis of intact proteins is the injection conditions, i.e. to have sample media that do not induce a significant deformation of chromatographic peaks while allowing the solubilization of the protein without causing its precipitation. Typically, the protein samples are prepared in 100% water and a very low volume was injected on the HILIC column to prevent from the peak deformation induced by the high percentage of water. Compared to the RPLC mode where the analytes are pre-concentrated at the head of the column, the high percentage of water is a major drawback for the retention of intact proteins on the HILIC columns. Reversed-phase LC (RPLC) for glycoform separation of hCG or FSH was used far more [15-19]. To get more information on the nature and the number of glycoforms, the coupling of RPLC with an MS detection is required as it was first carried out by Toll *et al.* who proposed an identification for 14 glycoforms of hCG $\alpha$  and obtained mass spectra for the  $\beta$ -subunit [16]. However, the ion trap quadrupole analyzer used in this study lacked mass accuracy and resolution which could lead to misidentifications and prevent unambiguous assignment. This is why we recently developed a RPLC separation method coupled to HRMS using a Q-TOF analyzer for the analysis of intact hCG [17]. It led to the detection of several hCG $\alpha$  glycoforms of two hCG-based drugs, Ovitrelle<sup>®</sup> and Pregnyl<sup>®</sup>. An identification of the five main glycoforms was proposed by mass matching, even if, unfortunately, the resolution of the Q-TOF analyzer was still insufficient due to the overlapping of some isotopic patterns corresponding to different glycoforms.

This is why in the current study, for the first time, a method involving a high mass resolution and a high mass accuracy Orbitrap analyzer was used for the analysis of hCG and FSH at the intact level to detect their  $\alpha$ -subunit glycoforms. The LC separation was performed at the nanoLC format to enhance the

sensitivity. Even if nowadays nanoLC is used in routine for the bottom-up approaches in proteomics field [20], there are only a few papers reported its use for the analysis of intact glycoforms of targeted proteins [21,22]. This nanoLC-(Orbitrap)HRMS method combined to an adapted data treatment allowed the detection of more than 30 hCG $\alpha$  glycoforms in the Ovitrelle<sup>®</sup> hCG-based drug (a recombinant hCG-based drug, rhCG) without overlapping of the isotopic patterns. After determining some figures of merit of the analytical method, the semi-relative quantification of each glycoform was next performed for the first time and the method was used to compare three batches of rhCG. The method was next successfully used for the qualitative and semi-quantitative glycosylation characterization of FSH $\alpha$  of Puregon<sup>®</sup>, a recombinant FSH-based drug (rFSH) and three batches were analyzed. The batch-to-batch studies of the glycoform heterogeneity demonstrated the utility of the nanoLC-HRMS for quality control of biopharmaceutical preparations.

## **2. Experimental Part**

### **2.1 Materials and Reagents**

HPLC-grade acetonitrile (ACN), formic acid (FA), and LC-MS-grade water were supplied by Carlo Erba (Val de Reuil, France). Puregon<sup>®</sup> and Ovitrelle<sup>®</sup> (Organon, Oss, The Netherlands) are drugs made of recombinant FSH and hCG from a CHO cell line; they will be written as rFSH and rhCG, respectively, in the rest of the manuscript. rFSH contains 900 UI of FSH in a 1.08 mL solution. rhCG contains 250  $\mu$ g of hCG resuspended in 500  $\mu$ L of water.

### **2.2 Sample preparation**

The hormone-based drug stock solutions were desalted and filtered for 6 min at 4°C with ultra-centrifugal units (Microcon) (Merck, Darmstadt, Germany) with a MW cut-off of 10 kDa. This step was performed to remove additives and excipients from the formulation before the analysis. Then the residues were washed 3 times with water. Finally, the samples were recovered from the inverted membrane by centrifugation for 4 min at 14,000 g (about 11,000 rpm) and the volumes were adjusted with water to obtain a final concentration of 160  $\mu$ g mL<sup>-1</sup> for each batch of FSH and 500  $\mu$ g mL<sup>-1</sup> for each batch of hCG. In this study, 3 batches of rFSH and 3 of rhCG were analyzed each in triplicate analysis.

### **2.3 NanoLC-MS system**

The analyses were carried out using a nanoLC system (NC 3500 RS, ThermoFisher Scientific, Courtaboeuf, France) coupled to an Orbitrap MS (Exactive plus, ThermoFisher Scientific). The column used was an

EASY-Spray PepMap RSLC C18 (150 x 0.075 mm internal diameter, 300 Å) packed with 5 µm particles (ThermoFisher Scientific). The mobile phase was composed of (A) 2% FA in H<sub>2</sub>O/ACN (98/2; v/v) and (B) 2% FA in H<sub>2</sub>O/ACN (1/9; v/v), and the column temperature was 60°C. The injection volume was set at 70 nL and the flow-rate at 300 nL min<sup>-1</sup>. The elution gradient starts from 10 to 30% of ACN in 30 min, followed by a rapid increase to 62% of ACN within 5 min, then a 5 min plateau, and finally a return to the initial 10% of ACN concentration and kept for 15 min for equilibration. The EASY-spray source parameters were optimized as follows: capillary voltage, 1.7 kV; drying temperature, 280°C; and sheath, auxiliary, and sweep gas flows were set at 0. The focusing voltages were tuned weekly for the transmission of the standard mixture positive ion calibration solution (ref 88323, ThermoFisher Scientific) and the negative ion calibration solution (ref 88324, ThermoFisher Scientific). The mass spectrometer was operated in the positive mode, in full MS, using a scan range from 1000 to 2500 m/z with a scan rate of 1 spectrum/s. The mass resolution was set to 140,000 at 200 m/z, the AGC target was 3e6, and the maximum injection time was 200 ms. The parameters for deconvolution used were mass tolerance for m/z of 30 mDa, and a minimal charge state of 5, with at least 3 different charge states to identify a compound, accuracy mass tolerance for deconvoluted MW was the 300 mDa. The deconvolution spectra were processed using Freestyle 1.3 SP2 (ThermoFisher Scientific).

### **3. Results and discussion**

#### **3.1 Development of the nanoLC-MS method**

We previously optimized a RPLC separation of hCG glycoforms using 3.6 µm core-shell C18-bonded silica-based particles packed in a column with conventional dimensions (150 × 2.1 mm) [17]. This method was here transferred to the nano format to enhance the sensitivity and detect minor glycoforms. The nanoLC parameters were optimized starting from our previous optimized conditions in LC. Due to the limited choice of nanoLC columns, an EASY-Spray PepMap RSLC C18 (150 x 0.075 mm) column was selected because it drastically reduces the extra-column band broadening, one of the most negative effects in nanoLC, by having electrospray emitter directly connected to the column, thus minimizing post-column dead volumes. In addition, this temperature-controlled column enables a perfect stability of the column temperature during the analysis compared to other nanoLC columns, since the capillary between the column and the MS source is heated. A column temperature of 65°C was previously optimized, but the value was set here at 60°C, since it is the maximal recommended value given by the manufacturer. A similar mobile phase was used consisting of a water-ACN mixture but acidified with 2% of FA instead of 0.1% to favor the multiple ionization of hCG glycoforms. The gradient previously optimized was slightly



adjusted to these new temperature and column to finally recover the chromatographic profile previously obtained with the conventional LC column (2.1 mm) [17]. The Orbitrap MS parameters were also optimized. Various spray voltage values were tested (2.5, 2.2, 2, 1.7, 1.5, and 1.2 kV) to minimize the possibility of in-source fragmentation. The final selected spray voltage was 1.7 kV, which enabled more glycoforms to be detected with the highest signal intensity and generating the fewest smaller  $m/z$  fragments while keeping a stable spray. rhCG was used as standard sample after a simple filtration step. Using these optimized conditions, the typical total ion current (TIC) obtained for the analysis of hCG is presented **Figure 1A**. hCG sample contains so many glycoforms that they cannot be baseline resolved giving rise to a large massif of peaks between 15 and 21.5 min, as it was already observed in LC [17]. However, the separation power of the MS analyzer can partially overcome it.

The objectives of transferring the analysis to the nanoLC format were to reduce sample consumption and improve sensitivity. With respect to sample consumption, the injected volume decreased from 5  $\mu\text{L}$  in LC to 70 nL in nanoLC, thus corresponding to a decrease in the injected quantity of a factor 14 as the hCG concentration was increased from 100 to 500  $\mu\text{g mL}^{-1}$ . Regarding the sensitivity, a theoretical gain of 784 based on the ratio of the square column diameters was calculated. The signal to noise ratio values were calculated on the TIC for the most intense peak obtained in LC [17] and in nanoLC (**Figure 1A**). It allowed the calculation of the experimental sensitivity enhancement and a value of about 40 was obtained, taking into account the different injected amount values, which is lower than expected but still interesting. Several hypotheses can explain this difference: (i) the relative high-injected volume in nanoLC, (ii) the presence of dead volumes in the system, both parameters causing a band broadening and then a decrease of peak heights, and (iii) the use of different MS detectors.

### **3.2 Nano-LC-HRMS data treatment for the identification of glycoforms**

Electrospray ionization led to multiple charged values for each molecular mass and an isotopic cluster for each nominal mass-to-charge ( $m/z$ ) value ratio, thus leading to highly complex mass spectra. For complex samples studied in MS-based proteomics, overlapping isotope patterns are commonly encountered and are problematic for accurate mass determination. To overcome these problems, the first step in spectral interpretation usually consists in spectral deconvolution, which converts the complex spectrum to a list of monoisotopic deconvoluted masses [23]. In the current study, two deconvolution methods were compared with the same algorithm in Freestyle (1.3) software.

We first applied the most commonly used deconvolution method by selecting the whole chromatographic peak (retention time between 15.5 and 21.5 min) at once and the resulting average

mass spectrum with the determined charge states is presented in **Figure 1B**. The deconvolution algorithm was next performed to determine the corresponding MWs. It allowed the detection of only 12 compounds with (M+H) ranging between 13,521.8 and 14,963.2 Da which correspond to hCG $\alpha$  glycoforms.

For the identification of glycoforms, the MWs of the detected putative glycoforms were compared to those of the primary amino acid sequence of hCG $\alpha$  plus the MW of all the combinations of two potential N-glycans already described in the literature for hCG $\alpha$  [16,24] and to the ones that we recently identified with a bottom-up approach [25]. From a practical point of view, the hCG $\alpha$  backbone sequence was thus converted to its elemental composition. Ten hydrogens were eliminated from the elemental composition due to the fact that five disulfide bridges are present in this subunit. Then the elemental composition of two identical glycans bearing two sialylated antennas (HexNAc(4)Hex(5)NeuAc(2)) was determined. These glycans were chosen because they are considered to be the most commonly encountered [16,24,25]. The combined elemental composition of hCG $\alpha$  and these two glycans (C<sub>84</sub>H<sub>158</sub>N<sub>6</sub>O<sub>75</sub>) was used in a MW calculator to obtain the theoretical isotopic profile of this glycoform ([M+H] = 14,598.2 Da) [26]. This value was present in the data of the current study as one of the most intense isotopes with a mass tolerance of 0.03 Da for the m/z and mass tolerance of 0.3 Da for the MW. The experimental isotopic profile (**Figure 2**) matches with the theoretical one (**Figure S1**) and allows the identification of the mentioned glycoform. The structure of this glycoform was the starting point of mass matching to determine the other structures. **Figure 3** represents the corresponding average deconvoluted mass spectrum with the 12 detected glycoforms with 9 of which were identified. The structures of the corresponding N-glycans present on each hCG $\alpha$  glycoform are also represented. The nanoLC-HRMS method allowed thus to confirm the identity of the glycoforms that could have been only assumed using our previously developed LC-(Q-TOF)MS method [17]. All these glycoforms were confirmed by a complementary bottom-up approach, which was applied to the same hCG sample [25].

The later average MS spectrum over 7 min of retention time enabled only the detection of the majority peaks. To improve the data processing to detect the minority forms, a second data treatment approach was considered. It consists in cutting the chromatogram in successive time slices of 30 s, starting from the retention time of 15 till 22 min, and also in slices of 50 s, starting from 15.5 till 21.5 min. The average deconvoluted mass spectrum for each slice was obtained resulting in the achievement of a list of M+H corresponding to detected glycoforms present in rhCG. Next, the results of both types of division were combined to reduce the risk of data loss at the boundaries of the performed cuts. A list of 28 M+H ranging between 13,050.6 and 15,254.4 Da corresponding to hCG $\alpha$  glycoforms with different retention

times was obtained, which increased by a factor of 2 the number of detected glycoforms by comparison with results from the first data treatment approach.

This second data treatment approach was applied to elucidate the structure of the two N-glycans present on each glycoform for 3 different batches of rhCG. **Table 1** presents the M+H of the detected glycoforms in these batches to compile the maximum information obtained by this method. It is worthwhile to mention that the retention time range where glycoforms were identified corresponds on the TIC to the central part of the peak massif with the highest intensities. The corresponding N-glycan structures for each hCG $\alpha$  glycoform were determined by mass matching and by comparison with results from the bottom-up approach [25] and with all the N-glycan structures identified for gonadotropins gathered by GlyGen database [27]. It was sometimes not possible to elucidate the glycoform structure that may result from a combination of glycosylation plus other PTMs. Of course, the previously 12 detected glycoforms were present but this second approach, even if it is non-automated and time-consuming, allowed also the detection of additional glycoforms that have sometimes a much less intense signal. Indeed, a ratio of 200 was obtained between the highest and the smallest intensity of the detected glycoforms. It thus provides a very much more detailed profile of hCG $\alpha$ .

**Table 1** also shows all the masses detected in each run of each batch. It appears that all the detected masses are detected in each run for a given batch. There are only one exception in batch 2 and one in batch 3, for which the mass value was detected only in two runs out of the three. For batch 3, it corresponds to a glycoform of low intensity and that has a retention time similar to 4 other glycoforms (the 5 glycoforms are eluted in 1.2 s), which may explain that in one run, the ionization process of this glycoform was not sufficiently efficient to detect it. For batch 2, it is more difficult to find an explanation because it concerns a glycoform with a higher intensity that does not co-elute with other ones. Nevertheless, regarding the total number of detected glycoforms and the complexity of the sample, it can be concluded that the direct nanoLC-HRMS analysis of the hCG-based drug in duplicate should be sufficient to get consistent hCG $\alpha$  glycoprofiling.

### 3.3 Figures of merit of the nanoLC-HRMS method

The repeatability of the nanoLC-HRMS analysis was investigated. The 3 batches of rhCG were analyzed, each in triplicate, and the resulting detected masses and their corresponding average retention time and RDS values are compiled in **Table S1** (see supplementary data). It appears that the RSDs of retention times are all inferior to 6.08%, with an average value of 0.4%, demonstrating a good repeatability of the analysis in terms of retention times. Next, the peak area of each glycoform was measured on its

corresponding extracted ion chromatogram (XIC) and normalized by the area of the glycoform leading to the highest area value measured on its own XIC which corresponds to the glycoform having a MW 14,307.1 Da. It is worthwhile to notice that the glycoform giving the highest area is not necessarily the most abundant in the initial hCG sample due to potential bias associated with the electrospray ionization process. The obtained average relative area and RSD of each detected mass are reported in **Figure 4** and all the data are compiled in **Table S1**. The RSD values are below 38%, with an average of 16%, which is acceptable for the analysis of intact proteins by nanoLC-HRMS [28].

Next, the linearity of the method was also studied for the highest intensity glycoforms for 5 concentrations of hCG ranging from 7.5 to 250  $\mu\text{g mL}^{-1}$  leading to correlation coefficient ( $R^2$ ) values between 0.991 and 0.997. All the  $R^2$  values and the equations of regression of each glycoform are gathered in **Table S2** (see supplementary data).

The determination of the limit of quantification (LOQ) was achieved by successive dilutions of the rhCG sample. It should be mentioned that if the global hCG concentration is known, the exact concentration of each glycoform present in the sample is unknown. By considering only the 11 glycoforms giving the highest signal in MS (fixed arbitrarily above 15% for their relative areas), they were all detected injecting 7.5  $\mu\text{g mL}^{-1}$  of the global hCG content with a signal to noise ratio measured on the XIC of each glycoform above 10. However, the injection of 2.5  $\mu\text{g mL}^{-1}$  of the global hCG content led to the absence of peaks. Therefore, the LOQ of the glycoforms giving the highest signal in MS is about 7.5  $\mu\text{g mL}^{-1}$  of the global hCG content.

### **3.4 Potential of the analytical method for batch-to-batch comparison**

The repeatability of the retention times and relative areas was assessed, which allowed the identification and semi-quantification of about 30 hCG $\alpha$  glycoforms. This analytical method was used to compare the 3 batches of rhCG as a proof-of-concept of its potential to characterize the glycosylation state of biopharmaceuticals in a simple and rapid manner. In **Table S1**, some qualitative differences between the three batches appear. First, the number of glycoforms detected in each batch varies: 28 in batch 1, 30 in batch 2, and 33 in batch 3. In more detail, the three batches shared 24 glycoforms with similar retention times while 10 glycoforms were only detected in one or two batches. As already mentioned, the glycoforms in a given batch are always detected in each run of the 3 repeated analysis, except for two that were detected in two runs only, which indicates that the result of 10 glycoforms that are different in the 3 batches is consistent and is not coming from a lack of repeatability of the analytical method. This difference between batches in terms of nature of some glycoforms exists.

Next, the three batches were compared semi-quantitatively based on the average relative area of each glycoform (**Table S1**). It is worthwhile to notice that only semi-quantitative comparison was possible due to the fact that the efficiency of the electrospray ionization process depends on the structure of the compounds and their co-elution. Therefore, it is not possible to directly correlate the relative area with the relative concentration of glycoforms in the sample because a glycoform could be better ionized than another one. **Figure 4** allows the comparison of the relative area of each detected glycoform ordered by increasing retention time. First the most abundant glycoform is the same in the three batches. Second, regarding the 11 most abundant glycoforms ( $\geq 15\%$ ), 7 have very close relative area in the three batches while 2 are not detected in batch 2. Next, the sum of the relative area of the 24 glycoforms that are common to the 3 batches are equal to 87, 96, and 78% of the sum of the relative areas of all the glycoforms in batch 1, 2, and 3, respectively. Therefore, the batches are very close from a semi-quantitative point of view, but some relevant differences allow their discrimination. Finally, this simple and effective method directly applied to the intact hCG-based drug without requiring time-consuming and expensive treatment steps enables the detection of numerous hCG $\alpha$  glycoforms and led to the  $\alpha$ -subunit profiling which could be promising for the comparison of other types of gonadotropin-based drugs.

### 3.5 Potential for the analysis of another gonadotropin

#### 3.5.1 Identification of FSH glycoforms

In order to evaluate the applicability of the developed nanoLC-HRMS method for the characterization of the  $\alpha$ -subunit of gonadotropin hormone, it was used for the analysis of a FSH-based drug (rFSH) by applying the same conditions previously optimized. **Figure 5A** presents the resulting TIC. The two deconvolution methods were tested starting first by the selection of the whole peak eluting from 17.5 to 20 min. **Figure 5B** presents the corresponding average deconvoluted mass spectrum. This resulted in the detection of 19 MWs corresponding to FSH $\alpha$  glycoforms ranging between 13,724.8 and 15,328.4 Da. With the second data treatment approach, by cutting the chromatographic peak in 30 s- and 50 s-slices and deconvoluting each slice separately, a total of 30 glycoforms was detected (see **Table S3** in supplementary data). This demonstrates again that this approach is definitively much more efficient to detect a higher number of  $\alpha$ -subunit glycoforms present in the gonadotropin-based drug, even if it is more time consuming and not yet automated than the approach based on the one-time data treatment of the totality of the elution profile.

Due to the fact that FSH and hCG share the same alpha subunit with the same number of possible N-glycosylation sites and same origin (CHO), the obtained masses of FSH were compared with the ones of hCG, the results that we got with the bottom-up approach done on hCG [25], and the GlyGen data base of all identified gonadotropin related N-glycans [27]. This allowed to propose N-glycans structures for 11 FSH $\alpha$  glycoforms (see **Table S3** in supplementary data), which demonstrates the ability of the nanoLC-HRMS method for the characterization of the glycosylation of the  $\alpha$ -subunit of gonadotropins analyzing them at the intact level.

### 3.5.2 Repeatability of the nanoLC-HRMS method for the analysis of FSH

The repeatability of the method was studied by repeating the analysis of a first batch of rFSH (n=3) and involving the second data treatment approach. The number of detected masses was compared and proved to be the same in each run except for one glycoform that was detected only in two runs out of the three. It corresponds to a glycoform of low intensity (relative area of 8%) and that has a retention time close to 4 other glycoforms (the 5 glycoforms are eluted in 1.2 s), which may explain that in one run, the ionization process of this glycoform was not sufficiently efficient to detect it. For each detected mass, the retention time was assigned and the average RSD of retention times was equal to 0.3% (n=3). Next, 2 additional batches of rFSH were analyzed in triplicate. **Table S4** (see supplementary data) shows all the masses detected in each run of each batch and, for each glycoform, their corresponding average retention time and average relative area plus their RSD values are given. The RSD values of the retention times were inferior to 1.2% with an average of 0.4 %, while that of the relative area are below 39% with an average of 17%.

### 3.5.3 Comparison of rFSH batches

As the repeatability of the nanoLC-HRMS method was assessed, the comparison of the 3 batches of rFSH was performed using the same approach as the one used previously with hCG. For ease of comparison, **Figure 6** presents the average relative area of each glycoform in each batch. First, from a qualitative point of view, the 3 batches shared 28 glycoforms. One glycoform was detected only in batch 3 and another one is missing in batch 2, but these glycoforms have very low relative area values when they are detected, which may explain why they were not detected in some cases. Therefore, the 3 batches are qualitatively very similar.

It can also be noticed that the most “abundant” glycoform (*i.e.* the one having the highest average area) is the same in the 3 batches. In addition, 19 glycoforms have a relative area higher than 15% and they

are all present in the three batches showing that the batches are quite homogeneous. The sums of the relative areas of the 28 glycoforms that are common to the three batches are equal to 95, 100, and 98% of the sum of the relative areas of all the glycoforms in batch 1, 2, and 3, respectively. Therefore, the batches are very close from a semi-quantitative point of view. These results demonstrate again the potential of the analysis by nanoLC-HRMS of the glycoproteins at the intact level to qualitatively compare the glycoprofiling of different samples in addition to the possibility of semi-quantitative assessment for quality control purposes for example.

### 3.6 rhCG vs rFSH

The nanoLC-HRMS method enabled the identification and semi-quantification of the  $\alpha$ -subunit glycoforms of the two gonadotropins hCG and FSH contained in rhCG and rFSH, respectively. Despite the fact that these two hormones share the same  $\alpha$ -subunit, their analysis led to the detection of different glycoforms. Only 13 of those were in common between the two as shown in **Figure 7A**, which means less than half of them. Therefore, both pharmaceutical drugs have different glycoprofiling of the  $\alpha$ -subunit.

A pie graph was constructed (**Figure 7B**) to compare semi-quantitatively the average relative area of each glycoform detected in the 3 batches of rhCG and rFSH analyzed in triplicate (n=9). It shows that the common glycoforms seem to be the most abundant ones, with the hypothesis that the MS signal is correlated to the initial concentration in the sample. Nevertheless, it appears that the composition of the two pharmaceutical preparations varies significantly with regard to the  $\alpha$ -subunit glycoforms, even if the differences in MS intensities may result from differences in MS ionization. Although this hypothesis cannot be eliminated, it seems unlikely that it will lead to such different profiles. Knowing that rFSH and rhCG are recombinant gonadotropins produced using the CHO cell line and the same DNA recombinant technology, it does not eliminate the fact that there are still slight differences in their production and purification procedures, leading to non-identical preparations, with variations in glycosylation that yield different sialic acid residue compositions for example [29]. This could explain the differences between the identified isoforms.

## 4. Conclusion

Special attention is being implemented in studying glycoproteins due to the fact that glycosylation has a strong impact on the properties of a glycoprotein-based drug. This work demonstrated for the first time the great potential of nano-LC coupled to HRMS (Orbitrap) for fast glycoprofiling of the  $\alpha$ -subunit of two gonadotropin-based drugs analyzed at the intact level after simple filtration, without any tedious sample

preparations. The good figures of merit of the method allow its use for identification and semi-quantification of the glycoforms. It could be useful at different stages of the life of these kinds of drugs, such as their development (discovery, preclinical, and clinical steps), production, storage, and delivery.

### **Acknowledgements**

This work has received the support of "Institut Pierre-Gilles de Gennes" (Laboratoire d'excellence, "Investissements d'avenir" program ANR-10-IDEX-0001-02 PSL and ANR-10-LABX-31), of PSL University, the Conseil Régional d'Île-de-France with the DIM Analytics and of the association "Al Ihssan".



## References

- [1] C. Nwabuobi, S. Arlier, F. Schatz, O. Guzeloglu-Kayisli, C. J. Lockwood, and U. A. Kayisli, hCG: Biological Functions and Clinical Applications, *Int J Mol Sci* 18 (2017) 2037, doi: 10.3390/ijms18102037.
- [2] J. G. Pierce and T. F. Parsons, Glycoprotein hormones: structure and function, *Annu Rev Biochem* 50 (1981) 465, doi: 10.1146/annurev.bi.50.070181.002341.
- [3] A. Stockell Hartree and A. G. C. Renwick, Molecular structures of glycoprotein hormones and functions of their carbohydrate components, *Biochem J* 287 (1992) 665, doi: 10.1042/bj2870665.
- [4] V. Kolli, K. N. Schumacher, and E. D. Dodds, Engaging challenges in glycoproteomics: recent advances in MS-based glycopeptide analysis, *Bioanalysis* 7 (2015) 113, doi: 10.4155/bio.14.272.
- [5] R. Zhu, L. Zacharias, K. M. Wooding, W. Peng, and Y. Mechref, Glycoprotein Enrichment Analytical Techniques, *Method Enzymol* 585 (2017) 397, doi: 10.1016/bs.mie.2016.11.009
- [6] E. Largy, F. Cantais, G. Van Vyncht, A. Beck, and A. Delobel, Orthogonal liquid chromatography–mass spectrometry methods for the comprehensive characterization of therapeutic glycoproteins, from released glycans to intact protein level, *J Chromatogr A* 1498 (2017) 128, doi: 10.1016/j.chroma.2017.02.072.
- [7] A. Palmigiano, A. Messina, L. Sturiale, and D. Garozzo, Advanced LC-MS Methods for N-Glycan Characterization, *Compr Anal Chem* 79 (2018) 147, doi: 10.1016/bs.coac.2017.06.009.
- [8] V. Dotz, R. Haselberg, A. Shubhakar, R. P. Kozak, D. Falck, Y. Rombouts, D. Reusch, G. W. Somsen, D. L. Fernandes, and M. Wuhrer, Mass spectrometry for glycosylation analysis of biopharmaceuticals, *TrAC Trends Anal Chem* 73 (2015) 1, doi: 10.1016/j.trac.2015.04.024.
- [9] J. Camperi, V. Pichon, and N. Delaunay, Separation Methods hyphenated to Mass Spectrometry for the Characterization of the Protein Glycosylation at the Intact Level, *J Pharm Biomed Anal* 178 (2020) 112921, doi: 10.1016/j.jpba.2019.112921.
- [10] P. L. Storrington, R. E. Gaines Das, J. W. M. Mulders, and M. Halder, Physicochemical Methods for Predicting the Biological Potency of Recombinant Follicle Stimulating Hormone: An International Collaborative Study of Isoelectric Focusing and Capillary Zone Electrophoresis, *Biologicals* 30 (2002) 217, doi: 10.1006/biol.2002.0333.
- [11] J. Camperi, B. De Cock, V. Pichon, A. Combes, J. Guibourdenche, T. Fournier, Y. Vander Heyden, D. Mangelings, and N. Delaunay, First characterizations by capillary electrophoresis of human Chorionic Gonadotropin at the intact level, *Talanta* 193 (2019) 77, doi: 10.1016/j.talanta.2018.09.095.
- [12] D. Thakur, T. Rejtar, B. L. Karger, N. J. Washburn, C. J. Bosques, N. S. Gunay, Z. Shriver, and G. Venkataraman, Profiling the Glycoforms of the Intact  $\alpha$  Subunit of Recombinant Human Chorionic Gonadotropin by High-Resolution Capillary Electrophoresis–Mass Spectrometry, *Anal Chem* 81 (2009) 8900, doi: 10.1021/ac901506p.
- [13] J. Camperi, V. Pichon, T. Fournier, and N. Delaunay, First profiling in hydrophilic interaction liquid chromatography of intact human chorionic gonadotropin isoforms, *J Pharm Biomed Anal* 174 (2019) 495, doi: 10.1016/j.jpba.2019.06.014.
- [14] J. Camperi, A. Combes, T. Fournier, V. Pichon, and N. Delaunay, Analysis of the human chorionic gonadotropin protein at the intact level by HILIC-MS and comparison with RPLC-MS, *Anal. Bioanal. Chem.*, accepted, doi: 10.1007/s00216-020-02684-8
- [15] R. Bassett, C. De Bellis, L. Chiacchiarini, D. Mendola, E. Micangeli, K. Minari, L. Grimaldi, M. Mancinelli, R. Mastrangeli, and R. Bucci., Comparative characterisation of a commercial human

- chorionic gonadotrophin extracted from human urine with a commercial recombinant human chorionic gonadotrophin, *Curr Med Res Opin* 21 (2005) 1969, doi: 10.1185/030079905X75005.
- [16] H. Toll, P. Berger, A. Hofmann, A. Hildebrandt, H. P. Lenhof, C. G. Huber, Glycosylation patterns of human chorionic gonadotropin revealed by liquid chromatography-mass spectrometry and bioinformatics, *ELECTROPHORESIS* 27 (2006) 2734, doi: 10.1002/elps.200600022.
- [17] J. Camperi, A. Combes, J. Guibourdenche, D. Guillarme, V. Pichon, T. Fournier, and N. Delaunay, An attempt to characterize the human Chorionic Gonadotropin protein by reversed phase liquid chromatography coupled with high-resolution mass spectrometry at the intact level, *J Pharm Biomed Anal* 161 (2018) 35, doi: 10.1016/j.jpba.2018.07.056.
- [18] R. F. Loureiro, J. E. de Oliveira, P. A. Torjesen, P. Bartolini, and M. T. C. P. Ribela, Analysis of intact human follicle-stimulating hormone preparations by reversed-phase high-performance liquid chromatography, *J Chromatogr A* 1136 (2006) 10, doi: 10.1016/j.chroma.2006.09.037.
- [19] A. Lombardi, C. Andreozzi, V. Pavone, V. Triglione, L. Angiolini, and P. Caccia, Evaluation of the oligosaccharide composition of commercial follicle stimulating hormone preparations, *ELECTROPHORESIS* 34 (2013) 2394, doi: 10.1002/elps.201300045.
- [20] S. R. Wilson, T. Vehus, H. S. Berg, and E. Lundanes, Nano-LC in proteomics: recent advances and approaches, *Bioanalysis* 7 (2015) 1799, doi: 10.4155/bio.15.92.
- [21] M. van Scherpenzeel, G. Steenbergen, E. Morava, R. A. Wevers, and D. J. Lefeber, High-resolution mass spectrometry glycoprofiling of intact transferrin for diagnosis and subtype identification in the congenital disorders of glycosylation, *Transl Res* 166 (2015) 639.e1, doi: 10.1016/j.trsl.2015.07.005.
- [22] J. F. M. Jacobs, R. A. Wevers, D. J. Lefeber, and M. van Scherpenzeel, Fast, robust and high-resolution glycosylation profiling of intact monoclonal IgG antibodies using nanoLC-chip-QTOF, *Clin Chim Acta* 461 (2016) 90, doi: 10.1016/j.cca.2016.07.015.
- [23] G. Xu, J. Stupak, L. Yang, L. Hu, B. Guo, and J. Li, Deconvolution in mass spectrometry based proteomics, *Rapid Commun Mass Spectrom* 32 (2018) 763, doi: 10.1002/rcm.8103.
- [24] V. Blanchard, R. A. Gadkari, G. J. Gerwig, B. R. Leeflang, R. R. Dighe, and J. P. Kamerling, Characterization of the N-linked oligosaccharides from human chorionic gonadotropin expressed in the methylotrophic yeast *Pichia pastoris*, *Glycoconj J* 24 (2007) 33, doi: 10.1007/s10719-006-9010-3.
- [25] S. Perchepied, N. Eskenazi, C. Giangrande, J. Camperi, T. Fournier, J. Vinh, N. Delaunay, and V. Pichon, Development of Immobilized Enzyme Reactors for the characterization of the glycosylation heterogeneity of a protein, *Talanta* 206 (2020) 120171, doi: 10.1016/j.talanta.2019.120171.
- [26] Isotope Distribution Calculator, Mass Spec Plotter, Isotope Abundance Graphs. [Online]. Available: <https://www.sisweb.com/mstools/isotope.htm>. [Accessed: 09-Mar-2020].
- [27] Glygen database, [www.glygen.org](http://www.glygen.org)
- [28] D. Liu, X. Han, X. Liu, M. Cheng, M. He, G. Rainer, H. Gao, X. Zhang, Measurement of ultra-trace level of intact oxytocin in plasma using SALLE combined with nano-LC-MS, *J Pharm Biomed Anal* 173 (2019) 62-67, doi: 10.1016/j.jpba.2013.04.031.
- [29] R. B. F. Leão and S. C. Esteves, Gonadotropin therapy in assisted reproduction: an evolutionary perspective from biologics to biotech, *Clinics* 69 (2014) 279, doi: 10.6061/clinics/2014(04)10.

**Table 1:** MW of each hCG $\alpha$  glycoform detected after nanoLC-HRMS analysis of rhCG (3 batches, 3 runs each) and use of the second data treatment approach with glycan structure suggestions for some of them in comparison with bottom-up data (marked with \*) or with GlyGen database (marked with \*\*), while when those confirmed by both are marked with \*\*\*. x: no structure to propose. The notation 1<sup>st</sup> and 2<sup>nd</sup> glycan is not correlated to their localization on the 2 N-glycosylation sites of hCG $\alpha$ .

Mass (Da)	Number of times the glycoforms was detected among the three runs			1 <sup>st</sup> N-Glycan	Proposed structure	2 <sup>nd</sup> N-Glycan	Proposed structure	Structure identification method
	Batch 1	Batch 2	Batch 3					
13050.6	3/3	3/3	3/3	x		x		
13434.8	3/3	3/3	3/3	x		x		
13521.8	0/3	0/3	3/3	HexNAc(3)Hex(5)* or HexNAc(5)Hex(4)** or HexNAc(4)Hex(5)** or HexNAc(5)Hex(4)NeuAc(1)**		HexNAc(4)Hex(5)NeuAc(1)* or HexNAc(4)Hex(3)NeuAc(1)** or HexNAc(3)Hex(5)NeuAc(1)** or HexNAc(4)Hex(3)**	*  **  **  **	
13609.8	3/3	3/3	3/3	HexNAc(3)Hex(4)NeuAc(1) or HexNAc(3)Hex(5)NeuAc(1) or HexNAc(3)Hex(4)NeuAc(1)		HexNAc(3)Hex(6)NeuAc(1) or HexNAc(3)Hex(5)NeuAc(1) or HexNAc(3)Hex(6)NeuAc(1)	*  **  **	
13650.9	3/3	3/3	2/3	HexNAc(3)Hex(4) or HexNAc(4)Hex(5)NeuAc(1) or HexNAc(3)Hex(6)NeuAc(1)		HexNAc(4)Hex(5)NeuAc(2) or HexNAc(3)Hex(4)NeuAc(1) or HexNAc(4)Hex(3)NeuAc(1)	*  **  **	
13724.9	3/3	3/3	3/3	HexNAc(4)Hex(5) or HexNAc(3)Hex(4) or HexNAc(5)Hex(6)		HexNAc(4)Hex(5)NeuAc(1) or HexNAc(5)Hex(6)NeuAc(1) or HexNAc(3)Hex(4)NeuAc(1)	*  **  **	

				or HexNAc(6)Hex(4)NeuAc(1)	or HexNAc(4)Hex(3)	**
13756.8	0/3	3/3	3/3	x	x	
13812.9	3/3	3/3	3/3	HexNAc(3)Hex(5) or HexNAc(3)Hex(5)NeuAc(1)	HexNAc(4)Hex(5)NeuAc(2) or HexNAc(4)Hex(5)NeuAc(1)	* **
13847.8	3/3	3/3	3/3	x	x	
13852.9	0/3	3/3	3/3	HexNAc(5)Hex(4)NeuAc(1)	HexNAc(3)Hex(5)NeuAc(1)	**
13942.9	3/3	3/3	3/3	HexNAc(5)Hex(4)	HexNAc(6)Hex(3)Fuc(1)	**
13974.0	3/3	3/3	3/3	HexNAc(3)Hex(6)NeuAc(1)	HexNAc(4)Hex(5)NeuAc(1)	*
14016.0	3/3	3/3	3/3	HexNAc(4)Hex(5)NeuAc(1)	HexNAc(4)Hex(5)NeuAc(1)	***
14024.1	3/3	0/3	3/3	x	x	
14032.0	3/3	3/3	3/3	HexNAc(5)Hex(4)NeuAc(1) or HexNAc(3)Hex(5)NeuAc(1) or HexNAc(4)Hex(4)Fuc(1)	HexNAc(5)Hex(4)Fuc(1) or HexNAc(5)Hex(6)Fuc(1) or HexNAc(6)Hex(4)NeuAc(1)	**
14048.9	0/3	3/3	3/3	HexNAc(3)Hex(6)	HexNAc(5)Hex(6)NeuAc(1)	**
14089.0	3/3	0/3	3/3	x	x	
14104.0	3/3	3/3	3/3	HexNAc(4)Hex(5)NeuAc(2)	HexNAc(3)Hex(5)NeuAc(1)	**
14139.0	3/3	3/3	3/3	x	x	
14148.0	3/3	3/3	3/3	x	x	
14265.1	3/3	3/3	3/3	x	x	
14307.1	3/3	3/3	3/3	HexNAc(4)Hex(5)NeuAc(1)	HexNAc(4)Hex(5)NeuAc(2)	***

14316.2	3/3	0/3	3/3	x	x	
14323.0	3/3	3/3	3/3	HexNAc(5)Hex(5)NeuAc(2) or HexNAc(5)Hex(4)	HexNAc(5)Hex(3)Fuc(1) or HexNAc(5)Hex(4)NeuAc(2)Fuc(1)	**
14339.0	0/3	3/3	3/3	x	x	
14381.1	3/3	3/3	3/3	HexNAc(4)Hex(5)NeuAc(1) or HexNAc(3)Hex(5) or HexNAc(5)Hex(4) or HexNAc(5)Hex(4)NeuAc(1)	HexNAc(5)Hex(6)NeuAc(1) or HexNAc(6)Hex(6)NeuAc(1) or HexNAc(6)Hex(4)NeuAc(2) or HexNAc(6)Hex(4)NeuAc(1)	***  **  **  **
14598.2	3/3	3/3	3/3	HexNAc(4)Hex(5)NeuAc(2)	HexNAc(4)Hex(5)NeuAc(2)	***
14607.2	3/3	3/3	3/3	HexNAc(5)Hex(5)NeuAc(2) or HexNAc(5)Hex(4)Fuc(1)	HexNAc(4)Hex(5)Fuc(1) or HexNAc(6)Hex(4)NeuAc(2)	**  **
14614.2	3/3	3/3	3/3	x	x	
14640.2	3/3	3/3	3/3	x	x	
14672.2	3/3	3/3	3/3	HexNAc(4)Hex(5)NeuAc(2)	HexNAc(5)Hex(6)NeuAc(1)	***
14688.2	0/3	2/3	3/3	HexNAc(5)Hex(4)Fuc(1)	HexNAc(6)Hex(4)NeuAc(1)	**
14963.3	3/3	3/3	3/3	HexNAc(4)Hex(5)NeuAc(2)	HexNAc(5)Hex(6)NeuAc(2)	*
15254.4	3/3	3/3	3/3	HexNAc(4)Hex(5)NeuAc(2)	HexNAc(5)Hex(6)NeuAc(3)	*

## Figure Caption

**Figure 1:** (A) TIC of hCG analyzed by nanoLC-HRMS. (B) Average mass spectrum corresponding to the totality of the chromatographic peak observed (retention time range: 15.5-21 min). Conditions: see section 2.3. rhCG, batch 1, run 1.

**Figure 2:** Zoom on average mass spectrum on the mass range [1825.6 – 1828.0 Da] showing the experimental isotopic profile ( $m/z = 1825.78$ ) corresponding to the most commonly described hCG $\alpha$  glycoforms (hCG $\alpha$  with 2 HexNAc(4)Hex(5)NeuAc(2) N-glycans).

**Figure 3:** Deconvoluted average mass spectrum, determined over the entire chromatographic peak observed in Figure 1, representing the 12 detected hCG $\alpha$  glycoforms and the N-glycan structures of the 9 glycoforms that were identified among the 12.

**Figure 4:** Comparison of the relative intensity of the glycoforms of hCG $\alpha$  present in 3 batches of rhCG analyzed by nanoLC-HRMS ( $n=3$  for each). The comparison is based on the average relative area of each glycoform characterized by its MW and ordered by their retention times.

**Figure 5:** A: TIC of FSH obtained by nanoLC-HRMS. B: average deconvoluted mass spectrum for the 17.5-20 min range. rFSH, batch 1, run 1. Conditions: see section 2.3.

**Figure 6:** Comparison of the glycoforms of FSH $\alpha$  present in 3 batches of rFSH analyzed by nanoLC-HRMS ( $n=3$  for each). The comparison is based on the average relative area of each glycoform characterized by its mass and ordered by their retention time.

**Figure 7:** A: Venn diagram comparing the number of the  $\alpha$ -subunit glycoforms identified by the analysis of rhCG and rFSH by nanoLC-HRMS. B: Pie charts representing the average relative area of each detected glycoform in (Top) rhCG and (Bottom) rFSH. For each drug, 3 batches were analyzed 3 times each ( $n=9$ ).

Figure 1

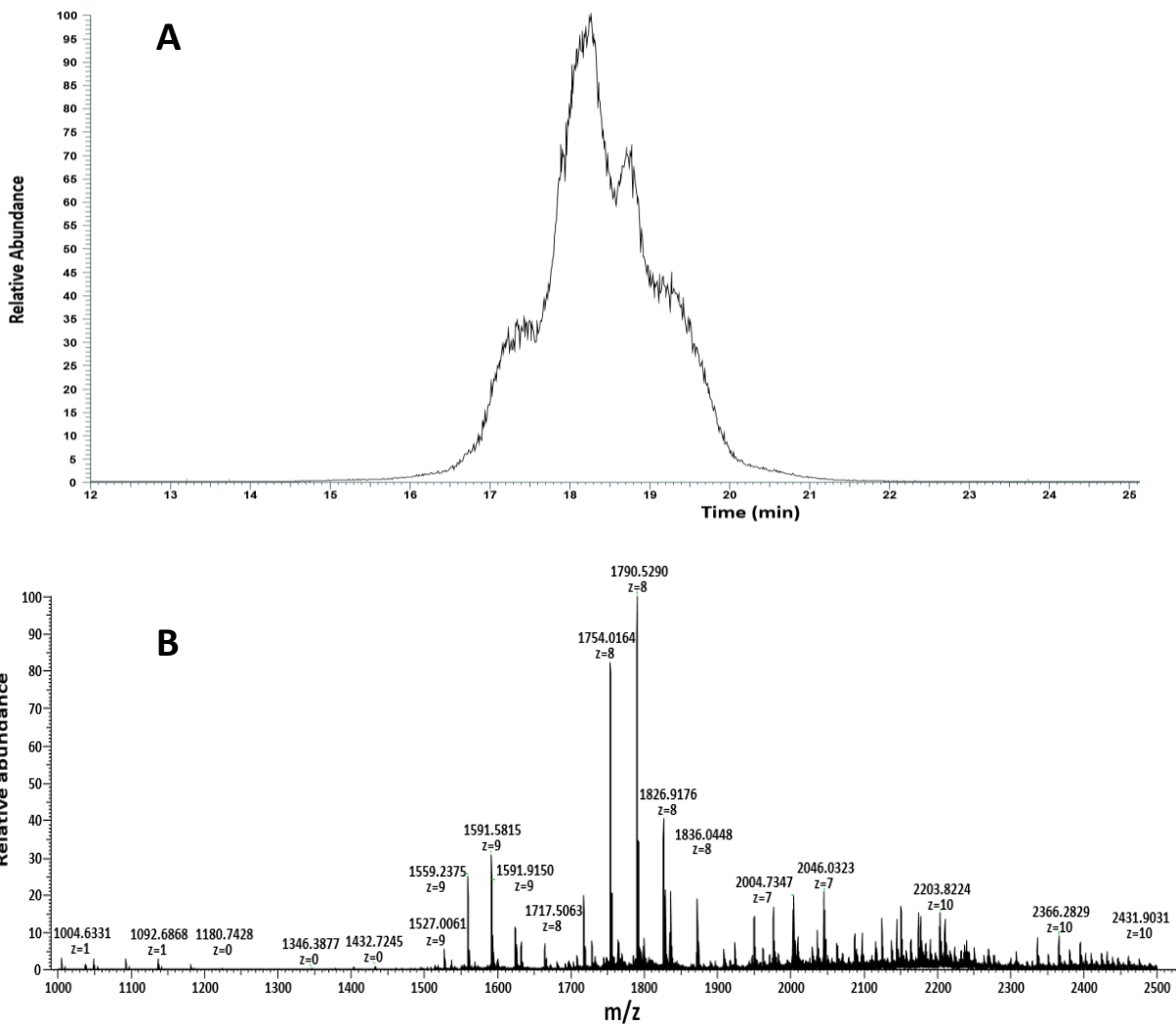


Figure 2

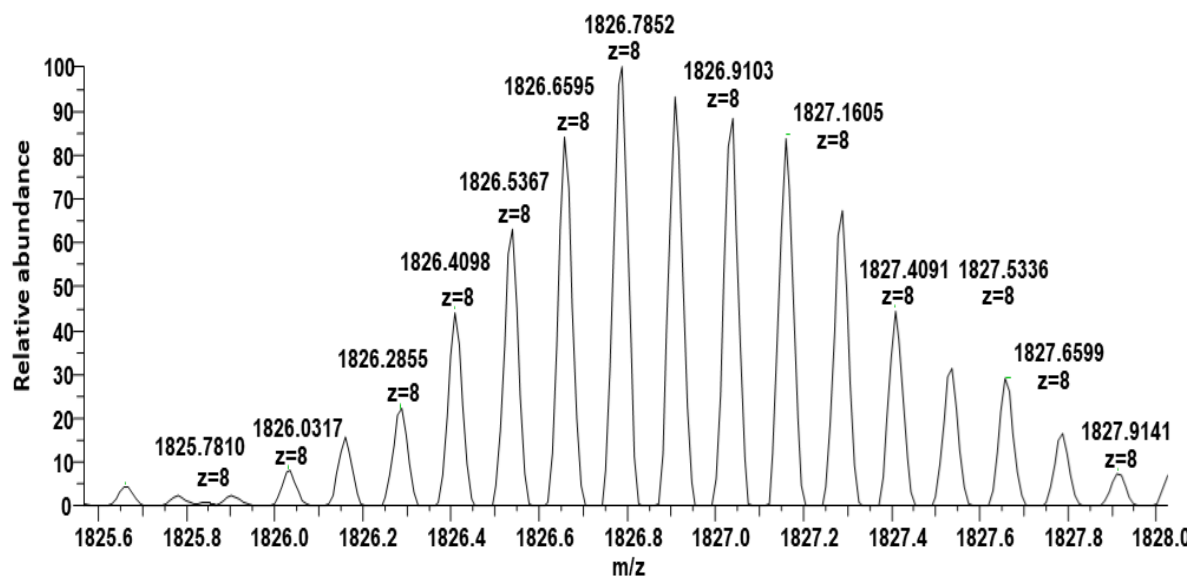




Figure 3

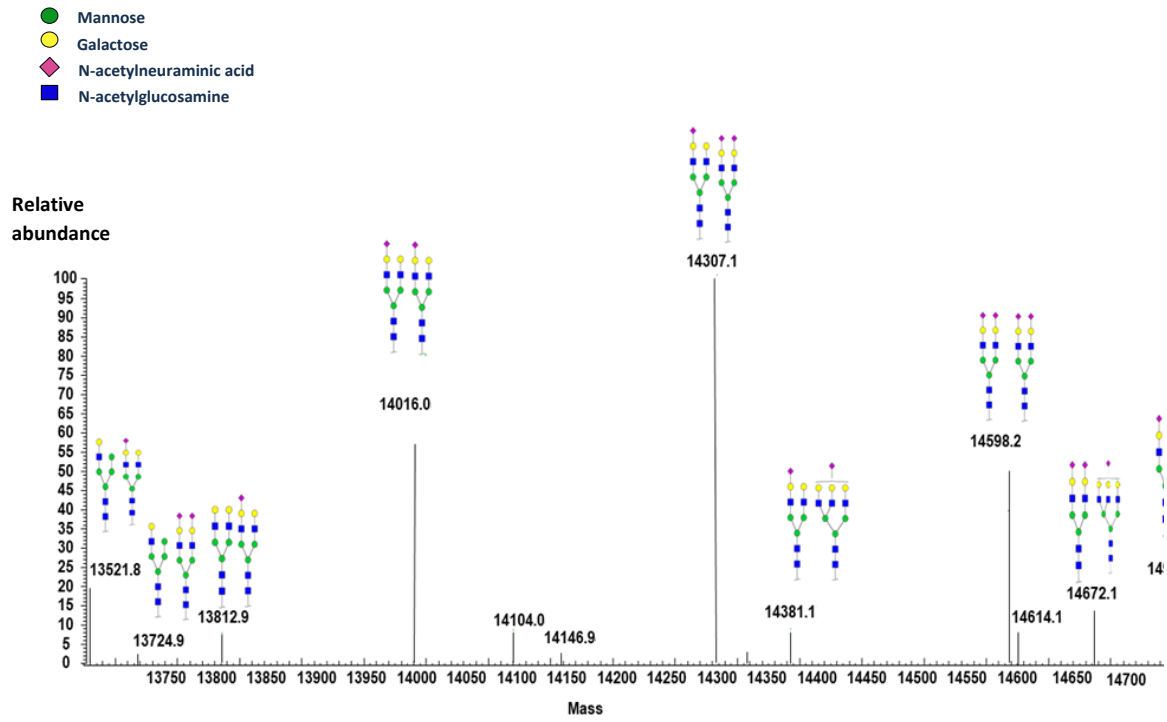


Figure 4

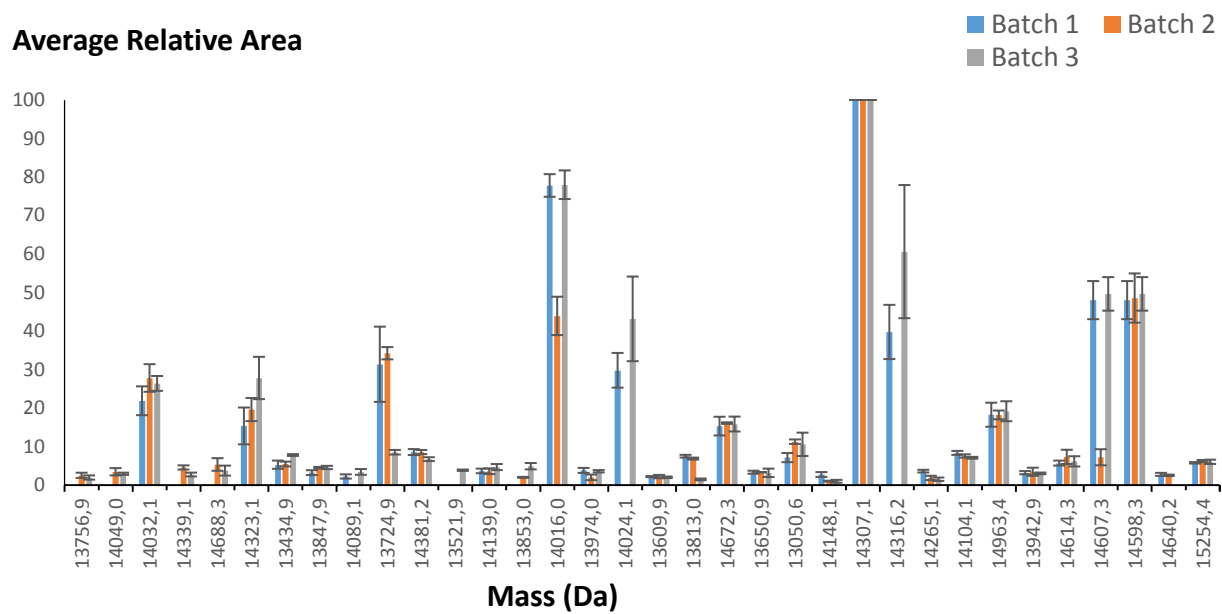


Figure 5

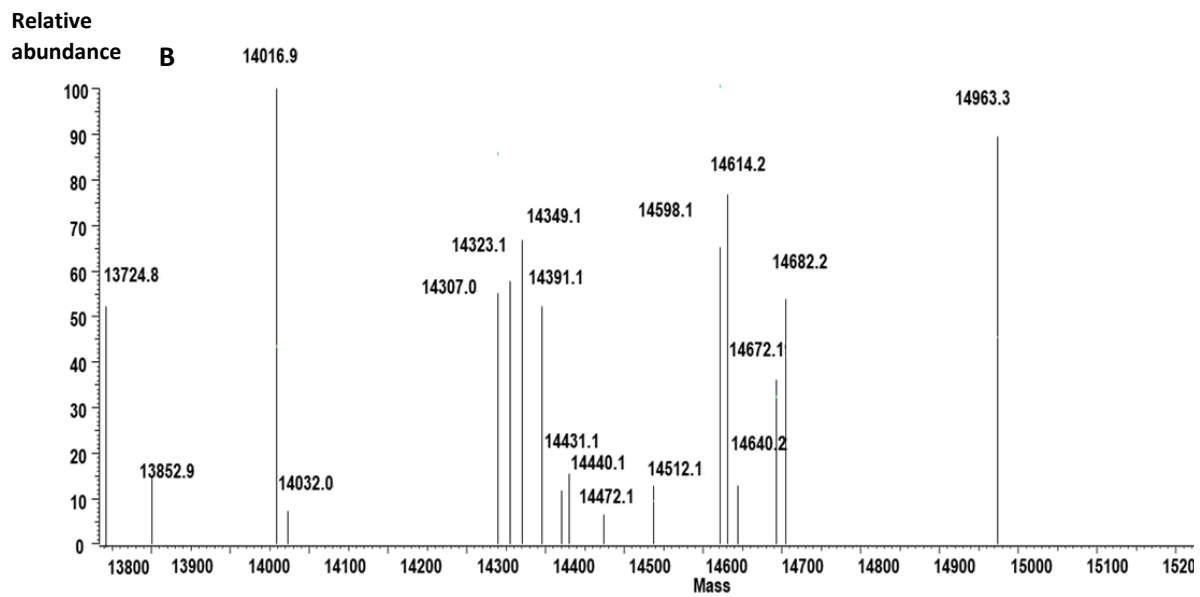
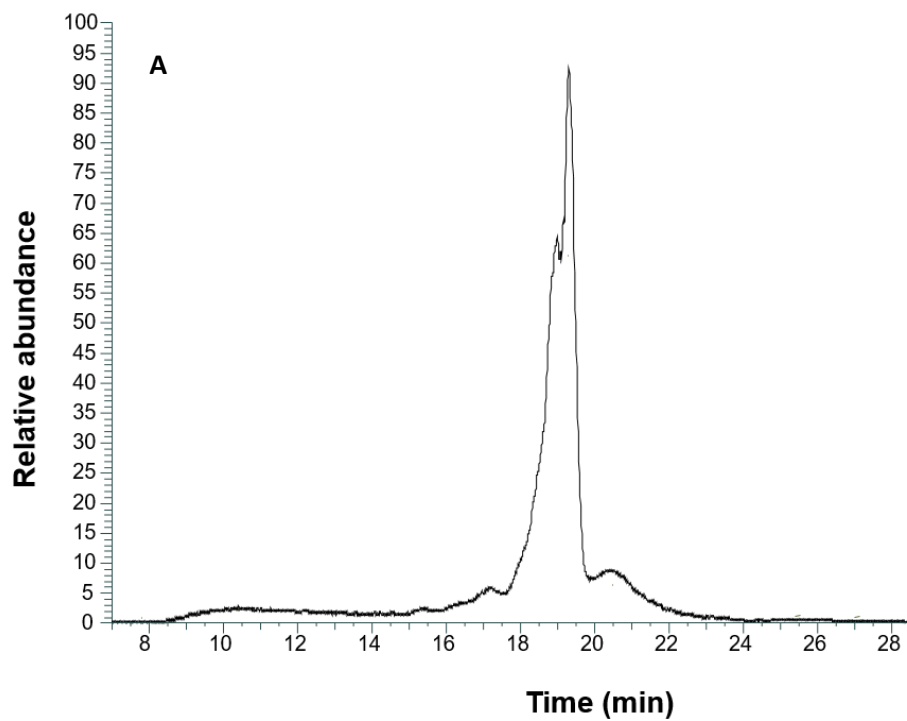


Figure 6

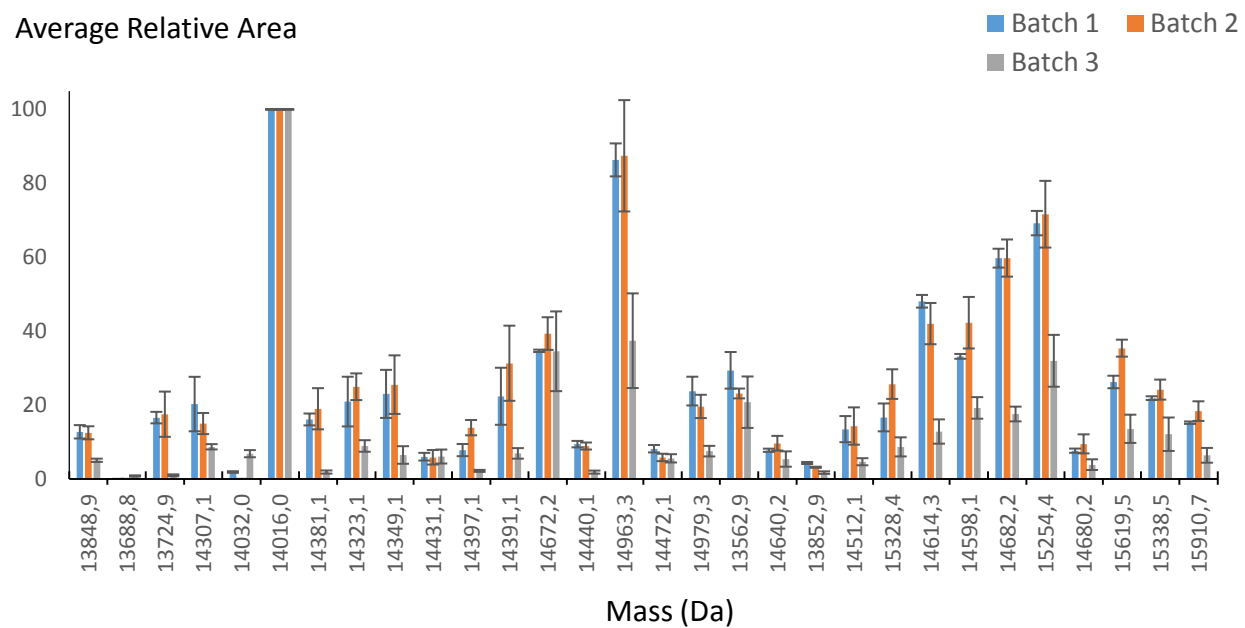


Figure 7

A



B

