



HAL
open science

Development of Immobilized Enzyme Reactors for the characterization of the glycosylation heterogeneity of a protein

Stan Perchepied, Nicolas Eskenazi, Chiara Giangrande, Julien Camperi, Thierry Fournier, Joelle Vinh, Nathalie N. Delaunay, Valérie Pichon

► To cite this version:

Stan Perchepied, Nicolas Eskenazi, Chiara Giangrande, Julien Camperi, Thierry Fournier, et al.. Development of Immobilized Enzyme Reactors for the characterization of the glycosylation heterogeneity of a protein. *Talanta*, 2020, 206, pp.120171. 10.1016/j.talanta.2019.120171 . hal-02330897

HAL Id: hal-02330897

<https://hal.science/hal-02330897v1>

Submitted on 24 Oct 2019

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.

Development of Immobilized Enzyme Reactors for the characterization of the glycosylation heterogeneity of a protein

Stan Perchepied¹, Nicolas Eskenazi², Chiara Giangrande², Julien Camperi¹, Thierry Fournier³,
Joëlle Vinh², Nathalie Delaunay¹, Valérie Pichon^{1,4,*}

¹ *Laboratoire des Sciences Analytiques, Bioanalytiques et miniaturisation – UMR Chimie Biologie Innovation, ESPCI Paris, CNRS, PSL University, 75005 Paris, France*

² *Spectrométrie de Masse Biologique et Protéomique, ESPCI Paris, CNRS, PSL University, 75005 Paris, France*

³ *Laboratory of PhysioPathology and PharmacoToxicology of the Human Placenta, UMR-S 1139 Inserm - University Paris Descartes, Sorbonne Paris Cité, Paris, France*

⁴ *Sorbonne Université, Paris, France*

*Corresponding author at: LSABM-ESPCI, 10 rue Vauquelin, 75005 Paris, France. E-mail address: valerie.pichon@espci.fr (V. Pichon); Tel: +33 140 794 772; +33 140 794 776

Keywords: Immobilized Enzyme Reactor; Glycoproteomics; Glycosylation mapping; Human chorionic gonadotropin; NanoLC-MS/MS

Abbreviations: **ACN**, acetonitrile; **BCA**, bicinchoninic acid assay; **BPC**, base peak chromatogram; **DTT**, dithiothreitol; **E/S**, enzyme to substrate ratio; **EIC**, extracted ion chromatogram; **FA**, formic acid; **Fuc**, fucose; **GalNAc**, N-acetylgalactosamine; **hCG**, human chorionic gonadotropin; **hCG α** , alpha subunit of the human chorionic gonadotropin; **hCG β** , beta subunit of the human chorionic gonadotropin; **Hex**, hexose; **HexNAc**, N-acetylhexoseamine; **IAM**, iodoacetamide; **IMER**, Immobilized Enzyme Reactor; **NeuAc**, N-acetylneuraminic acid; **NeuGc**, N-glycolylneuraminic acid; **MS/MS**, tandem mass spectrometry; **P-IMER**, pepsin-based immobilized enzyme reactor; **PTM**, post-translational modification; **r-hCG**, recombinant human chorionic gonadotropin; **TFA**, trifluoroacetic acid; **T-IMER**, trypsin-based immobilized enzyme reactor; **TRIS**, Trizma hydrochloride; **u-hCG**, human chorionic gonadotropin isolated from urine of pregnant women

Abstract

The mapping of post-translational modifications (PTMs) of proteins can be addressed by bottom-up proteomics strategy using proteases to achieve the enzymatic digestion of the biomolecule. Glycosylation is one of the most challenging PTM to characterize due to its large structural heterogeneity. In this work, two Immobilized Enzyme Reactors (IMERs) based on trypsin and pepsin protease were used for the first time to fasten and improve the reliability of the specific mapping of the N-glycosylation heterogeneity of glycoproteins. The performance of the supports was evaluated with the digestion of human Chorionic Gonadotropin hormone (hCG), a glycoprotein characterized by four N- and four O-glycosylation sites, prior to the analysis of the digests by nanoliquid chromatography coupled to tandem mass spectrometry (nanoLC-MS/MS). Firstly, the repeatability of the nanoLC-MS/MS was evaluated and a method to control the identification of the identified glycans was developed to validate them regarding the retention time of glycopeptides in reversed phase nanoLC separation. The repeatability of the digestion with trypsin-based IMER was evaluated on the same hCG batch and on three independent batches with common located glycans up to 75%. Then, the performance of the IMER digestions was compared to in-solution digestions to evaluate the qualitative mapping of the glycosylation. For the first time, the complementarity of trypsin and pepsin was illustrated for the glycosylation mapping. The potential of IMERs for was illustrated with the comparison of two hCG-based drugs, Ovitrelle® and Pregnyl®.

1. Introduction

The human proteome is composed of a large and dynamic number of proteins as it has been estimated to more than one million proteoforms [1], while the human genome has only approximately 20,000 protein-coding genes [2]. This diversity is partially explained by post-translational modifications (PTMs) that correspond to covalent modifications of chemical moieties on a protein after its translation. These numerous modifications (phosphorylation, acetylation, amidation, glycosylation ...) can alter the activity of the modified proteins and efficient analytical methods are required to elucidate the composition of each proteoform [3–5]. Glycosylation is one of the most challenging PTM to characterize because of its heterogeneity (N- or O-glycosylation) and its high relative molecular weight (up to several kDa) [6]. Still, there is a need to investigate the glycosylation heterogeneity of a protein because it is tightly related to the protein folding, stability or immune response, and thus abnormal glycosylation is often found on the pathways to genetic disorders. Indeed, it was observed, in 2017, that more than half of genetic disorders may be due to abnormal N-glycosylation [7]. Their structure has the advantage to be conserved and well-known: it consists in a common core structure $\text{Man}\alpha 1-6(\text{Man}\alpha 1-3)\text{Man}\beta 1-4\text{GlcNAc}\beta 1-4\text{GlcNAc}$ with GlcNAc-initiated antennas or Man residues bound on the terminal $\text{Man}\alpha$ residues. A sole $\beta 1-4$ -linked GlcNAc residue attached to the β -mannose of the N-glycan core is defined as bisecting. According to the type of antennae, one can distinguish so-called high mannose, hybrid, and complex N-glycans.

The characterization of glycoproteins can be considered by their analysis at the intact level that has the advantage to preserve the integrity of the analyzed proteoforms and the combination of their potential glycosylation sites [8]. Due to the potential high number of glycoforms, they first need to be separated, using of liquid chromatography (LC) or capillary electrophoresis to ensure their ionization in mass spectrometry (MS). Yet, this route faces challenges such as the lack of resolution and their low solubility in organic solvents involved for their separation, and their potential bad ionization in MS [9]. Moreover, high-resolution high-accuracy MS is required to help the identification of such heterogeneous molecules.

An alternative for glycosylation mapping is to use a glycosidase (*e.g.* PNGase F) that cleaves the bond between asparagine and the first GlcNAc (*i.e.* between the peptide and the N-glycan), releasing the N-glycans [3]. N-glycan mapping, or glycomics, will produce a global

view of the N-glycosylation diversity of a single protein, biological fluid, cell type or cell compartment. Note that the N-glycosylation population can be obtained with the use of pronase E that can digest the protein in very small peptides that can be easily analysed for glyco-mapping [10,11]. Yet, one should notice that this approach is very expensive and if more than one glycosylation site is present on the protein, or if the sample contains several proteins, the information about the location is lost.

The third alternative to characterize a glycoprotein is the so-called bottom-up approach and is the most used to date [12]. Each protein is treated to generate proteolytic peptides using endoproteases that are next analyzed and N-glycans can be localized on the detected N-glycopeptides by sequencing using MS/MS fragmentation. The digestion with the protease is carried out during 3 to 16 hours (overnight) in solution with sufficiently high enzyme/substrate (E/S) ratio to decrease the duration of digestion. Yet, autoproteolysis of proteases increases with high E/S ratio, thus explaining the ratio of about 1:50 that can be found in the literature [13,14]. The development of Immobilized Enzyme Reactors (IMERs) overcomes this limitation by immobilizing proteases. High E/S ratio up to 4,000 have been reported [15], leading to a digestion time of less than one hour for greater digestion yields. These IMERs also have the advantage to be reusable and easily automated. Numerous IMERs based on the immobilization of trypsin [4,16–20], and to a lesser extent of pepsin [15,21,22], have been reported in the literature. Yet, the focus was set on the identification of proteins and few information was given on PTMs. Therefore, the performance of the support in terms of digestion yield, repeatability of the digestion, comparison with digestions in solution were evaluated on peptides without consideration of PTMs. The objective of this work was to prepare and use protease-based IMERs to map the glycosylation heterogeneity of a glycoprotein.

We selected a highly glycosylated protein, the human chorionic gonadotropin (hCG), which is the hormone specific to pregnancy. Carbohydrate moieties represent \pm 30% of its molecular weight. hCG is a heterodimer of the α subunit (hCG α) and the β subunit (hCG β) with two N-glycosylation sites on each subunit (respectively noted $^{\alpha}$ N₅₂, $^{\alpha}$ N₇₈, $^{\beta}$ N₁₃, and $^{\beta}$ N₃₀ in order of appearance in the amino acid sequence) and four O-glycosylation sites on hCG β [23,24]. Four hCG β have been reported on Uniprot database (P0DN86, P0DN87, Q6NT52, A6NKQ9). Several studies used an overnight digestion in solution with trypsin for the characterization

of N-glycosylation heterogeneity of a recombinant hCG [13,25]. Recently, Zhu *et al.* were able to map the glycosylation heterogeneity of the digested protein on each glycosylation site. They manually implemented thresholds for the automated validation of the glycopeptides identification. In this study, IMERs were used for the first time for glycosylation characterization and localization, hCG being used as a model glycoprotein. Two complementary endoproteases, trypsin and pepsin, were used to obtain more information on the glycosylation mapping. The performances and potential of the IMERs were evaluated with the digestion of two hCG-based drugs. Validation of N-glycans structures based on both chromatographic and mass spectrometry data was performed after nanoLC-MS/MS analysis of the digest.

2. Materials and methods

2.1. Reagents and analytes

Trypsin from bovine pancreas ($\geq 10,000$ BAEE units mg^{-1} protein), pepsin from porcine gastric mucosa (3,200-4,500 units mg^{-1} protein), cyanogen bromide-activated Sepharose 4B, dithiothreitol (DTT), iodoacetamide (IAM), sodium chloride (99%), Trizma hydrochloride (TRIS) (99%), sodium acetate (99%), sodium azide (99%), glycine (99%), hydrochloric acid (1 M), 0.5 mL centrifugal filters (10 kDa), and bicinchoninic acid assay (BCA) protein assay reagents were purchased from Sigma Aldrich (Saint-Quentin Fallavier, France). Potassium chloride (99.5%), sodium bicarbonate (99.5%), disodium hydrogen phosphate (99%), formic acid (FA) (99%), and potassium dihydrogen phosphate (99.5%) were ordered from VWR (Fontenay-sous-Bois, France). LC/MS Grade acetonitrile (ACN) and trifluoroacetic acid (TFA) were purchased from Fischer Chemical (Illkirch, France). Deionized water was obtained from a Milli-Q purification system (Millipore, Molsheim, France).

Two hCG-based drugs were used: Ovitrelle[®] (Organon, Oss, The Netherlands) and Pregnyl[®] (Serono Europe Ltd, London, UK). Ovitrelle[®] (recombinant protein from Chinese Hamster Ovary cells, r-hCG) was conditioned as a 500 $\mu\text{g mL}^{-1}$ solution in water. Pregnyl[®] (hCG purified from urine of pregnant women, u-hCG) was obtained as a lyophilized powder (5,000 UI). Considering that 1 UI corresponds to 0.092 μg [26], the drug was solubilized to a final

concentration of $500 \mu\text{g mL}^{-1}$ in water. The hCG samples were aliquoted and stored at -20°C for later use.

2.2. Preparation of IMERs

The procedure was adapted from previous studies from our group [15,27]. Briefly, 40 mg of Sepharose was swollen with 2 mL of HCl (1 mM). The Sepharose was washed twice either with 1.8 mL of bicarbonate buffer (NaHCO_3 0.1 M, NaCl 0.5 M, pH 8.3) for trypsin immobilization or 1.8 mL of acetate buffer ($\text{CH}_3\text{CO}_2\text{Na}$ 0.1 M, NaCl 0.5 M, pH 5.8) for pepsin immobilization. 1.7 mL of enzyme solution (4 mg mL^{-1}) was incubated for 18 h at 4°C . Then, the enzyme-coupled Sepharose was introduced into a precolumn (2 mm i.d. x 20 mm, $2 \mu\text{m}$, Upchurch, Interchim, Asnières, France). The supernatant was kept for further evaluation of the grafting yield.

Trypsin-based IMER (T-IMER) was washed with 48 mL of TRIS buffer (0.1 M, pH 8.0). Acetate buffer ($\text{CH}_3\text{CO}_2\text{Na}$ 0.1 M, NaCl 0.4 M, pH 4.0) and bicarbonate buffer were successively percolated through the IMER thrice to remove unbound trypsin. Finally, the IMER was stored at 4°C in PBS buffer (NaCl 10 mM, Na_2HPO_4 10 mM, KCl 1 mM, KH_2PO_4 1 mM, NaN_3 0.1% (w/w), pH 7.4).

Pepsin-based IMER (P-IMER) was washed with 48 mL of glycine buffer (0.2 M, pH 5.0). Acetate buffer ($\text{CH}_3\text{CO}_2\text{Na}$ 0.1 M, NaCl 0.5 M, pH 5.8) and an acidic solution (HCl 0.1 mM, NaCl 0.5 M) were successively percolated through the precolumn thrice to remove unbound pepsin. Finally, the IMER was stored at 4°C in FA solution (pH 3.0, 0.2% v/v).

The amount of bound proteases was evaluated with an indirect bicinchoninic acid assay [28]. The calibration curves (5 points) were prepared from 0.1 to 2 mg mL^{-1} of each protease. $10 \mu\text{L}$ of supernatant were introduced in a well of a 96-microwell plate. Then, $100 \mu\text{L}$ of working reagent (bicinchoninic acid and copper (II)) were added. The reaction was performed during 30 min at 37°C under agitation at 300 rpm. After cooling down at room temperature, the absorbance was measured at 562 nm in every well with a UV-spectrophotometer (SpectraMax M2, Molecular Devices, St Gregoire, France). The grafting yield was determined by the ratio between the amount of protease recovered in the supernatant with the amount introduced.

2.3. Sample preparation

The two hCG-based drug stock solutions (r-hCG and u-hCG) were first filtered on Microcon 10 kDa membrane filter for 45 min at 12,000 G to remove small size additives from drug formulation. 150 μ L of TRIS buffer (TRIS 50 mM, CaCl₂ 10 mM, pH 8.0) was added twice to clean the drug. The proteins were reduced in 5 mM DTT at 56°C for 40 min and alkylated in 25 mM IAM at room temperature for 20 min in the dark. Denatured hCG was washed three times on centrifugal membranes (5 min at 16,800 G) with the digestion buffer of each protease: TRIS buffer (TRIS 50 mM, CaCl₂ 10 mM, pH 8.0) for trypsin, FA pH 2.2 (0.25% v/v) for pepsin was added. The final volume was adjusted with the same buffer to get a final concentration of hCG of 500 μ g mL⁻¹.

2.4. Digestion procedures

2.4.1. Digestion on IMERs

The protocol of digestion was adapted from previous studies from our group [15,27]. Briefly, 2.5 μ g of hCG were introduced in a 5 μ L sample loop set on a six-port switching valve connected to an isocratic pump (LC-10 AS, Shimadzu, Champs-sur-Marne, France) and to the IMER set in an independent oven (Croco-cil oven, Interchim, Montluçon, France). Then, the protein was transferred on the IMER with the digestion buffer at 50 μ L min⁻¹. The digestion was carried-out during 30 min at 37°C using a « stop-flow » approach. Finally, 100 μ L of the digestion buffer were percolated to collect both peptides and glycopeptides.

2.4.2. Digestion in solution

Samples were digested in solution in the same buffers as IMERs with an enzyme-to-substrate ratio of 1:100. After overnight incubation at 37°C in a Thermomixer (Eppendorf, France), the reaction was quenched with 1 μ L of TFA. The digests were subsequently diluted 20-folds in either TRIS buffer for trypsin digests, or FA buffer for pepsin digests, to match the dilution introduced by the IMER protocol. Digests were stored at 4°C before injection until further analysis.

2.5. LC-MS analysis

The nano-LC separation of the digests was achieved on a RSLCnano UltiMate 3000 system (ThermoFisher Scientific, Le Pecq, France). Peptides were loaded and desalted onto a C18

trapping column (Acclaim PepMap100, 300 μm i.d. x 5 mm, 5 μm , ThermoFisher Scientific, Le Pecq, France) at a flow rate of 15 $\mu\text{L min}^{-1}$ in buffer C: 0.05 % TFA in $\text{H}_2\text{O}/\text{ACN}$ 98/2 (v/v) for 5 min, before to be transferred to a C18 column (Acclaim PepMap100 C18, 75 μm i.d. x 50 cm, 3 μm , 100 \AA , ThermoFisher Scientific) at a flow rate of 220 nL min^{-1} with a gradient composed of A (0.1 % FA in $\text{H}_2\text{O}/\text{ACN}$ 98/2 (v/v)) and B (0.1 % FA in $\text{H}_2\text{O}/\text{ACN}$ 10/90 (v/v)) with 2 to 40 % B in 60 min followed by a rapid increase to 90 % B in 1 min. The detection was carried out with a nano-ESI QqOrbitrap hybrid mass spectrometer (Q Exactive, ThermoFisher Scientific) with the top10 DDA mode: one full MS scan (m/z range 400-2000, resolution 70,000 at 200 m/z , maximum injection time 100 ms, AGC target 3e6) followed by ten MS/MS acquisitions (Top 10, isolation windows 2 m/z , maximum injection time 120 ms, AGC target 1e5, fixed first mass 90 m/z , resolution 17,500 at 200 m/z , normalized collision energy 25, systematic exclusion of $z=1$ ions and dynamic exclusion of 10 s).

2.6. Data analysis

The glycopeptide identification and annotation were performed using PMI-ByonicTM (build 2.15.7), on a restricted database consisted of 5 annotated sequences from Swiss-Prot database associated to hCG sub-units (UniProt accession numbers: P0DN86; P01215; P0DN87; A6NKQ9; Q6NT52). The glycan database was the ByonicTM "N-glycan 309 mammalian no sodium" database. Cleavage residue was set to: i) C-terminal cutter of KR aminoacids allowing up to 2 missed cleavages for trypsin digestions; ii) non-specific for pepsin digestion. Error tolerance was set to 10 ppm in MS mode and 20 mDa in MS/Ms mode. Modifications were set as follows: Carbamidomethylation (Nterm, C, K), Oxidation (M), and Deamidation (N, Q) as common1; N-glycan (N) as rare1. Maximum number of modifications set to 2 common modifications, 1 rare modification.

3. Results and discussion

3.1. Development of IMERs

3.1.1. Selection of the enzymes

The aim of this work was to study the potential and the complementarity of protease-based IMERs to fasten and improve the reliability of N-glycosylation heterogeneity mapping by

testing them with a model glycoprotein, *i.e.* hCG, a four times N-glycosylated protein. Trypsin was chosen because it is the most common protease for bottom-up proteomics. It has well-known cleavage sites at the C-terminal side of two amino acids (arginine and lysine) that produces peptides with protonable C-terminus amino-acid, and has a high proteolytic activity while being very stable under various solvent conditions. A computer simulation of expected peptides was carried out with ExPASy server (<https://www.expasy.org/>) for hCG α and all four human gonadotropin hCG β (Table S.1). It is worth noting that the algorithm of the software did not take into consideration R.P trypsin cuts (« Keil rules »). Yet, Rodriguez *et al.* [29] have shown that this cleavage site is not less recurrent than R.C that is considered by the server. Therefore, even if not considered by the software, R.P cuts were further investigated.

One tryptic cleavage site was close to $^{\alpha}\text{N}_{52}$. To avoid loss of information caused by steric hindrance due to glycosylation, pepsin was also immobilized on a second support because pepsin and trypsin do not have the same specificity. Pepsin cleaves proteins at C-terminal side of hydrophobic amino acids [21].

3.1.2. Preparation of the IMERs

The selected enzymes were then grafted on Sepharose because this solid support has low non-specific interactions with proteins according to a procedure already described by our group [15,27]. With evaluation of the amount of enzymes in the supernatant by BCA, the grafting yield of trypsin was estimated at 34%, giving a grafting density of trypsin in the precolumn up to 1.40 nmol μL^{-1} . Meanwhile, pepsin was grafted on Sepharose beads at 25%. Thus, the grafting density of pepsin was estimated at 0.78 nmol μL^{-1} .

It is worth noting that these grafting yields are consistent with our previous results: trypsin and pepsin were immobilized on the same support with a grafting yield of 30% [27] and 24% [15], respectively. In all cases, the grafting yield of trypsin was found higher than the one of pepsin for the same amount of proteases incubated for immobilization. This difference can be explained by the difference of molecular mass between the two (26.6 and 34.6 kDa, respectively). Indeed, as trypsin has a lower molecular mass, the kinetic of diffusion is further increased and the steric hindrance for its immobilization is lower.

3.2. Method developed for the identification of N-glycans by LC-MS in the digest

To evaluate the potential of these IMERs for the glycosylation mapping of hCG, it was first necessary to develop a data processing of the chromatograms and the MS spectra. For this, a first digestion was carried out using the T-IMER by transferring a solution of r-hCG (5 μ L, 2.5 μ g) on the IMER in the conditions previously described by our group for the digestion of Cytochrome C [27]. However, unlike cytochrome C, hCG has five disulfide bonds on hCG α and 6 on hCG β . Therefore, a reduction /alkylation step was carried out to hCG prior to its enzymatic digestion to try to prevent from miss-cleavages. The digestion was further performed in stop flow mode according to previous studies by our group to increase the digestion time and thus the digestion yield. Considering the relative molecular weight of the protein to digest a digestion time of 30 min was chosen.

The digest was analyzed in nanoLC-MS/MS with a routinely used method in proteomics involving a C18-based stationary phase and a gradient from 2% of ACN up to about 40% and next 90%. The nanoLC format lowers the limits of detection of peptides in ESI-MS by reducing the diameter of the separation column. The detection of lower concentrations was also favored by a C18 trapping precolumn that allows the injection of high volumes of sample (up to 2 μ L in this work) on the nanoLC-MS device. The repeatability of the LC-MS analysis was studied and the relative standard deviations (RSD) of the retention times of the most intense peaks were below 0.5%, thus exhibiting a high repeatability of the analysis.

Data processing was performed using ByonicTM software. ByonicTM result files display glycan composition as unresolved stereoisomers such as "Hex" or "HexNAc". Knowing that N-glycan biosynthesis is highly conserved in humans, we completed the structural identity of those monosaccharides to be either galactose (Gal), Mannose (Man) or N-acetylglucosamine (GlcNAc). Manual validation of the MS/MS spectra (in regard of the quality of the fragmented spectrum and the peak annotations) allowed us to propose a scoring threshold for each glycosylated peptide. The parameters chosen for the validation of glycopeptide identification are summarized in Table 1. As it is known, monosaccharides like fucose can rearrange between antennas in CID [30]. While we did not perform fucose-specific linkage analysis (such as low collision energy MS/MS), our assumption of core fucosylated glycans was often supported by Y-ion Fuc-GlcNAc-peptide such as in the spectrum provided in supplementary material (Figure S.1), matching annotated human fucosylated N-glycans.

Table 1. *ByonicTM score threshold used for the validation of the glycopeptide identification.*

Enzyme	Glycosylation site	Byonic TM Score
Pepsin	α N ₅₂	> 150
	β N ₁₃	> 80
Trypsin	α N ₅₂	> 30
	α N ₇₈	> 15
	β N ₁₃	> 10
	β N ₃₀	> 150

To ensure the identification of glycan moieties, we combined the MS and the chromatographic data. As the retention time in reversed-phase LC of a glycopeptide is mostly driven by the peptide hydrophobicity, the combination of one given glycosylation site and one given peptide was considered. All the glycans that have been identified as bound onto these particular peptides by the software formed glycopeptides whose retention times were studied by taking into account their respective Extracted Ions Chromatograms (EIC). For instance, EICs of glycopeptides from β N₁₃ (CRPIN₁₃LAVEK) are shown on Figure 1. It was observed that the addition of a N-acetylneuraminic acid (NeuAc) on the glycan causes an increase in the retention time of the glycopeptides (*k* increases by 20% from the monosialylated to the bisialylated glycopeptide). This phenomenon in reversed phase mode was already reported in the literature and the explanation proposed was that: (i) the hydrophobicity of the glycopeptide containing an additional NeuAc increases due to the presence of an acetoamide group [31]; (ii) there are some interactions between the acidic moieties of the monosaccharides and active sites on the stationary phase as the same phenomenon was observed both in reversed phase and HILIC mode [32,33].

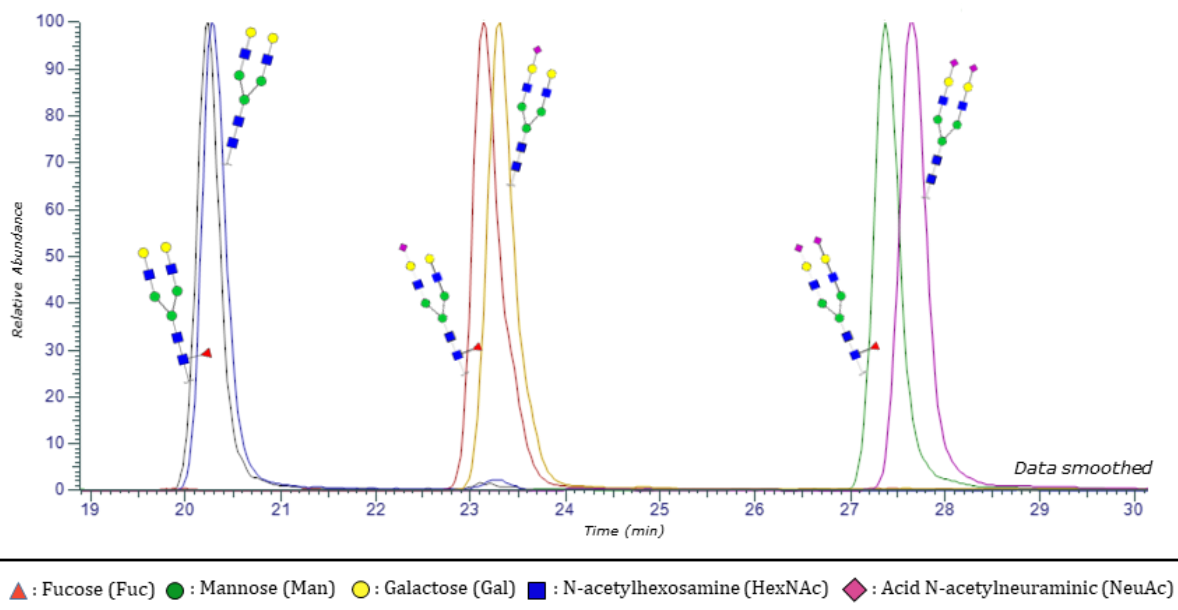


Figure 1. Extracted Ion Chromatograms of glycoforms of CRPIN₁₃ATLAVEK sequence (⁶N₁₃) produced from r-hCG digestion with T-IMER and analysis by LC-MS/MS exhibiting the influence of sialic acids and fucose on the retention time of the glycopeptide (only its glycan moiety was represented).

Figure 1 also shows the influence of the addition of a fucose (Fuc) that decreases, to a lesser extent (k decreases by 1.5% for the biantennary bisialylated glycopeptide), the retention time of the glycopeptides. The presence of three hydroxyl groups with fucose should contribute to increase the hydrophilicity of the fucosylated glycopeptides [33]. Similarly, the influence of the addition of hexose (Hex) and N-acetylglucosamine (GlcNAc) on the retention time of the resulting glycopeptides was studied. It was shown that the retention time decreases with the addition of one of these neutral monosaccharides because of their hydrophilicity.

Therefore, taking into account the retention time of glycopeptides should help to confirm or rule out some identifications made by the software. As an example, a fragmentation spectrum that has been attributed to the NV₅₂TSESTCCVAK (^αN₅₂) peptide with the N-glycan HexNAc₄Hex₆Fuc₁ (G1) was identified by the software at 17.33 min. Yet, the presence of NeuAc was strongly suggested by the identification of a diagnostic fragment ion corresponding to the trisaccharide GlcNAc₁Gal₁NeuAc₁ (m/z 657.59 Da). The EIC of G1 has shown that its retention time was the same than the same peptide glycosylated with GlcNAc₄Gal₂Man₃NeuAc₁ (G2) (see Figure 2A). As the addition of a NeuAc strongly increases the retention time of the associated glycopeptide, both glycopeptides cannot result from the digestion of our sample. As the retention time was close to the one of monosialylated

glycopeptides, only the presence of G2 was confirmed. Moreover, the MS/MS spectrum did not show any confirmation of the presence of fucose.

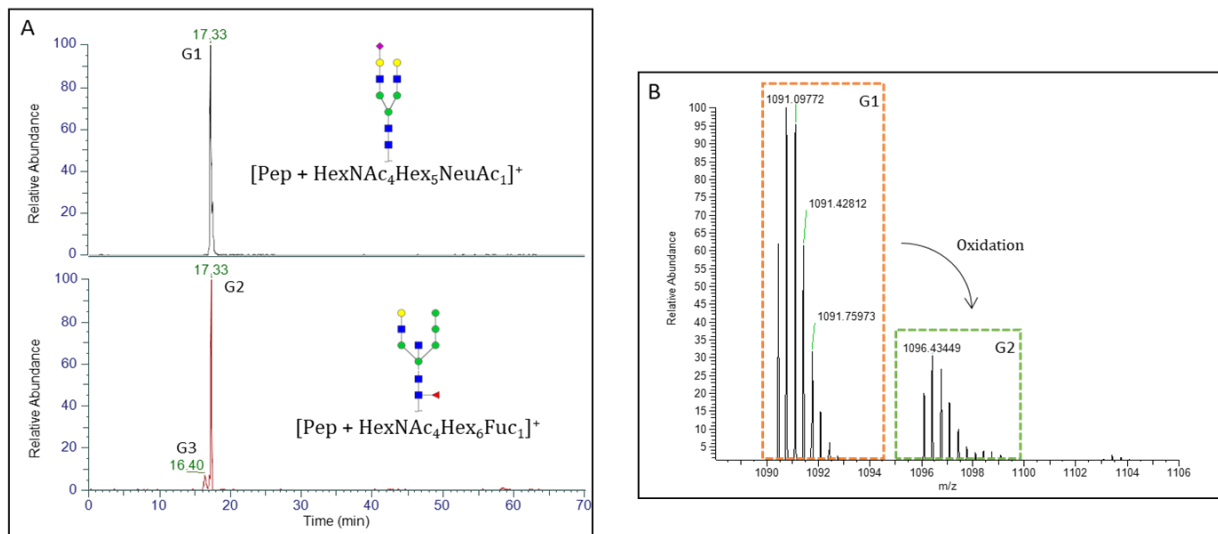


Figure 2. (A) Extracted Ions Chromatograms of two glycopeptides identified on α N52 by ByonicTM software obtained by the nanoLC-MS/MS analysis of the T-IMER digest of r-hCG. (B) Isotopic MS profile combined from 16.08 to 16.6 min.

This misidentification may come from the software that did not take into account an oxidation either on the peptide or on the NeuAc (accounting for the 16 Da mass shift), thus misidentified the glycopeptide as the molecular weight of a Fuc and an Hex corresponds to the one of a NeuAc and an oxidation with water. Indeed, the isotopic MS profile confirmed this oxidation thanks to similar isotopic signature (see Figure 2B).

On Figure 2A down, no fragmentation spectrum was found for the peak of G3, the observed isomer ($t_r = 16.40$ min). By taking into account its retention time (Figure 2A) and its isotopic signature, this compound could be identified as an isobar of G2 with the same peptidic and glycosidic moiety composition, but having another oxidized function. The difference in retention time could also be explained by a difference on the linkage of the neuraminic acid (e.g. α 2,3 or α 2,6 linkage) [34]. As the contribution of the location of NeuAc on the retention time is not well understood, the identification has not been confirmed. It is possible to add a derivatization step to differentiate them by MS. Linkage-specific derivatization is mostly achieved on release glycans rather than proteolytic glycopeptides because it also induces peptide modification. This could be very informative when performed between digestion and LC-MS/MS. Furthermore, recent publication [35] shows that oxonium ions alone can also be informative on sialic acid linkage, thus eliminating the need of linkage-specific

derivatization. However, the full characterization of the glycan composition was not judged necessary to compare sample handling and digestion protocols on IMERs.

Therefore, G3 was not taken into consideration, although ByonicTM identified this glycopeptide with a score high enough to pass our first validation step. Adding the retention time enabled to eliminate false positives from the identified glycopeptides in agreement with previous reports [36–38].

Similarly, another kind of doubtful identification was proposed by ByonicTM. For instance, a glycan composition corresponding to HexNAc₄Hex₇ on the peptide containing ^aN₅₂ was proposed. Yet, the fragmentation spectrum and the analysis of the retention time led to think that a sialic acid was also present on the glycan. Knowing that the molecular weight of two Hex corresponds to the one of a sialic acid and two oxidations with water, it led to the identification of the HexNAc₄Gal₂Man₃NeuAc₁ glycan with two oxidations located either on the peptide or on the glycan. As the previous misidentification, the glycopeptide was not considered for further data processing because of the ambiguity of the location of the oxidation either on the peptide or of the NeuAc.

3.3. Evaluation of the trypsin-based IMER

3.3.1. Repeatability of the digestion for the glycosylation mapping

Once having developed this method to control the acceptability of the identification provided by ByonicTM, the potential of protease-based IMERs for the mapping of glycosylation was studied for the first time. As the digestion efficiency of protease based IMERs for the glycosylation mapping was not known, a digestion time of 30 min was chosen [15]. In order to compare digestions on IMERs and in solution, the temperature of the digestion was set at 37°C on IMERs.

T-IMER was first studied because it is based on the most used protease. It was used to digest r-hCG. The analysis of the digest was performed in triplicate. A chromatogram of the resulting digestion is available in Figure S.2A. All the identified glycans in the digest with the previously developed method are reported in Table S.2 in the left column. Some different results were observed between the three analysis runs (R1, R2, and R3) of the same digest of the batch 1 (B1) of r-hCG. Hence, only glycans identified on a specific location in at least two out of three runs were further considered as relevant for a given digestion, and most of

misidentifications or unresolved compositions (uncertainty between NeuGc and peptidic oxidation) were discarded while focusing on the most frequently detected glycopeptides. No glycopeptide identified encompasses the $^{\alpha}\text{N}_{78}$ glycosylation site due to potential steric hindrance of the carbohydrate moiety. Moreover, the identification of glycopeptides from $^{\beta}\text{N}_{30}$ was not possible because of the lack of sufficient data from the fragmentation spectra as the resulting peptide on this site is heavy

The performance of the digestion with T-IMER was then evaluated by carrying out the digestion of a batch of r-hCG drug (batch 1) in triplicate (D1, D2, and D3) and each digest was analyzed in triplicate in nanoLC-MS/MS. To study the repeatability of the digestion, the identified glycans were considered (see Table S.2, in the mid column). No glycan identified on $^{\alpha}\text{N}_{52}$ was fucosylated while glycopeptides from $^{\beta}\text{N}_{13}$ have a higher variability in terms of identified glycans (31 different glycans). For the three digestions, the most intense peaks were attributed to biantennary and triantennary complex N-glycans. The three independent digestions of the r-hCG batch 1 gave 81% of common N-glycans out of 42 identified, exhibiting good performances in terms of repeatability of the T-IMER digestions for the glycosylation mapping.

It is noteworthy that all glycans identified on $\text{PIN}_{13}\text{ATLAVEK}$ ($^{\beta}\text{N}_{13}$) were also identified on $\text{CRPIN}_{13}\text{ATLAVEK}$ ($^{\beta}\text{N}_{13}$). Yet, further analyses were only performed with the latter even though it contains a missed cleavage, because more glycans were identified on this peptide.

It should also be specified that peptides resulting from the trypsin auto-digestion were found in the first digests after the IMER synthesis. The intensity of two of those peptides (SSGTSYPDVLK and LGEDNINVVEGNEQFISASK) was then monitored through the different digests and it was observed that it decreased by a 1.6 factor between the first digestion and the third. This led to think that the post-immobilization washing steps could be further optimized to remove non-grafted trypsin from the IMER before its use.

To further study the repeatability of the digestion on T-IMER, three independent batches of hCG were considered (B1, B2, and B3). The analysis of each digest was only achieved in duplicate to save time. Therefore, in order to stick at the method previously developed, the presence of a glycopeptide was validated if identified in two out of the two runs. The

identified glycans can be found in Table S.2 in the right column. Common located N-glycans were calculated at 75% out of 40 identified, exhibiting an inter-batch repeatability of the digestion with T-IMER for the glycosylation mapping. The lower repeatability for inter-batch than a unique batch may be explained by the analysis of the digests only in duplicate. Indeed, some glycans may be excluded during the validation process, while a third run might have allowed to confirm their presence. Small variations between hCG batches may have also contributed to this lower repeatability.

3.3.2. Comparison of T-IMER with in-solution digestion

After the evaluation of the repeatability of the digestion on T-IMER in terms of number of glycans identified on a specific location, the performance of T-IMER digestion was compared to conventional in-solution digestion using the same sample of hCG to confirm the benefit of IMERs. As the digestions on IMERs and in solution were carried out with 2.5 µg of hCG, by considering an average molecular mass of 35 kDa [24] and taking into account the grafting yield, the E/S ratio was evaluated up to 1,430 for T-IMER. This E/S ratio was roughly 150,000 fold higher than the one of digestion in solution (1:100), thus explaining the short digestion time used with IMERs (30 min).

In order to inject the same quantity of hCG between the two digestion modes, the digest in solution was diluted twenty fold to match the concentration resulting from the digestion on T-IMER. The results in terms of numbers of glycans identified on a specific location are presented in Figure 3. Both digestions shared 93.5% of common N-glycans (42 out of 45) with only a 30 min digestion on IMER instead of overnight in solution. Even though, the loss of triantennary glycans was noticed on $^{\alpha}\text{N}_{52}$. It is worthwhile to notice that both digestion modes involving trypsin led to identified glycans only on $^{\alpha}\text{N}_{52}$ and $^{\beta}\text{N}_{13}$ glycosylation sites and no glycan identified on $^{\alpha}\text{N}_{78}$ and $^{\beta}\text{N}_{30}$. Once again, the most intense peaks were attributed to glycopeptides containing biantennary and triantennary complex N-glycans for both digestion modes. In the end, the digestions on T-IMER led to similar characterization of the glycosylation heterogeneity of r-hCG and the same peptides, without missed cleavages (R.P cuts were considered insignificant), were identified. Nevertheless, it should also be noted that to avoid autoproteolysis, modified proteases can be used. Yet, such enzymes are usually more expensive, so the immobilization of trypsin could lead to their reusability, which would dramatically decrease the price of a digestion.

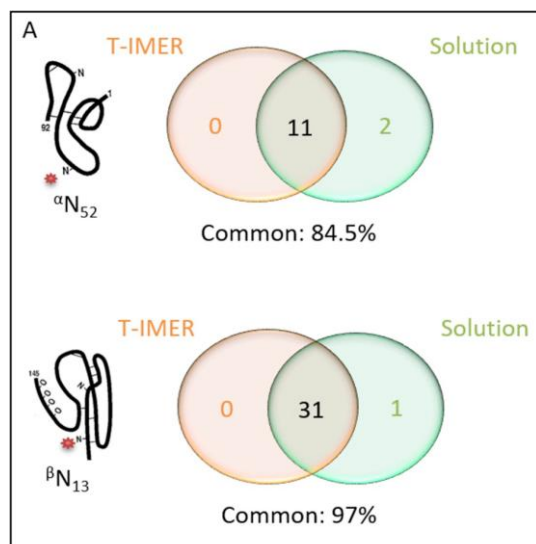


Figure 3. Comparison of the numbers of N-glycans identified on a specific glycosylation site after the digestion of r-hCG with trypsin-based IMER and in solution.

3.4. Digestions with pepsin-based IMER

The digestions with T-IMER gave partial information on the N-glycosylation heterogeneity of the hCG because of the absence of glycans identified on two N-glycosylation sites (αN_{78} and βN_{30}). To complete this characterization, a digestion with P-IMER was considered. The repeatability of the pepsic digestion was already studied by our group [15]. Therefore, P-IMER was used to carry out the digestion of an hCG batch (batch 1). A chromatogram of the resulting digestion is available in Figure S.2B. The comparison of the peptides with a glycosylation site successfully identified after digestion with T-IMER and P-IMER is presented in Table S.3. The digestion with P-IMER led to more heterogeneous glycopeptides for one glycosylation site than with T-IMER, as the proteolysis site-specificity of pepsin is lower than that of trypsin. Besides, the comparison of the glycosylation mapping obtained using T-IMER and P-IMER is depicted in Figure 5. It should be mentioned that the structure propositions are primarily based on the sugar composition inferred from common annotated structures and fragmentation spectra. One should notice that for αN_{52} and βN_{13} , there are more glycans identified with trypsin than pepsin due to the overall variability in terms of peptide that tends to decrease the signal intensity of each glycopeptide since one given glycan can be present on several peptides varying in length. Therefore, due to the lower signal-to-noise ratio, the probability to have an informative fragmentation spectrum decreases. Nevertheless, the digestion with P-IMER allowed the identification of glycans on αN_{78} and βN_{30} , thus highlighting the complementary of trypsin and pepsin. Moreover, common glycans

localized on a specific N-glycosylation site were evaluated down to 39% between both enzymes, highlighting the complementarity of the two enzymes to identify a maximum number of N-glycans. Although this was only a qualitative study, the results were also compared to what can be found in literature [39,40]. It was reported that hCG contains mostly complex glycans, especially hCG β . In this study, the combination of pepsin and trypsin allowed the identification of 48% of complex glycans on hCG β .

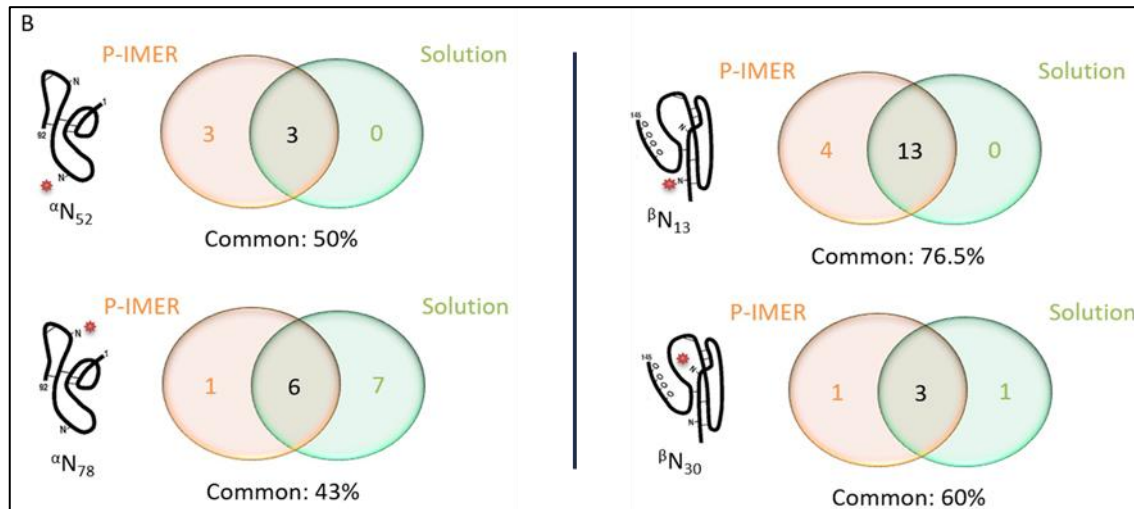


Figure 4. Comparison of the numbers of N-glycans identified on a specific glycosylation site after the digestion of r-hCG with pepsin-based IMER and in solution.

The performance of the digestion with P-IMER was also compared with a pepsin digestion in solution. As for T-IMER, the E/S ratio was evaluated and a value up to 800 was calculated for P-IMER. Even if the E/S ratio is lower than for T-IMER, it is still 80,000 fold higher than the ratio used for digestion in solution (1:100). The results in terms of glycans identified on a specific location are presented in Figure 4. Noticeably, according to the glycosylation site, either IMER or in-solution digestion was more efficient. As an example, IMER led to the identification of three additional glycans on α N₅₂, whereas the digestion in solution led to seven additional glycans on α N₇₈. Unlike tryptic digestions, both digestions shared only 59.5% of common glycans localized on a specific N-glycosylation site. Among the common identified glycans located on a given site, biantennaries and triantennaries were the most intense. Yet, less intense glycans (*e.g.* hybrids and high mannose) were specifically identified with the pepsin digestion either in-solution or with P-IMER (see Figure 4). This was surprising because pepsin digestion provides more variable peptides that get the glycans, so there is a possibility of decrease of signal for one site of glycosylation. One explanation could be the

specificity of each enzyme to cleave the protein with potential steric hindrance of the sugar moieties.

As pepsin and trypsin are complementary, it would have been interesting to immobilize both enzymes on the same support. Yet, the optimum pH for digestion with trypsin is 8.0 while it is 2.0 for pepsin. Therefore, if co-immobilization is considered one of the enzymes would be inactive. Moreover, the activity of pepsin is decreasing rapidly and is practically zero above a pH of 5.5. Yet, if in a pH range from 5.5 to 7.0-7.5 the pepsin is stable and a simple return to a lower pH of 2.0 allows to recover a full activity of the enzyme, a pH greater than 7.5 irreversibly inactivates the pepsin [41] which will not allow its use once grafted.

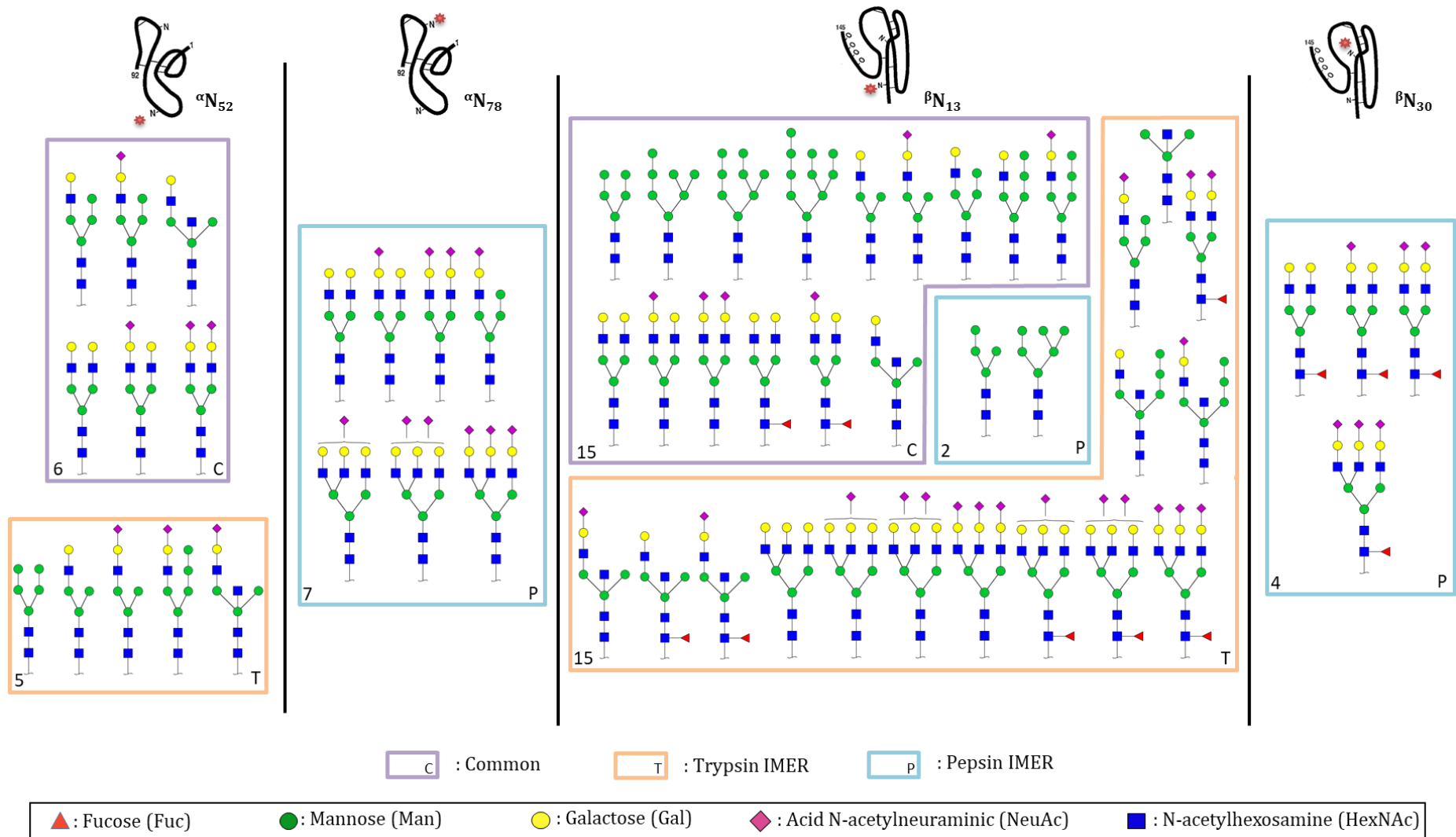


Figure 5. Glycans identified on glycopeptides after digestions on trypsin- and pepsin-based IMERs of r-hCG and analysis by LC-MS/MS. (Common identified N-glycans between both enzymes: 39%).

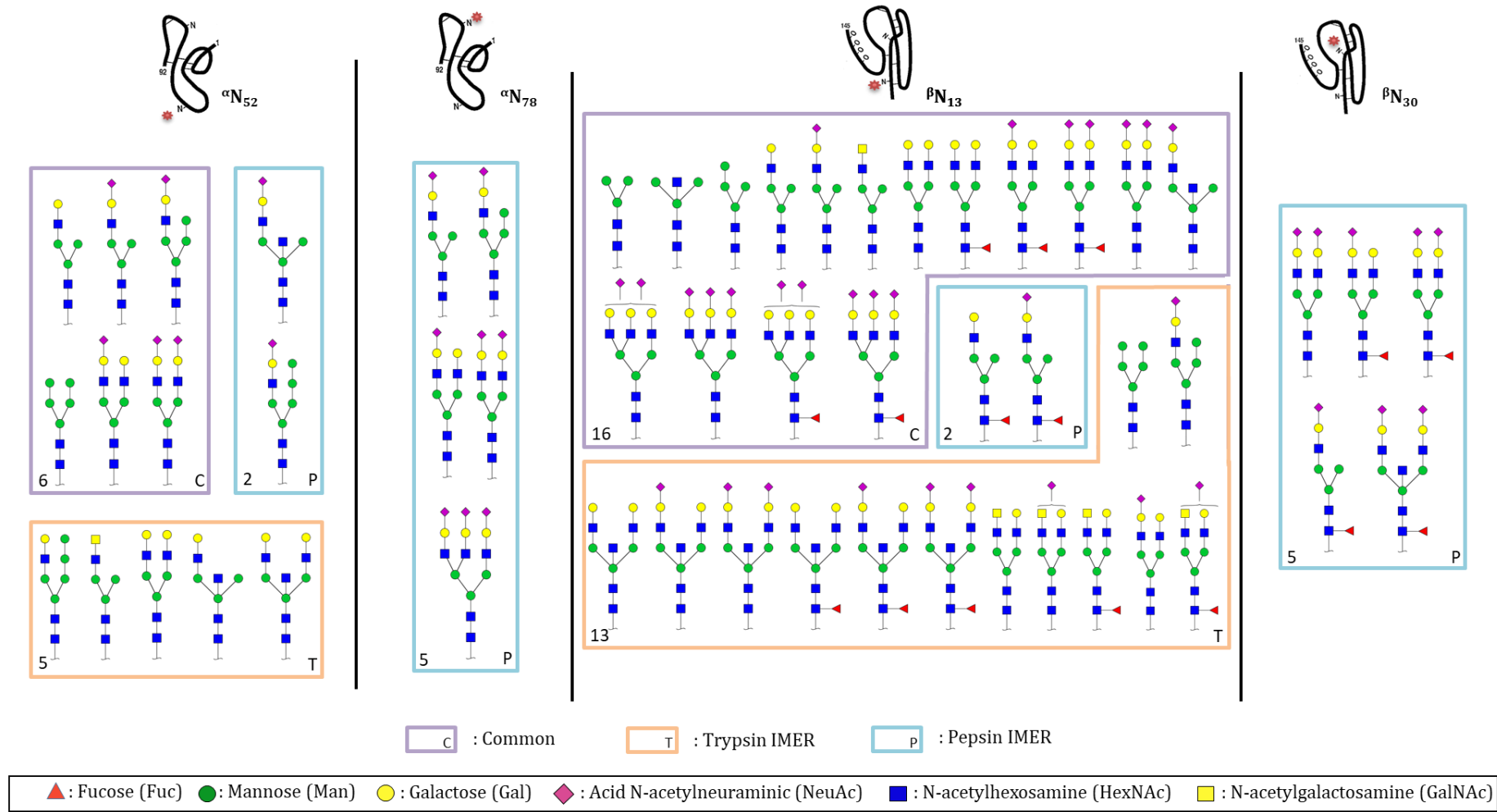


Figure 6. Glycans identified on glycopeptides after digestions on trypsin- and pepsin-based IMERs of u-hCG and analysis by LC-MS/MS. (Common identified N-glycans between both enzymes: 41%).

3.5. Potential of the IMERs for the glycosylation mapping

Having established the repeatability of the IMER digestion and the complementarity of our two IMERs for N-glycan mapping, the potential of the supports was further evaluated for the discrimination of hCG varying in production way. This is why a second hCG-based drug was considered. The source of production between the two drugs is different as r-hCG is a recombinant one expressed in Chinese Hamster Ovary cells while u-hCG is extracted from urine of pregnant women. Therefore, one can expect that they have two different glycosylation profiles. It is of interest to characterize the glycosylation profile because it is related to the biological activity of the protein [42]. The potential of the two IMERs for their digestion combined with LC-MS/MS analysis was evaluated in order to see if the fast N-glycosylation mapping led to significantly different results.

The analysis of the digests was carried out in duplicates and only glycans identified in both runs were further considered. The glycosylation mapping of u-hCG is reported in Figure 6. Once again, one should note the complementarity of trypsin and pepsin for the glycosylation mapping of u-hCG as only pepsin allowed the identification of glycans on αN_{78} and βN_{30} . Surprisingly, it seemed at first that the total number of glycans identified on the glycosylation sites was roughly the same between the two drugs. It is also noteworthy that the most intense peaks obtained with both digestions were attributed to complex N-glycans, thus being in good agreement with previous studies [39,40].

To compare the two drugs in detail, the characteristics of the N-glycans identified thanks to the digestions with both IMERs were compared for each drug and each glycosylation site (Table 2). It appears that both drugs have some similar characteristics. As an example, fucosylated glycans are not found on hCG α while most glycans identified on βN_{30} are fucosylated. The amount of glycans lacking sialic acids is also quite similar for r-hCG and u-hCG on each glycosylation site. With the exception on βN_{13} , the number of biantennary glycans is about the same for both drugs. But, Table 2 also demonstrates that both drugs have significantly different N-glycosylation mappings. For instance, glycans containing GalNAc were only identified in u-hCG. Moreover, the number of biantennary glycans was higher for u-hCG than for r-hCG.

Table 2. Comparison of the N-glycosylation mapping of r-hCG (batch 1) and u-hCG based on a relative qualitative characterization of the N-glycans identified on each N-glycosylation site after digestions on T-IMER and P-IMER and analysis by LC-MS/MS.

α N52	r-hCG	u-hCG	β N13	r-hCG	u-hCG
High mannose	9% (1)	8% (1)	High mannose	19% (6)	10% (3)
Monoantennary	64% (7)	62% (8)	Monoantennary	41% (13)	26% (8)
Biantennary	27% (3)	31% (4)	Biantennary	19% (6)	52% (16)
Triantennary	-	-	Triantennary	22% (7)	13% (4)
Not sialylated	46% (5)	54% (7)	Not sialylated	50% (16)	42% (13)
Fucosylated	-	-	Fucosylated	25% (8)	39% (12)
GalNAc containing	-	8% (1)	GalNAc containing	-	16% (5)
Total number	11	13	Total number	32	31

α N78	r-hCG	u-hCG	β N30	r-hCG	u-hCG
High mannose	14% (1)	-	High mannose	-	-
Monoantennary	-	40% (2)	Monoantennary	-	20% (1)
Biantennary	43% (3)	40% (2)	Biantennary	75% (3)	80% (4)
Triantennary	43% (3)	20% (1)	Triantennary	25% (1)	-
Not sialylated	14% (1)	-	Not sialylated	25% (1)	-
Fucosylated	-	-	Fucosylated	100% (4)	80% (4)
GalNAc containing	-	-	GalNAc containing	-	-
Total number	7	5	Total number	4	5

* Number of corresponding identified N-glycans

4. Conclusions

Two IMERs prepared by immobilizing either trypsin or pepsin were applied to the N-glycosylation mapping of a model highly glycosylated protein, *i.e.* hCG. After the evaluation of the repeatability of the analysis of a digest using T-IMER by nanoLC-MS/MS, each identified glycopeptide was validated by considering its retention time. This allowed to invalidate in-source fragmentation of two identifications that differ from a single monosaccharide thanks to differences in retention time of the glycopeptides. The identification of 81% of common N-glycans of a given batch of hCG highlighted the repeatability of the IMER digestion. Finding also 93.5% of common N-glycans after an

overnight digestion in solution and only 30 min using the IMER highlighted also the great potential of IMERs for glycosylation mapping.

The use of a pepsin-based IMER on the same hCG batch next demonstrated the complementarity of both IMERs. Indeed, despite trypsin allowed the identification of more glycans than pepsin on a specific glycosylation location, two of the four N-glycosylation sites were only characterized with a digestion with pepsin. The potential of these IMERs for glycosylation mapping was also demonstrated with the digestion of two hCG-based drugs leading to significantly different results. At last, IMERs could notably increase the performances of the whole analytical procedure due to their possible re-use, the low digestion time, and their possible integration in an on-line fully automated LC-MS set-up that decrease the analysis cost and improve accuracy.

Acknowledgement

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, and the Conseil Régional d'Île-de-France.

REFERENCES

- [1] O. Nørregaard Jensen, Modification-specific proteomics: characterization of post-translational modifications by mass spectrometry, *Curr. Opin. Chem. Biol.* 8 (2004) 33–41. doi:10.1016/j.cbpa.2003.12.009.
- [2] I.H.G.S. Consortium, Finishing the euchromatic sequence of the human genome, *Nature*. 431 (2004) 931–945.
- [3] Y. Jmeian, L.A. Hammad, Y. Mechref, Fast and Efficient Online Release of N-Glycans from Glycoproteins Facilitating Liquid Chromatography–Tandem Mass Spectrometry Glycomic Profiling, *Anal. Chem.* 84 (2012) 8790–8796. doi:10.1021/ac301855v.
- [4] K. Meller, P. Pomastowski, M. Szumski, B. Buszewski, Preparation of an improved hydrophilic monolith to make trypsin-immobilized microreactors, *J. Chromatogr. B.* 1043 (2017) 128–137. doi:10.1016/j.jchromb.2016.08.032.
- [5] F. Brothier, V. Pichon, Miniaturized DNA aptamer-based monolithic sorbent for selective extraction of a target analyte coupled on-line to nanoLC, *Anal. Bioanal. Chem.* 406 (2014) 7875–7886. doi:10.1007/s00216-014-8256-z.
- [6] Y. Zhang, B.R. Fonslow, B. Shan, M.-C. Baek, J.R. Yates, Protein Analysis by Shotgun/Bottom-up Proteomics, *Chem. Rev.* 113 (2013) 2343–2394. doi:10.1021/cr3003533.
- [7] B.G. Ng, H.H. Freeze, Perspectives on Glycosylation and Its Congenital Disorders, *Trends Genet.* (2018). doi:10.1016/j.tig.2018.03.002.
- [8] A. Tholey, A. Becker, Top-down proteomics for the analysis of proteolytic events - Methods, applications and perspectives, *Biochim. Biophys. Acta BBA - Mol. Cell Res.* 1864 (2017) 2191–2199. doi:10.1016/j.bbamcr.2017.07.002.
- [9] T.E. Angel, U.K. Aryal, S.M. Hengel, E.S. Baker, R.T. Kelly, E.W. Robinson, R.D. Smith, Mass spectrometry-based proteomics: existing capabilities and future directions, *Chem. Soc. Rev.* 41 (2012) 3912. doi:10.1039/c2cs15331a.
- [10] C. Temporini, E. Perani, E. Calleri, L. Dolcini, D. Lubda, G. Caccialanza, G. Massolini, Pronase-Immobilized Enzyme Reactor: an Approach for Automation in Glycoprotein Analysis by LC/LC–ESI/MSⁿ, *Anal. Chem.* 79 (2007) 355–363. doi:10.1021/ac0611519.
- [11] B.H. Clowers, E.D. Dodds, R.R. Seipert, C.B. Lebrilla, Site Determination of Protein Glycosylation Based on Digestion with Immobilized Nonspecific Proteases and Fourier Transform Ion Cyclotron Resonance Mass Spectrometry, *J. Proteome Res.* 6 (2007) 4032–4040. doi:10.1021/pr070317z.

- [12] N. Leymarie, J. Zaia, Effective Use of Mass Spectrometry for Glycan and Glycopeptide Structural Analysis, *Anal. Chem.* 84 (2012) 3040–3048. doi:10.1021/ac3000573.
- [13] H. Zhu, C. Qiu, A.C. Ruth, D.A. Keire, H. Ye, A LC-MS All-in-One Workflow for Site-Specific Location, Identification and Quantification of N-/O- Glycosylation in Human Chorionic Gonadotropin Drug Products, *AAPS J.* 19 (2017) 846–855. doi:10.1208/s12248-017-0062-z.
- [14] A. Kecskemeti, A. Gaspar, Preparation and characterization of a packed bead immobilized trypsin reactor integrated into a PDMS microfluidic chip for rapid protein digestion, *Talanta.* 166 (2017) 275–283. doi:10.1016/j.talanta.2017.01.060.
- [15] M. Bonichon, A. Combès, C. Desoubries, A. Bossée, V. Pichon, Development of immunosorbents coupled on-line to immobilized pepsin reactor and micro liquid chromatography–tandem mass spectrometry for analysis of butyrylcholinesterase in human plasma, *J. Chromatogr. A.* 1526 (2017) 70–81. doi:10.1016/j.chroma.2017.10.033.
- [16] M. Hedström, M. Andersson, I.Yu. Galaev, B. Mattiasson, Fast on-column protein digestion with subsequent peptide mapping using tandem mass spectrometry with information dependent acquisition, *J. Chromatogr. A.* 1080 (2005) 117–123. doi:10.1016/j.chroma.2005.04.069.
- [17] M. Naldi, U. Černigoj, A. Štrancar, M. Bartolini, Towards automation in protein digestion: Development of a monolithic trypsin immobilized reactor for highly efficient on-line digestion and analysis, *Talanta.* 167 (2017) 143–157. doi:10.1016/j.talanta.2017.02.016.
- [18] S. Moore, S. Hess, J. Jorgenson, Characterization of an immobilized enzyme reactor for on-line protein digestion, *J. Chromatogr. A.* 1476 (2016) 1–8. doi:10.1016/j.chroma.2016.11.021.
- [19] J. Sproß, A. Sinz, A Capillary Monolithic Trypsin Reactor for Efficient Protein Digestion in *Online* and *Offline* Coupling to ESI and MALDI Mass Spectrometry, *Anal. Chem.* 82 (2010) 1434–1443. doi:10.1021/ac9025362.
- [20] T. Šlechtová, M. Gilar, K. Kalíková, S.M. Moore, J.W. Jorgenson, E. Tesařová, Performance comparison of three trypsin columns used in liquid chromatography, *J. Chromatogr. A.* 1490 (2017) 126–132. doi:10.1016/j.chroma.2017.02.024.
- [21] L. Geiser, S. Eeltink, F. Svec, J.M.J. Fréchet, In-line system containing porous polymer monoliths for protein digestion with immobilized pepsin, peptide preconcentration and nano-liquid chromatography separation coupled to electrospray ionization mass spectroscopy, *J. Chromatogr. A.* 1188 (2008) 88–96. doi:10.1016/j.chroma.2008.02.075.
- [22] J. Carol-Visser, M. van der Schans, A. Fidder, A.G. Hulst, B.L.M. van Baar, H. Irth, D. Noort, Development of an automated on-line pepsin digestion–liquid chromatography–tandem mass spectrometry configuration for the rapid analysis of protein adducts of

- chemical warfare agents, *J. Chromatogr. B.* 870 (2008) 91–97. doi:10.1016/j.jchromb.2008.06.008.
- [23] T. Fournier, J. Guibourdenche, D. Evain-Brion, Review: hCGs: Different sources of production, different glycoforms and functions, *Placenta*. 36 (2015) S60–S65. doi:10.1016/j.placenta.2015.02.002.
- [24] L. Cole, hCG, Hyperglycosylated hCG, Pituitary hCG, Cancer hCG and Fetal hCG, *J. Pregnancy Child Health*. 03 (2015). doi:10.4172/2376-127X.1000222.
- [25] H. Toll, P. Berger, A. Hofmann, A. Hildebrandt, H. Oberacher, H.P. Lenhof, C.G. Huber, Glycosylation patterns of human chorionic gonadotropin revealed by liquid chromatography-mass spectrometry and bioinformatics, *Electrophoresis*. 27 (2006) 2734–2746. doi:10.1002/elps.200600022.
- [26] P. Berger, A.J. Laphorn, The molecular relationship between antigenic domains and epitopes on hCG, *Mol. Immunol.* 76 (2016) 134–145. doi:10.1016/j.molimm.2016.06.015.
- [27] A. Cingöz, F. Hugon-Chapuis, V. Pichon, Evaluation of various immobilized enzymatic microreactors coupled on-line with liquid chromatography and mass spectrometry detection for quantitative analysis of cytochrome c, *J. Chromatogr. A.* 1209 (2008) 95–103. doi:10.1016/j.chroma.2008.08.120.
- [28] P.K. Smith, R.I. Krohn, G.T. Hermanson, A.K. Mallia, F.H. Gartner, M.D. Provenzano, E.K. Fujimoto, N.M. Goeke, B.J. Olson, D.C. Klenk, Measurement of protein using bicinchoninic acid, *Anal. Biochem.* 150 (1985) 76–85. doi:10.1016/0003-2697(85)90442-7.
- [29] J. Rodriguez, N. Gupta, R.D. Smith, P.A. Pevzner, Does Trypsin Cut Before Proline?, *J. Proteome Res.* 7 (2008) 300–305. doi:10.1021/pr0705035.
- [30] M. Wührer, C.A.M. Koeleman, C.H. Hokke, A.M. Deelder, Mass spectrometry of proton adducts of fucosylated N-glycans: fucose transfer between antennae gives rise to misleading fragments, *Rapid Commun. Mass Spectrom.* 20 (2006) 1747–1754. doi:10.1002/rcm.2509.
- [31] G. Palmisano, M.R. Larsen, N.H. Packer, M. Thaysen-Andersen, Structural analysis of glycoprotein sialylation – part II: LC-MS based detection, *RSC Adv.* 3 (2013) 22706. doi:10.1039/c3ra42969e.
- [32] P. Kozlik, M. Sanda, R. Goldman, Nano reversed phase versus nano hydrophilic interaction liquid chromatography on a chip in the analysis of hemopexin glycopeptides, *J. Chromatogr. A.* 1519 (2017) 152–155. doi:10.1016/j.chroma.2017.08.066.
- [33] O. Ozohanics, L. Turiák, A. Puerta, K. Vékey, L. Drahos, High-performance liquid chromatography coupled to mass spectrometry methodology for analyzing site-specific N-glycosylation patterns, *J. Chromatogr. A.* 1259 (2012) 200–212. doi:10.1016/j.chroma.2012.05.031.

- [34] G.C.M. Vreeker, M. Wuhrer, Reversed-phase separation methods for glycan analysis, *Anal. Bioanal. Chem.* 409 (2017) 359–378. doi:10.1007/s00216-016-0073-0.
- [35] C. Pett, W. Nasir, C. Sihlbom, B.-M. Olsson, V. Caixeta, M. Schorlemer, R.P. Zahedi, G. Larson, J. Nilsson, U. Westerlind, Effective Assignment of α 2,3/ α 2,6-Sialic Acid Isomers by LC-MS/MS-Based Glycoproteomics, *Angew. Chem. Int. Ed.* 57 (2018) 9320–9324. doi:10.1002/anie.201803540.
- [36] M. Pabst, J.S. Bondili, J. Stadlmann, L. Mach, F. Altmann, Mass + Retention Time = Structure: A Strategy for the Analysis of *N*-Glycans by Carbon LC-ESI-MS and Its Application to Fibrin *N*-Glycans, *Anal. Chem.* 79 (2007) 5051–5057. doi:10.1021/ac070363i.
- [37] Z. An, Q. Shu, H. Lv, L. Shu, J. Wang, F. Yang, Y. Fu, N-Linked Glycopeptide Identification Based on Open Mass Spectral Library Search, *BioMed Res. Int.* 2018 (2018) 1–11. doi:10.1155/2018/1564136.
- [38] W.-F. Zeng, M.-Q. Liu, Y. Zhang, J.-Q. Wu, P. Fang, C. Peng, A. Nie, G. Yan, W. Cao, C. Liu, H. Chi, R.-X. Sun, C.C.L. Wong, S.-M. He, P. Yang, pGlyco: a pipeline for the identification of intact N-glycopeptides by using HCD- and CID-MS/MS and MS3, *Sci. Rep.* 6 (2016). doi:10.1038/srep25102.
- [39] L. Valmu, H. Alfthan, K. Hotakainen, S. Birken, U.-H. Stenman, Site-specific glycan analysis of human chorionic gonadotropin -subunit from malignancies and pregnancy by liquid chromatography--electrospray mass spectrometry, *Glycobiology.* 16 (2006) 1207–1218. doi:10.1093/glycob/cwl034.
- [40] A. Gervais, Y.-A. Hammel, S. Pelloux, P. Lepage, G. Baer, N. Carte, O. Sorokine, J.-M. Strub, R. Koerner, E. Leize, A. Van Dorsselaer, Glycosylation of human recombinant gonadotrophins: characterization and batch-to-batch consistency, *Glycobiology.* 13 (2003) 179–189. doi:10.1093/glycob/cwg020.
- [41] D.W. Piper, B.H. Fenton, pH stability and activity curves of pepsin with special reference to their clinical importance., *Gut.* 6 (1965) 506–508. doi:10.1136/gut.6.5.506.
- [42] L. Diaz-Cueto, J. Barrios-de-Tomasi, C. Timossi, J.P. Mendez, A. Ulloa-Aguirre, More in-vitro bioactive, shorter-lived human chorionic gonadotrophin charge isoforms increase at the end of the first and during the third trimesters of gestation, *Mol. Hum. Reprod.* 2 (1996) 643–650. doi:10.1093/molehr/2.9.643.

Supplementary Material

Table S.1. *In silico* simulation of the expected glycopeptides from human hCG for a digestion with trypsin. (N-glycosylation sites are tagged in bold). Data obtained from <https://www.expasy.org/>.

α subunit	Resulting peptide sequence	Peptide mass [Da]
$^{\alpha}\text{N}_{52}$	N ₅₂ VTSESTCCVAK	1241.397
$^{\alpha}\text{N}_{78}$	VEN ₇₈ HTACHCSTCYHK	1896.100
β subunit	Resulting peptide sequence	Peptide mass [Da]
$^{\beta}\text{N}_{13}$	CRPIN ₁₃ ATLAVEK	1314.566
$^{\beta}\text{N}_{30}$	EGCPVCITVN ₃₀ TTICAGYCPTMTR	2433.871

Glycosylation site ⁶ N ₁₃ (continued)	Proposed structure	Repeatability of the analysis			Repeatability of the digestion			Repeatability batch-to-batch		
		R1	R2	R3	D1	D2	D3	R1	R2	R3
HexNAc(4)Hex(5)		X	X	X	X	X	X	X	X	X
HexNAc(4)Hex(5)Fuc(1)		X	X	X	X	X	X	X	X	X
HexNAc(4)Hex(5)Fuc(1)NeuAc(1)		X	X	X	X	X	X	X	X	X
HexNAc(4)Hex(5)Fuc(1)NeuAc(2)		X	X	X	X	X	X	X	X	X
HexNAc(4)Hex(5)NeuAc(1)		X	X	X	X	X	X	X	X	X
HexNAc(4)Hex(5)NeuAc(2)		X	X	X	X	X	X	X	X	X
HexNAc(4)Hex(6)							X			
HexNAc(4)Hex(6)NeuAc(1)		X		X	X		X			
HexNAc(5)Hex(6)						X	X	X		X
HexNAc(5)Hex(6)Fuc(1)NeuAc(1)						X	X	X		X
HexNAc(5)Hex(6)Fuc(1)NeuAc(2)		X	X	X	X	X	X	X	X	X
HexNAc(5)Hex(6)Fuc(1)NeuAc(3)		X	X	X	X	X	X	X	X	X
HexNAc(5)Hex(6)NeuAc(1)		X	X	X		X	X	X	X	X

Table S.3. Peptides (> 5 amino acids) with glycosylation site identified by Byonic™ after a digestion on T-IMER and on P-IMER of r-hCG (batch 1) and analysis by LC-MS/MS.

αN_{52}	Peptide sequence after digestion on T-IMER	Peptide sequence after digestion on P-IMER
	N ₅₂ VTSESTCCVAK	SRAYPTPLRSKKTMLVQKN ₅₂ VTSE RAYPTPLRSKKTMLVQKN ₅₂ VTSES PTPLRSKKTMLVQKN ₅₂ VTSESTCC VQKN ₅₂ VTSESTCCVAKSYNRVTVMGGF VQKN ₅₂ VTSESTCCVAKSYNRVTVMGG VQKN ₅₂ VTSESTCCVAKSYNRVTVM VQKN ₅₂ VTSESTCCVAKSYNRV VQKN ₅₂ VTSESTCCVAKSYNRVTV VQKN ₅₂ VTSESTCCVAKSYNRVTVMG VQKN ₅₂ VTSESTCCVAKSYN VQKN ₅₂ VTSESTCCVAKS VQKN ₅₂ VTSESTCC QKN ₅₂ VTSESTCCV KN ₅₂ VTSESTCCVAKSYNRVTVM
αN_{78}	Peptide sequence after digestion on T-IMER	Peptide sequence after digestion on P-IMER
	-	NRVTVMGGFKVEN ₇₈ HTACHCSTCY MGGFKVEN ₇₈ HTACHCSTC

GGFKVEN₇₈HTACHCSTCY

GGFKVEN₇₈HTACHCSTCYHKS

^αN₇₈
(continued) Peptide sequence after digestion on T-IMER Peptide sequence after digestion on P-IMER

GGFKVEN₇₈HTACHCSTC

GFKVEN₇₈HTACHCSTCY

GFKVEN₇₈HTACHCSTC

KVEN₇₈HTACHCSTCY

KVEN₇₈HTACHCSTC

^βN₁₃ Peptide sequence after digestion on T-IMER Peptide sequence after digestion on P-IMER

CRPIN₁₃ATLAVEK

TWASKEPLRPRCRPIN₁₃ATL

PIN₁₃ATLAVEK

ASKEPLRPRCRPIN₁₃ATL

SKEPLRPRCRPIN₁₃ATL

SKEPLRPRCRPIN₁₃ATLA

EPLRPRCRPIN₁₃ATL

PLRPRCRPIN₁₃AT

PLRPRCRPIN₁₃ATLA

PRCRPIN₁₃ATLAVEKEG

N₁₃ATLAV

^βN₃₀ Peptide sequence after digestion on T-IMER Peptide sequence after digestion on P-IMER

^βN₃₀

-

ITVN₃₀TTIC

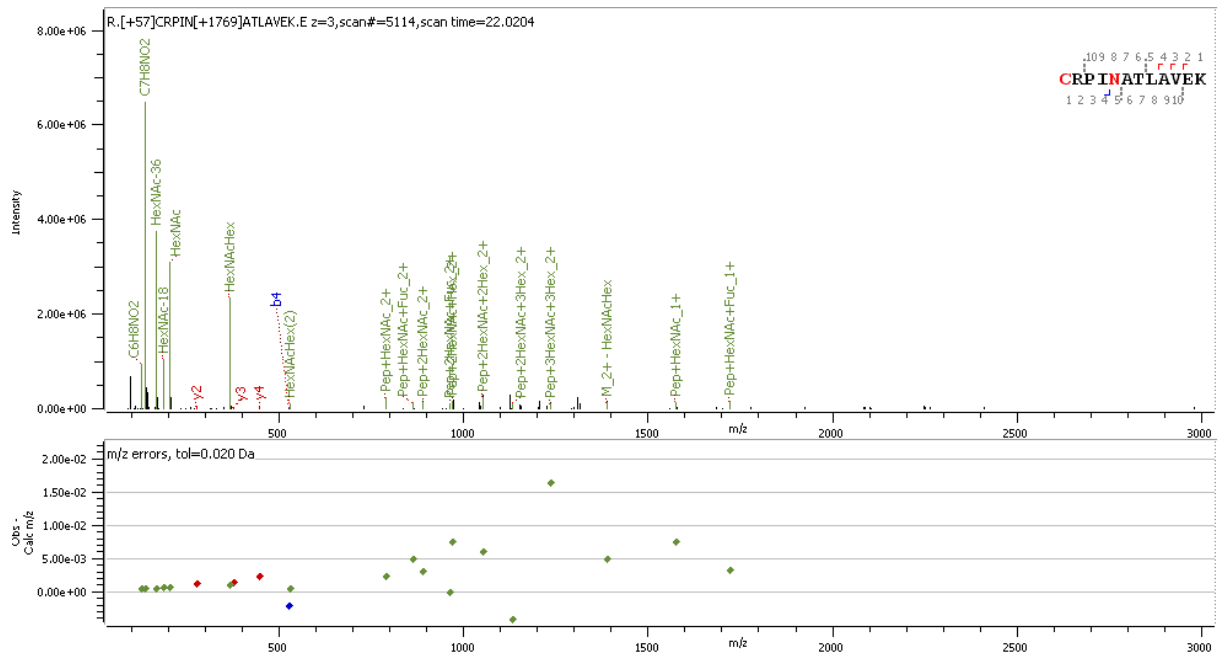


Figure S.1. MS/MS spectrum of a glycopeptide identified on θ N13 after a tryptic digestion

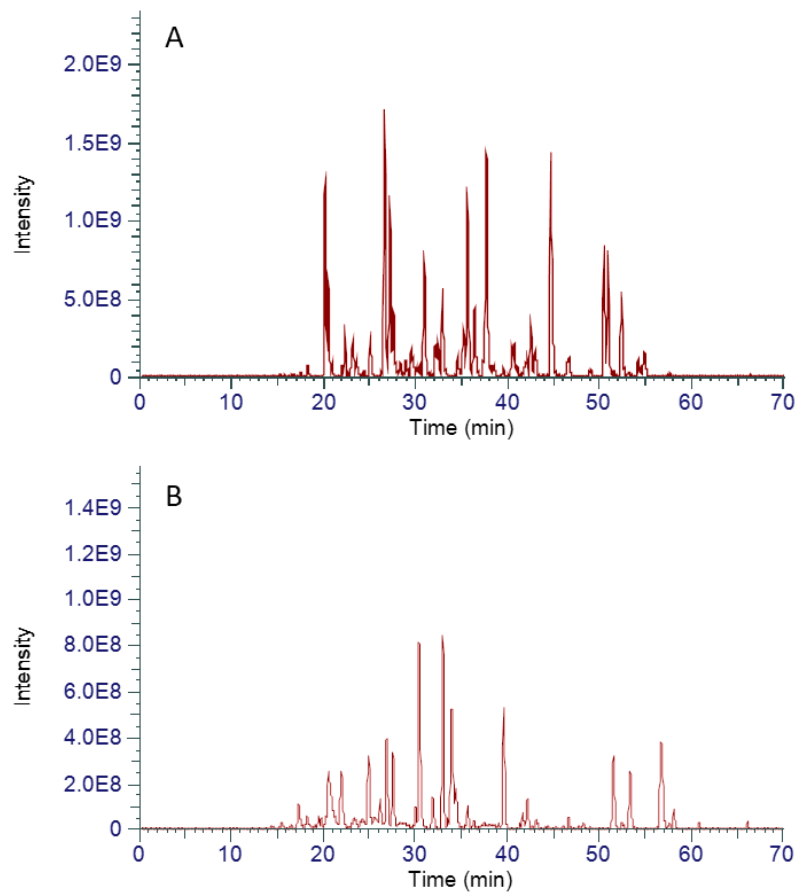


Figure S.2. Base Peak Chromatogram obtained after a digestion of 2.5 μ g of r-hCG on (A) T-IMER and (B) P-IMER followed by a nanoLC-MS/MS analysis